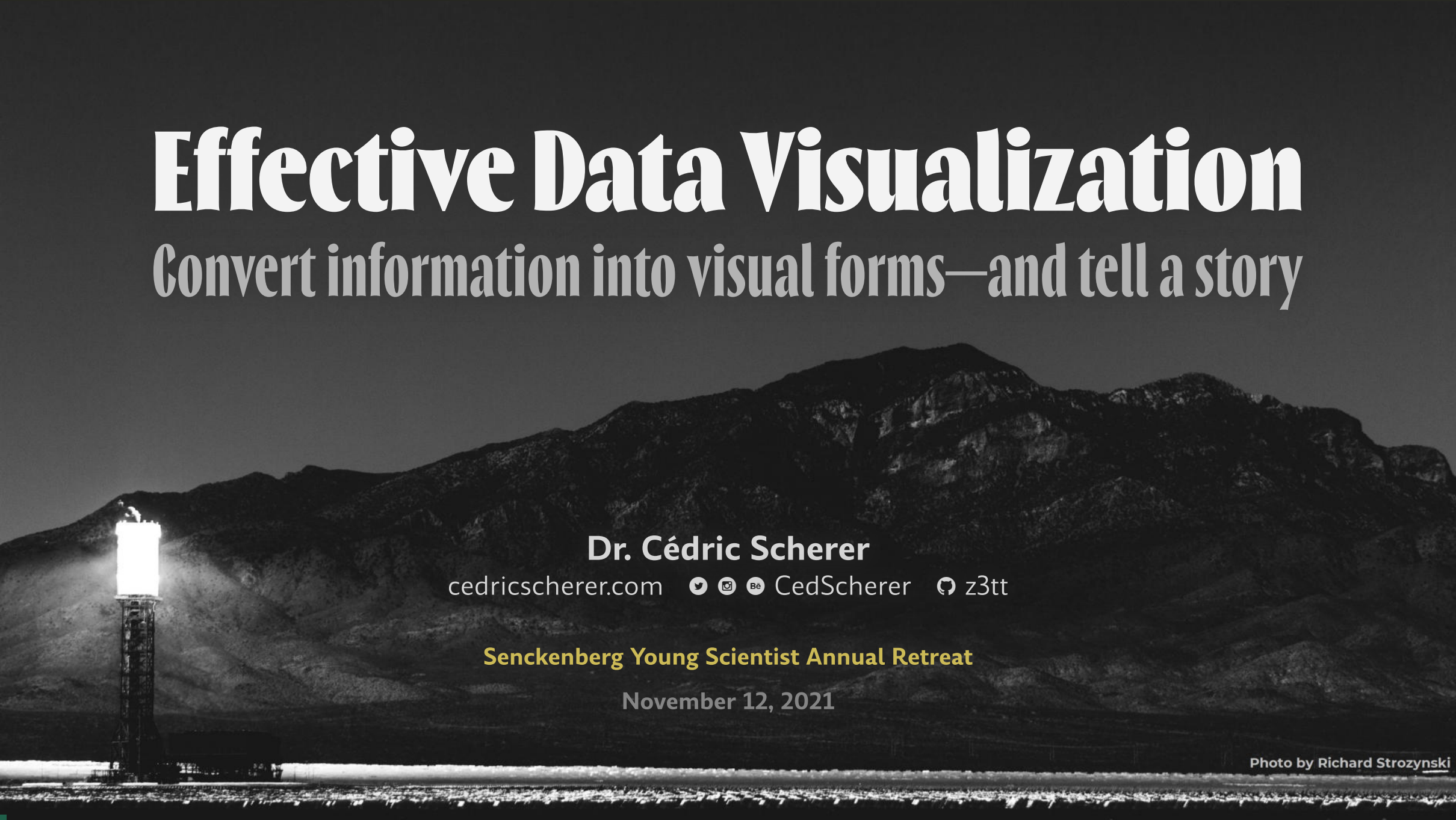
Effective Data Visualization

Convert information into visual forms—and tell a story



Dr. Cédric Scherer

cedricscherer.com CedScherer z3tt

Senckenberg Young Scientist Annual Retreat

November 12, 2021

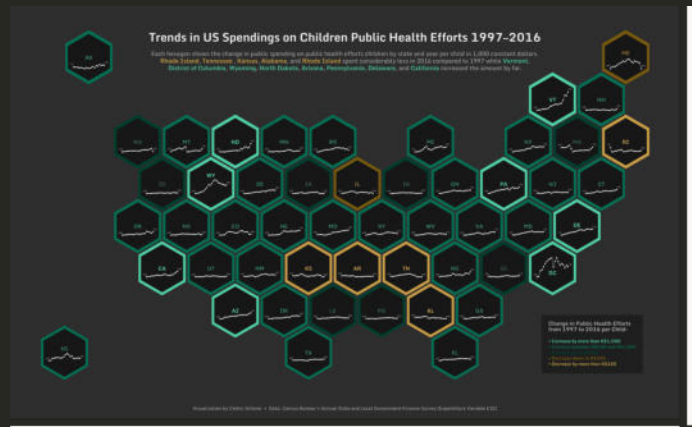
Photo by Richard Strozynski

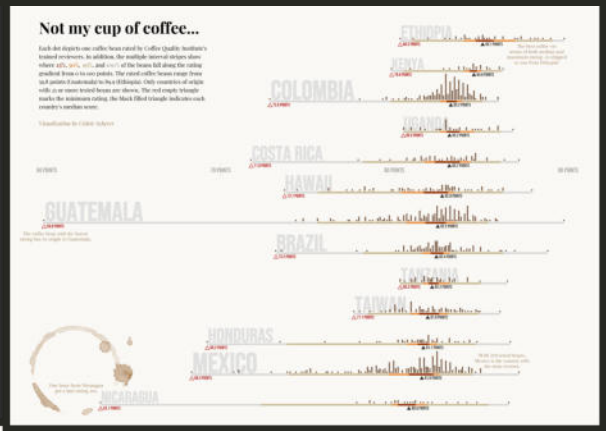
CÉDRIC SCHERER

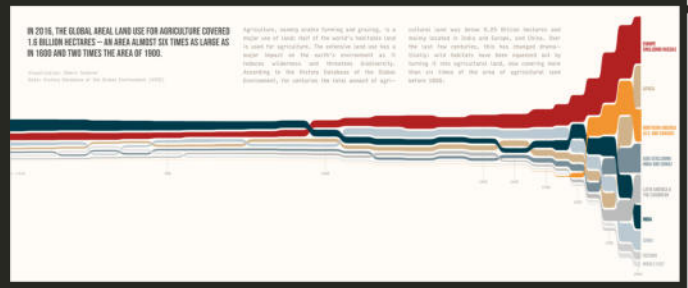
Independent Data Visualization Specialist
Computational Ecologist at IZW Berlin

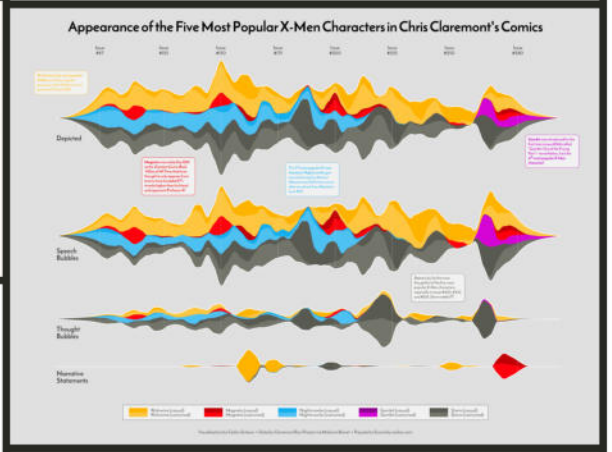
Consulting **Coaching** **Coding**

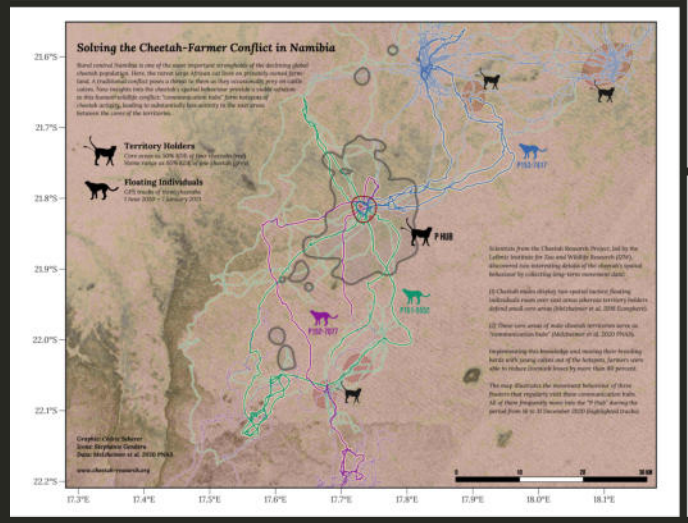
cedricscherer.com @CedScherer @z3tt @cedscherer

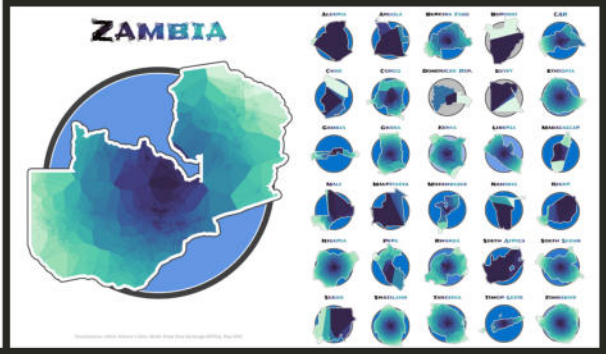












SCIENTIFIC AMERICAN

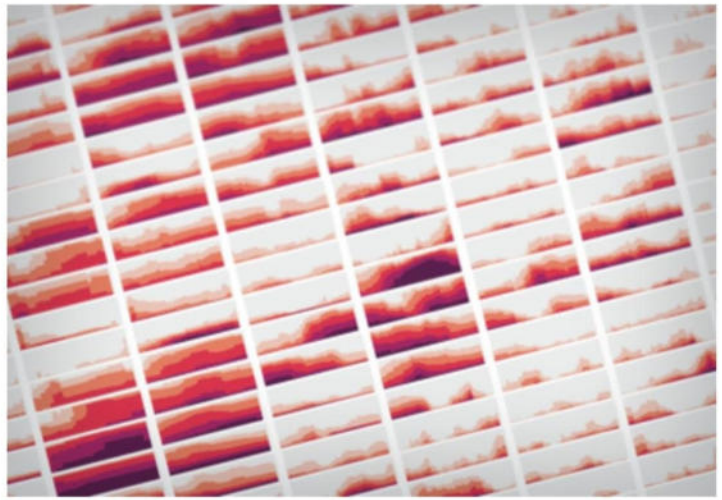
Subscribe

CLIMATE CHANGE

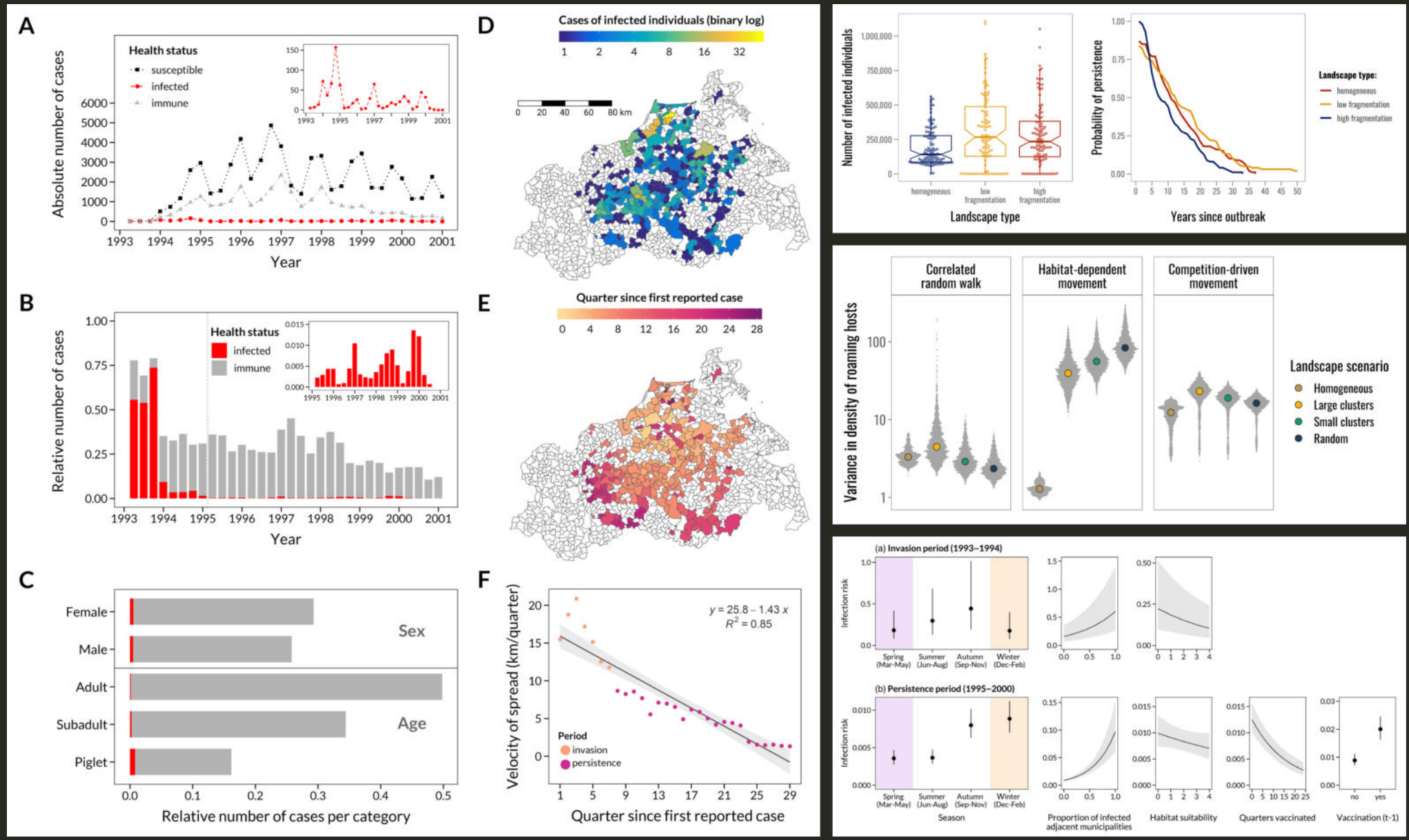
Climate Change Drives Escalating Drought

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

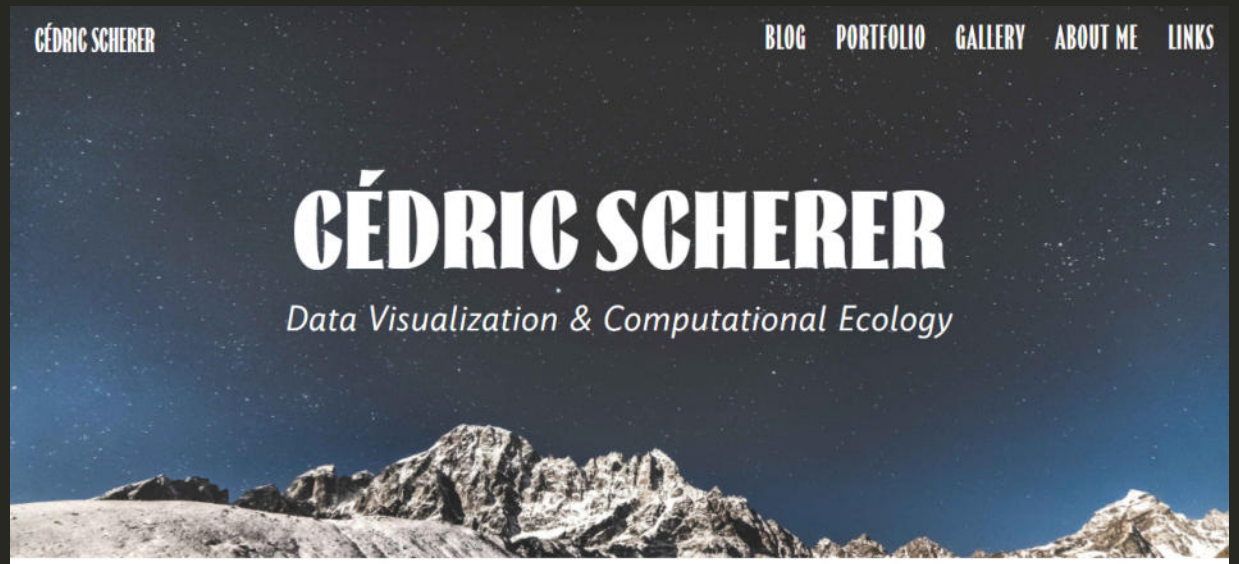
By Clara Moskowitz, Cédric Scherer, Georgios Karamanis | Scientific American November 2021 Issue



Credit: Cédric Scherer and Georgios Karamanis


A
Health status
susceptible
infected
immune
Absolute number of cases
Year
B
Health status
infected
immune
Relative number of cases
Year
C
Female
Male
Sex
Adult
Subadult
Piglet
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

cedricscherer.com



The World's Countries Colored by Their First Letter

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

Posted by Cédric • Friday, August 27, 2021

A Quick How-to on Labelling Bar Graphs in ggplot2

Bar charts are likely the most common chart type out there and come in several varieties. Most notably, Direct labels can increase accessibility of a bar graph. I got a request how one can add percentage labels inside the bars and how to highlight specific bars with {ggplot2}. This short tutorial shows you multiple ways how to do so.

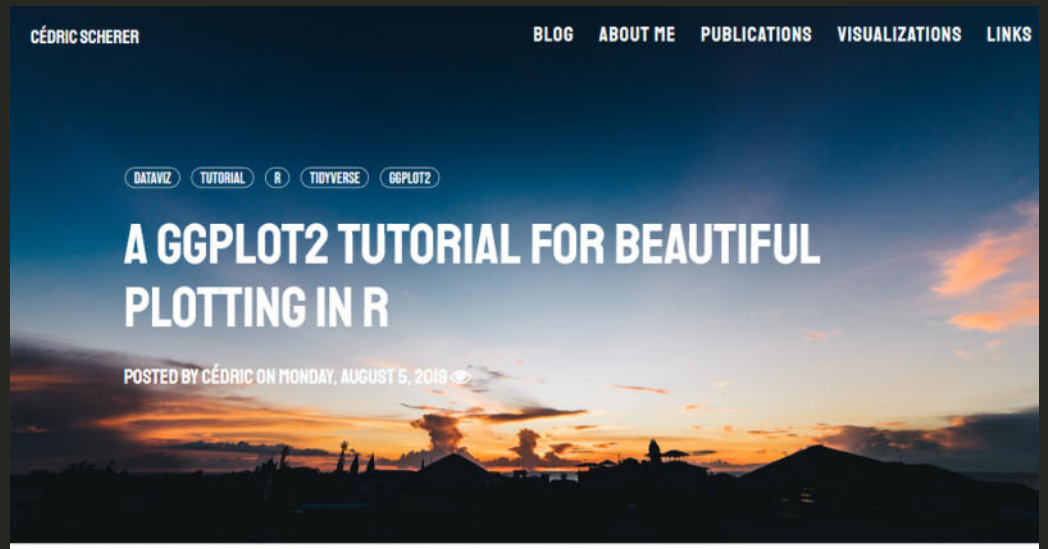
Posted by Cédric • Monday, July 5, 2021

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

QUICK LINKS

ggplot2 Tutorial

cedricscherer.com



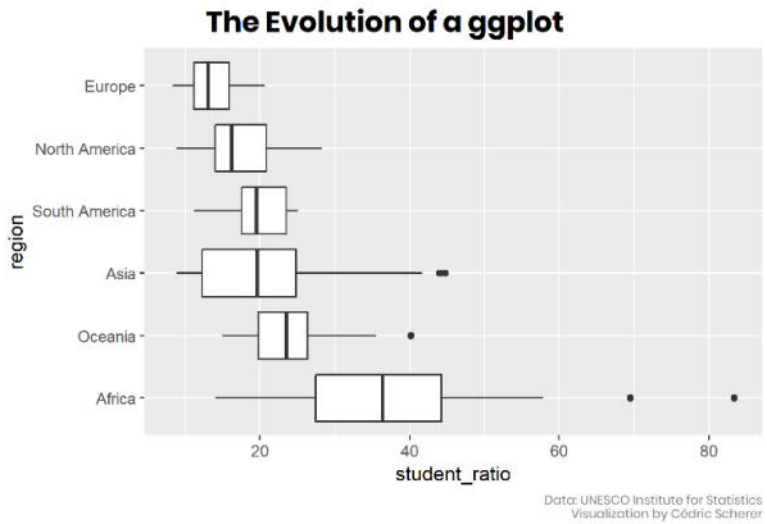
Last update: 2019-11-01

INTRODUCTORY WORDS

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

Begin of 2016, I had to prepare my PhD introductory talk and I started using `{ggplot2}` to visualize my data since I never liked the syntax and style of base plots in R. Because I was short on time, I plotted these figures by try'n'error and with the help of lots of googling. The resource I came always back to was a blog entry called Beautiful plotting in R: A ggplot2 cheatsheet by Zev Ross, posted on 4. August 2014, updated last in January 2016. After giving the talk which contained some quite beautiful plots thanks to the blog post, I decided to go through this tutorial step-by-step. I learned so much from it and directly started modifying the codes and over the time I added some additional code snippets, chart types and resources.

Since the blog entry by Zev Ross was not updated for some years, I hosted the updated version on my GitHub. Now it finds its proper place on this homepage! (Plus I added some updates, for example the fantastic `{patchwork}` and `{gganimate}` packages. And pie charts



- Aim of this Tutorial
- Data Preparation
- The Default Boxplot

Data Visualizations

are any graphical representation
of information and data

Data Visualizations

convert information into visual forms as quantifiable features

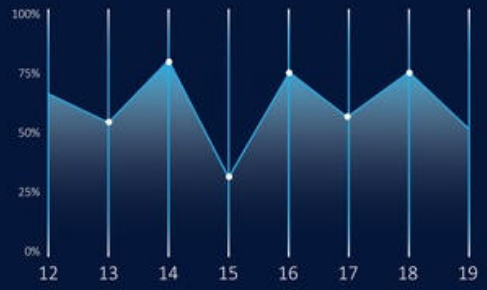
Data Visualizations

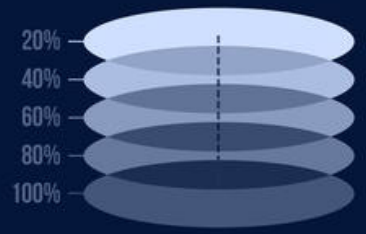
help to amplify cognition, gain insights, discover, explain, and make decisions

Data Visualizations

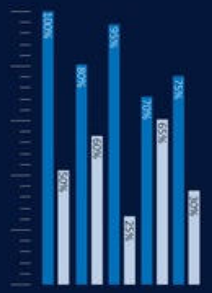
are part art and part science

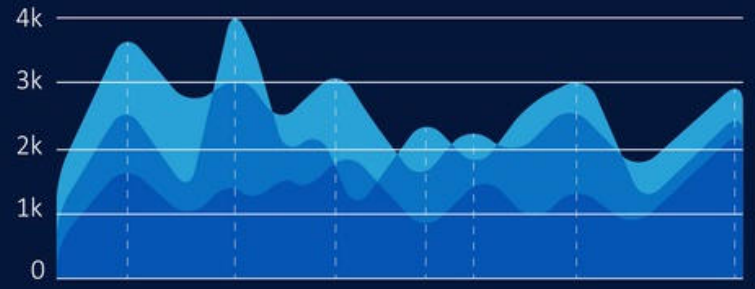

45 29 26 79
68 84 95 41
78 27 29 62
68 24 45 62
26 36 48 84
52 34 27 27
36 38 26 65
50 100 150 200
400 200 0
7%
3%
19%
1200 1000 800 600 400 200 0
1 2 3 4 5 6 7 8 9
100 90 80
-10
40 30 20 10
5 6 7 8 9 10 11 12
764 1004
955 950
764 1004
3%

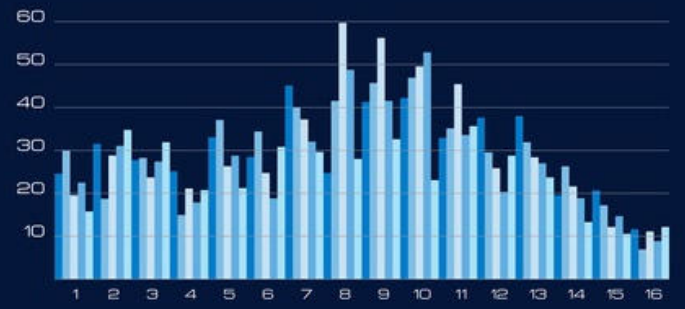

100%
75%
50%
25%
0%
12
13
14
15
16
17
18
19

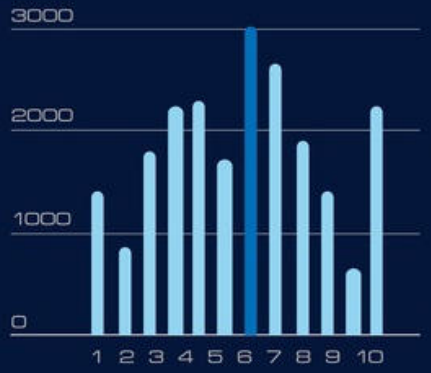

20%
40%
60%
80%
100%

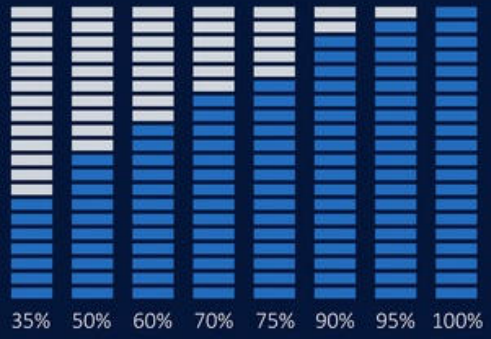

30%
50%
70%
100%




4k
3k
2k
1k
0

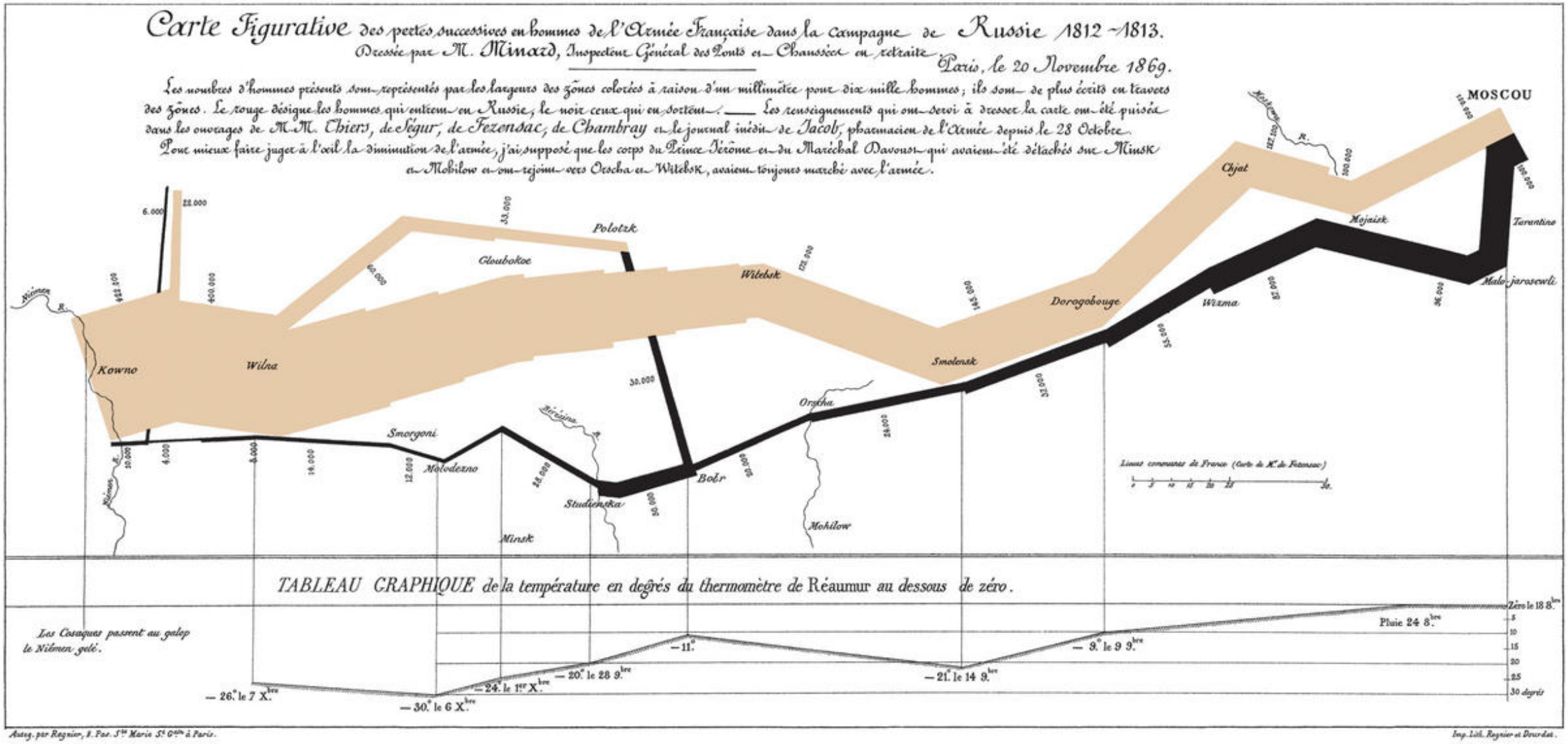

60
50
40
30
20
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16


3000
2000
1000
0
1 2 3 4 5 6 7 8 9 10

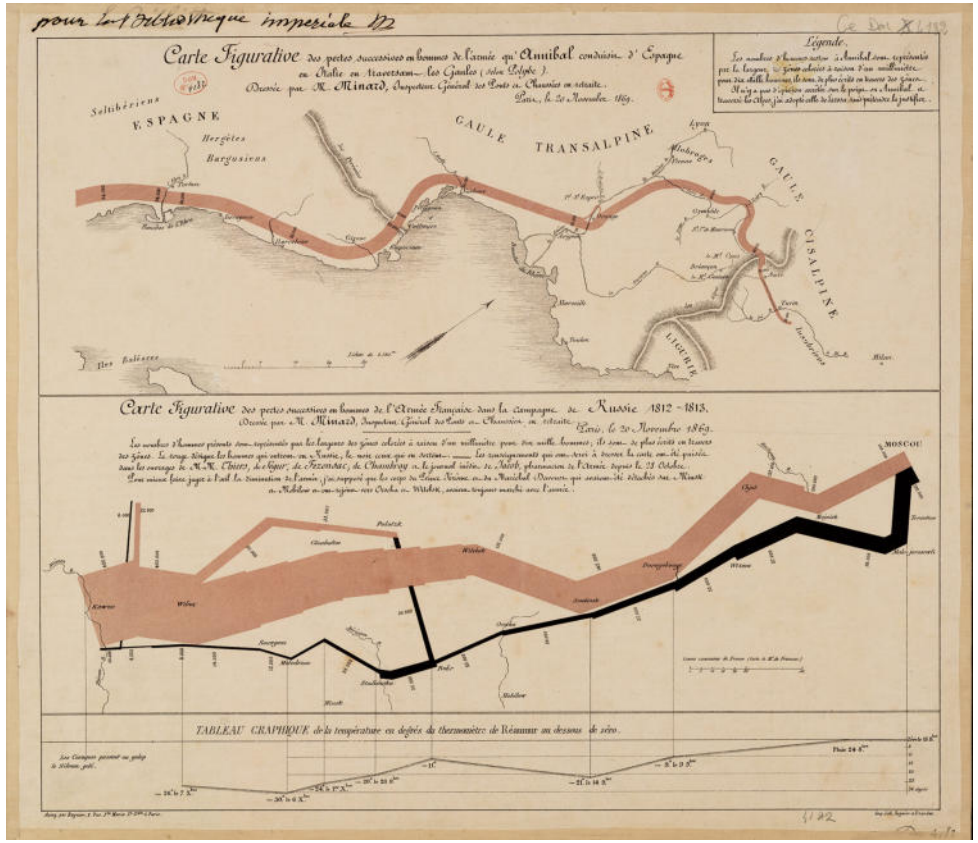

35%
50%
60%
70%
75%
90%
95%
100%


4k
3k
2k
1k
0
JAN
FEB
MAR
APR
MAY


Volume
Average 50
Aug
Sep
Oct
20
5
9
16
24
30
7
13
20
28
4
11
18
40.00K
30.00K
20.00K
10.00K
0.00



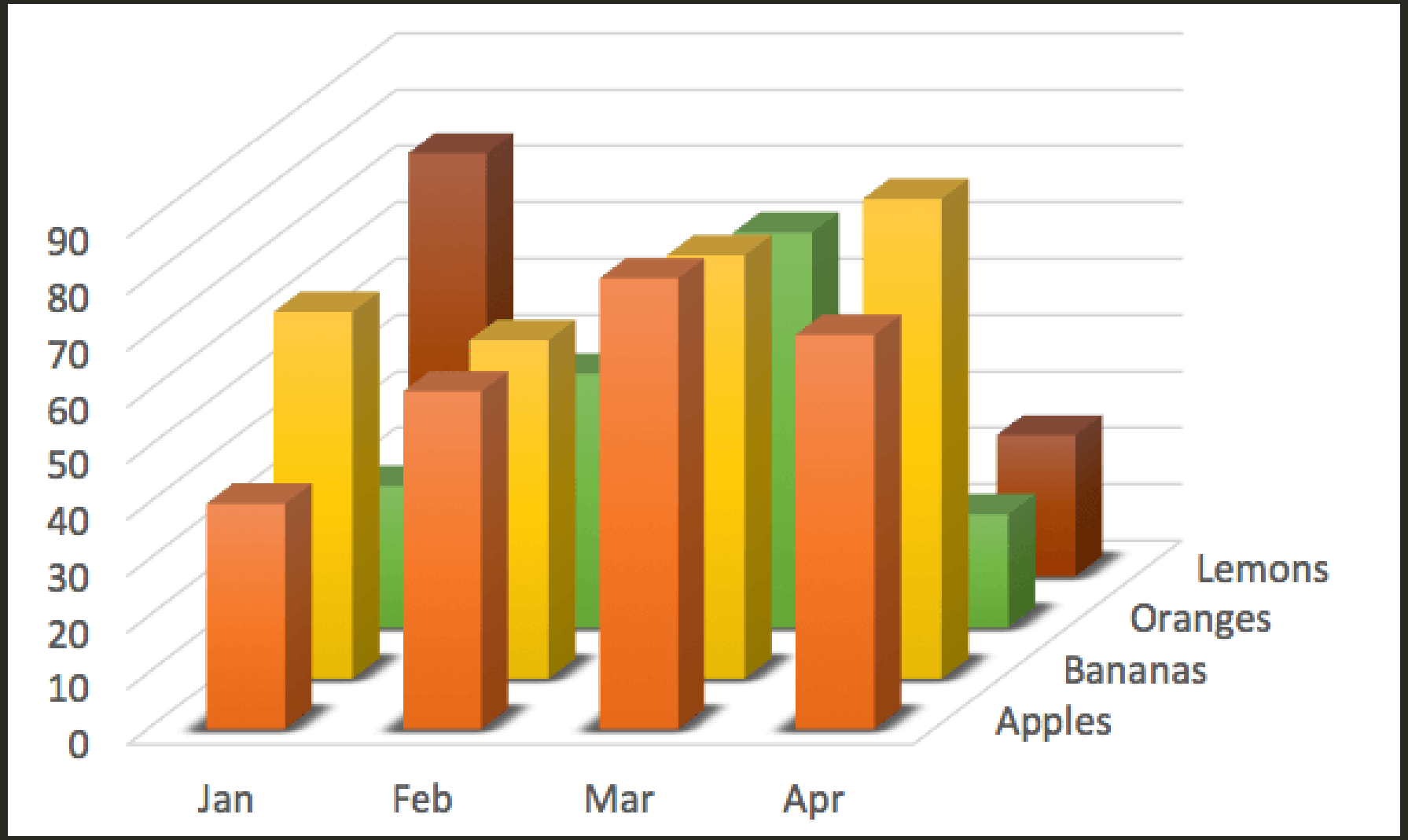
Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 by Charles Joseph Minard



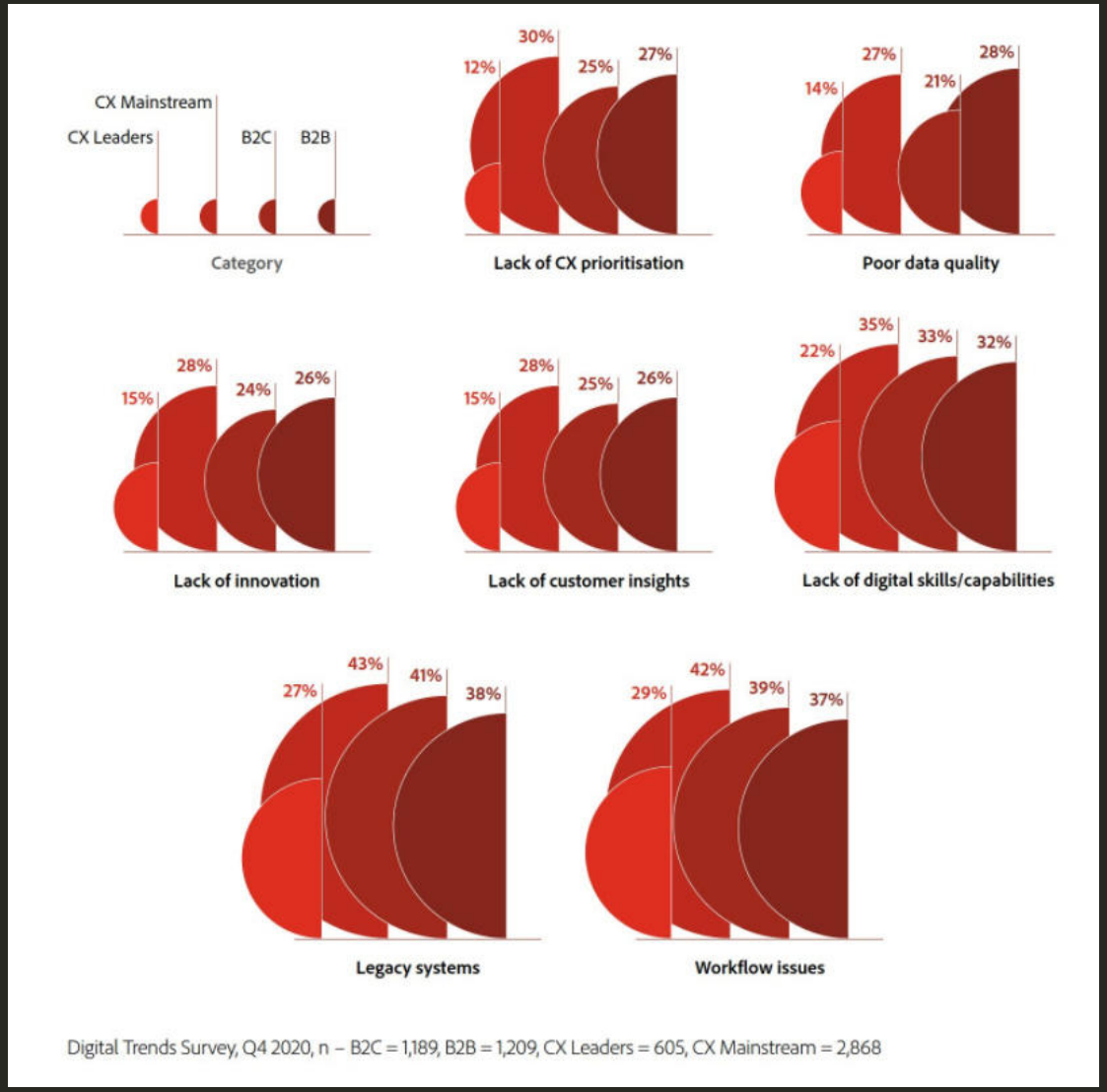
Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 and Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe) by Charles Joseph Minard

- displays the progress of the troops of Hannibal (218 BC) and Napoleon (1812–1813) in the form of a stream
- often considered as the **best statistical graphic ever drawn**

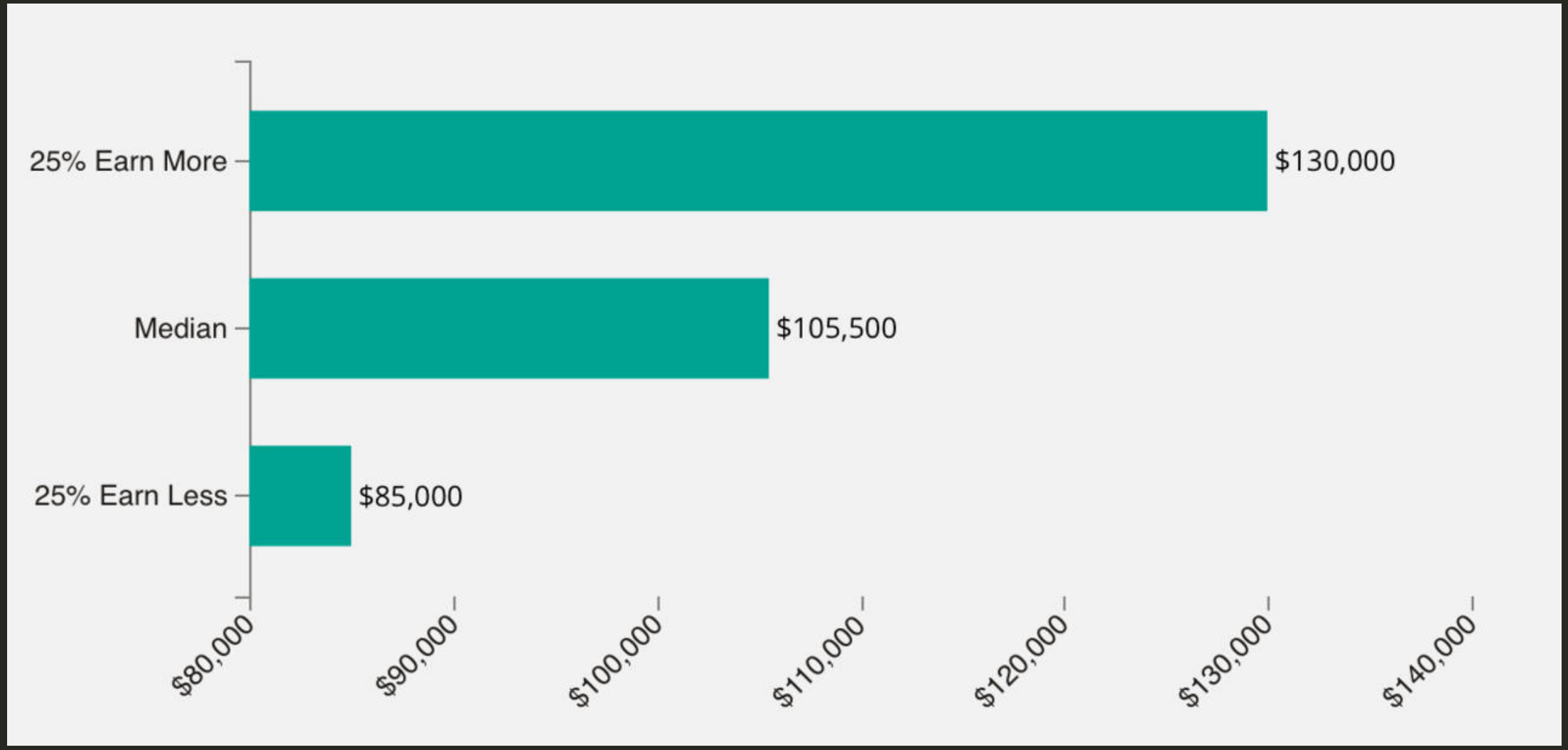
What Makes It a Bad Data Visualization?



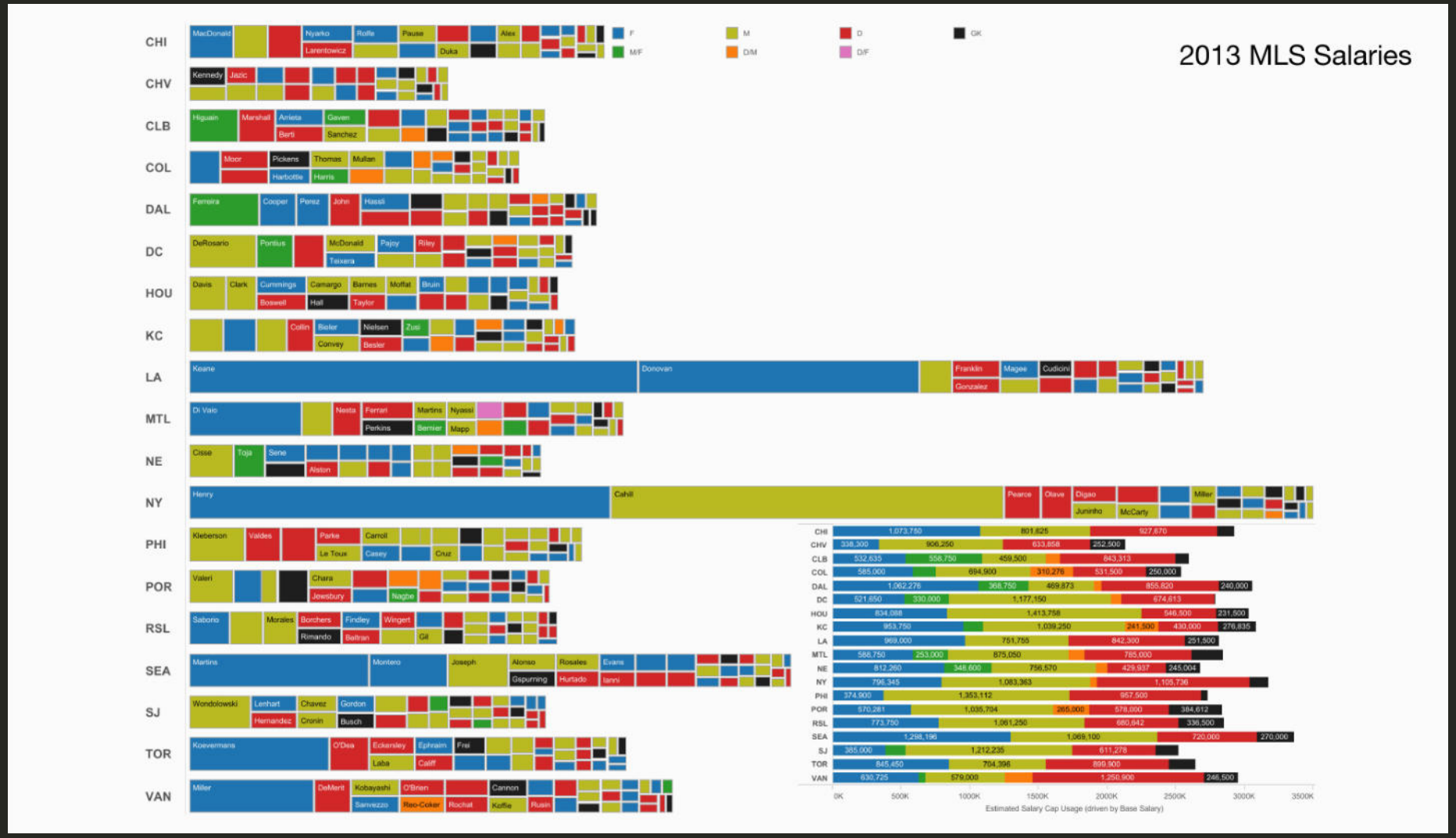
What Makes It a Bad Data Visualization?



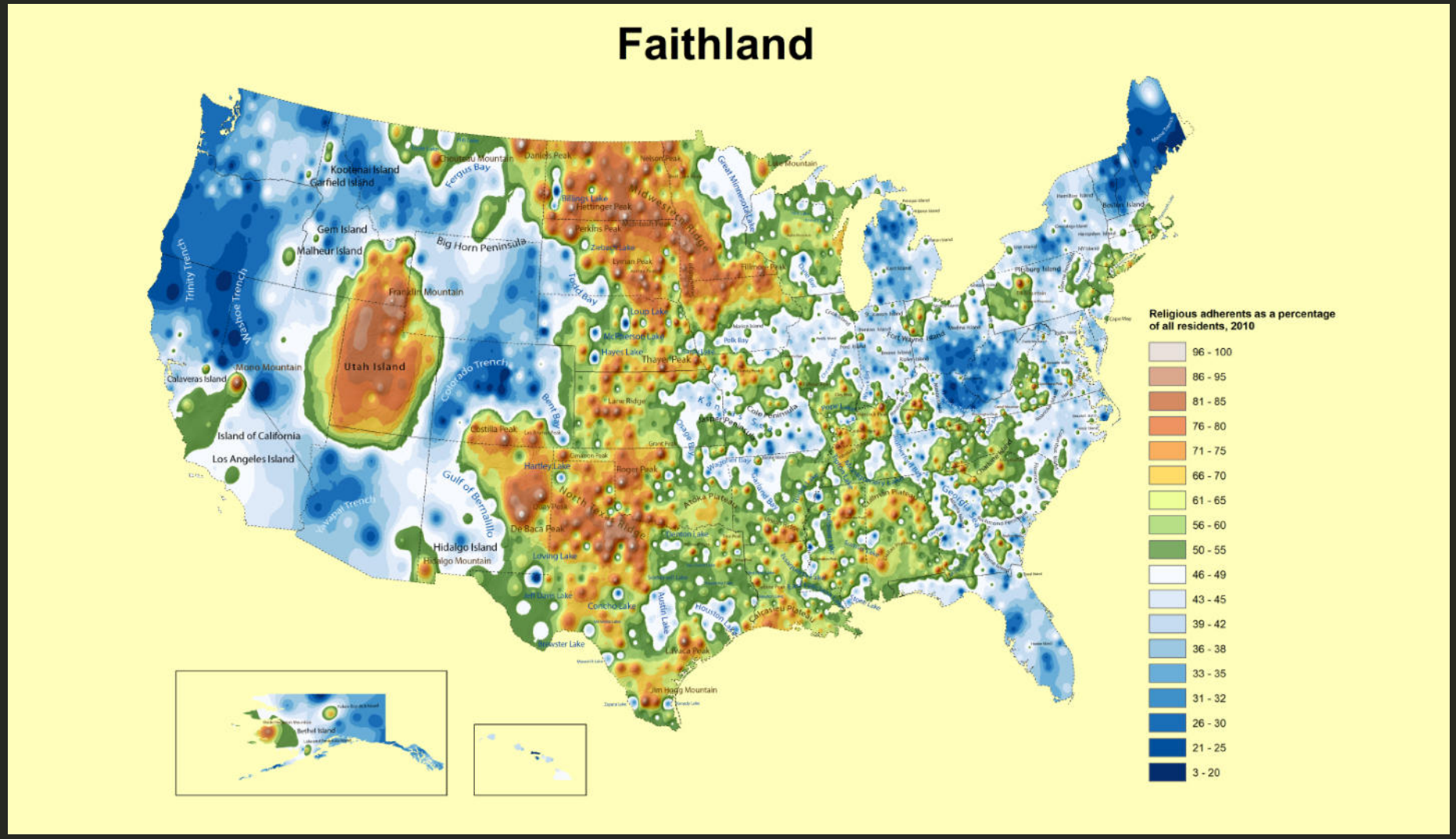
What Makes It a Bad Data Visualization?



What Makes It a Bad Data Visualization?



What Makes It a Bad Data Visualization?



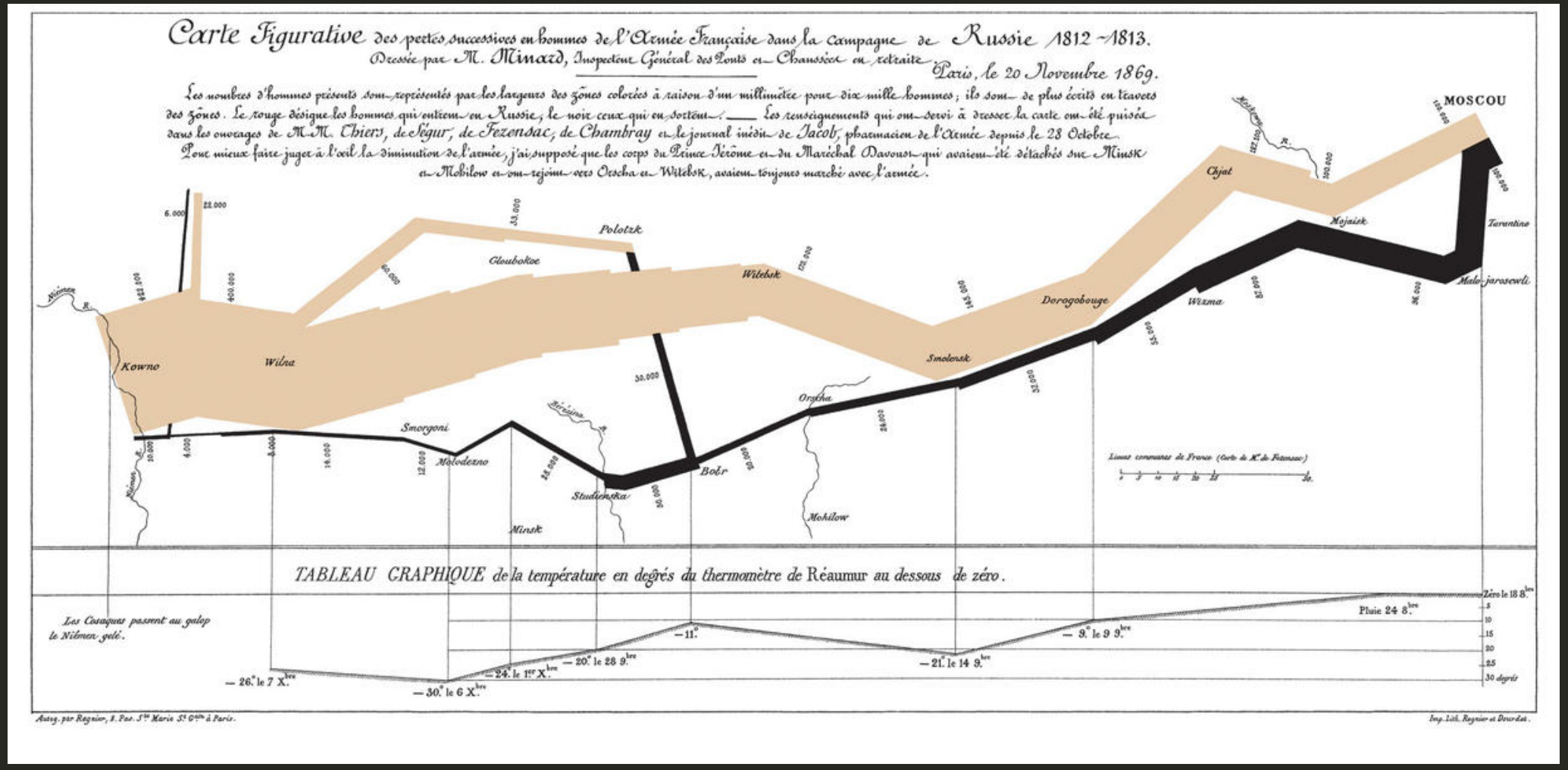
What Makes It a Bad Data Visualization

→ **Substantive problems** (bad data/story)

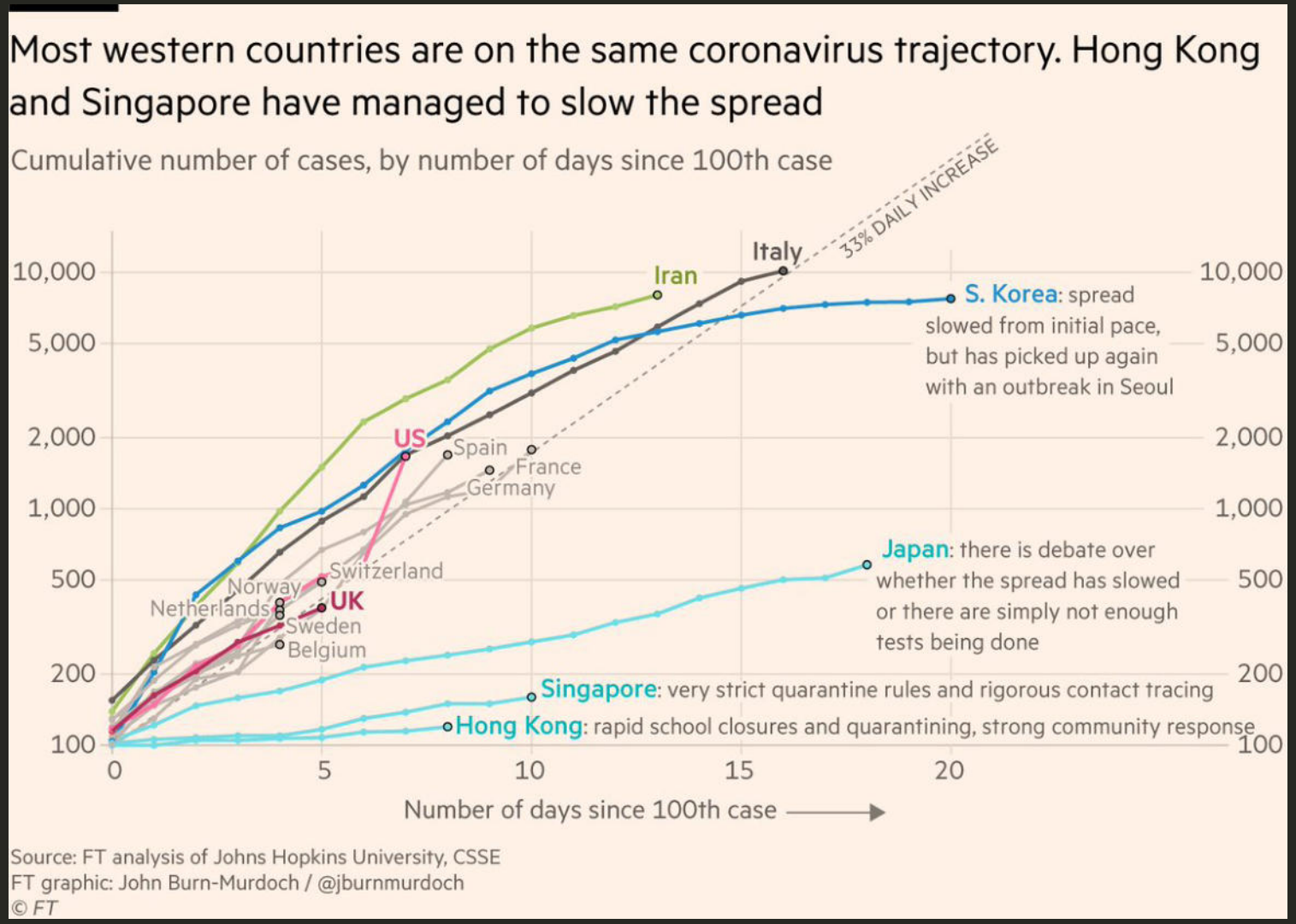
→ **Aesthetic problems** (bad design)

→ **Perceptual problems** (bad perception)

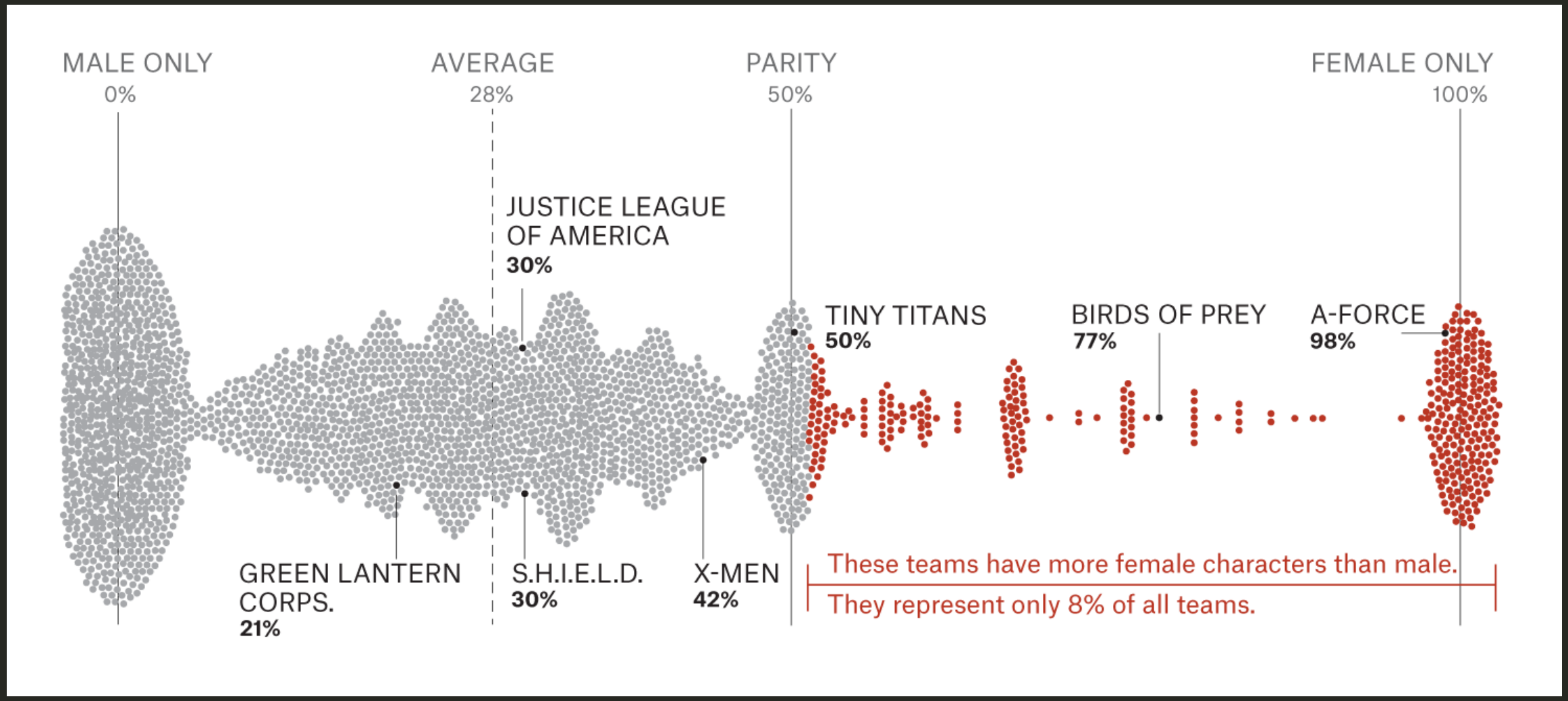
What Makes It a Good Data Visualization?



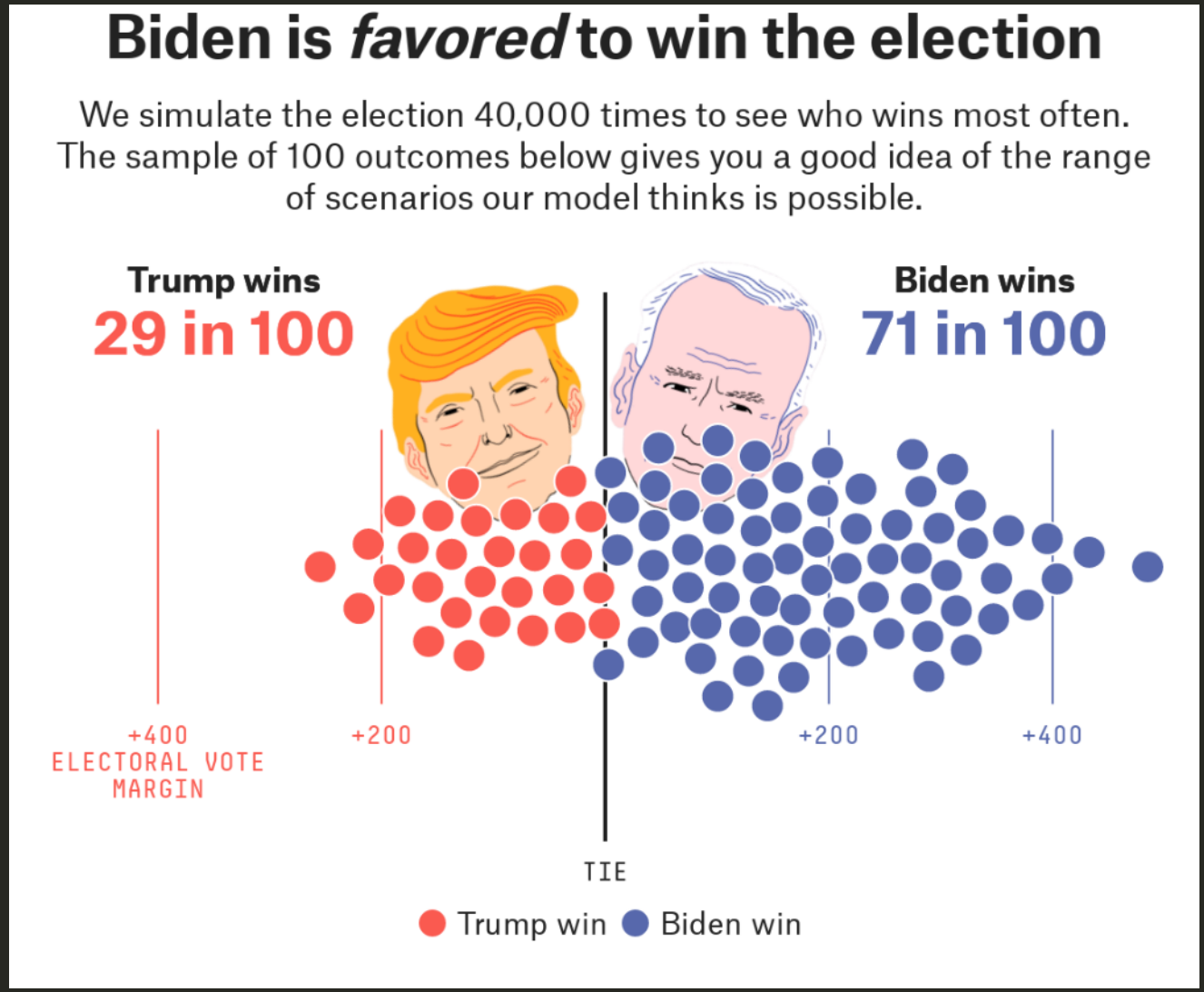
What Makes It a Good Data Visualization?



What Makes It a Good Data Visualization?



What Makes It a Good Data Visualization?



What Makes It a Good Data Visualization

→ **Information** (Integrity)

→ **Story** (Interestingness)

→ **Goal** (Usefulness)

→ **Visual Form** (Beauty)

Information

Understand your data and be accurate

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:** precision and errors

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

Source: inhomelandsecurity.com/risk-management-and-black-swan-events

Don't form a **singular statement**:

"The swan is white"

Instead, confront yourself with a **falsifiable universal statement**:

"All swans are white"

Know Your

Types of Data

Know Your
Types of Data

Quantitative Data
versus
Qualitative Data

Quantitative Data (Numerical)

Illustration by Allison Horst

Qualitative Data (Categorical)

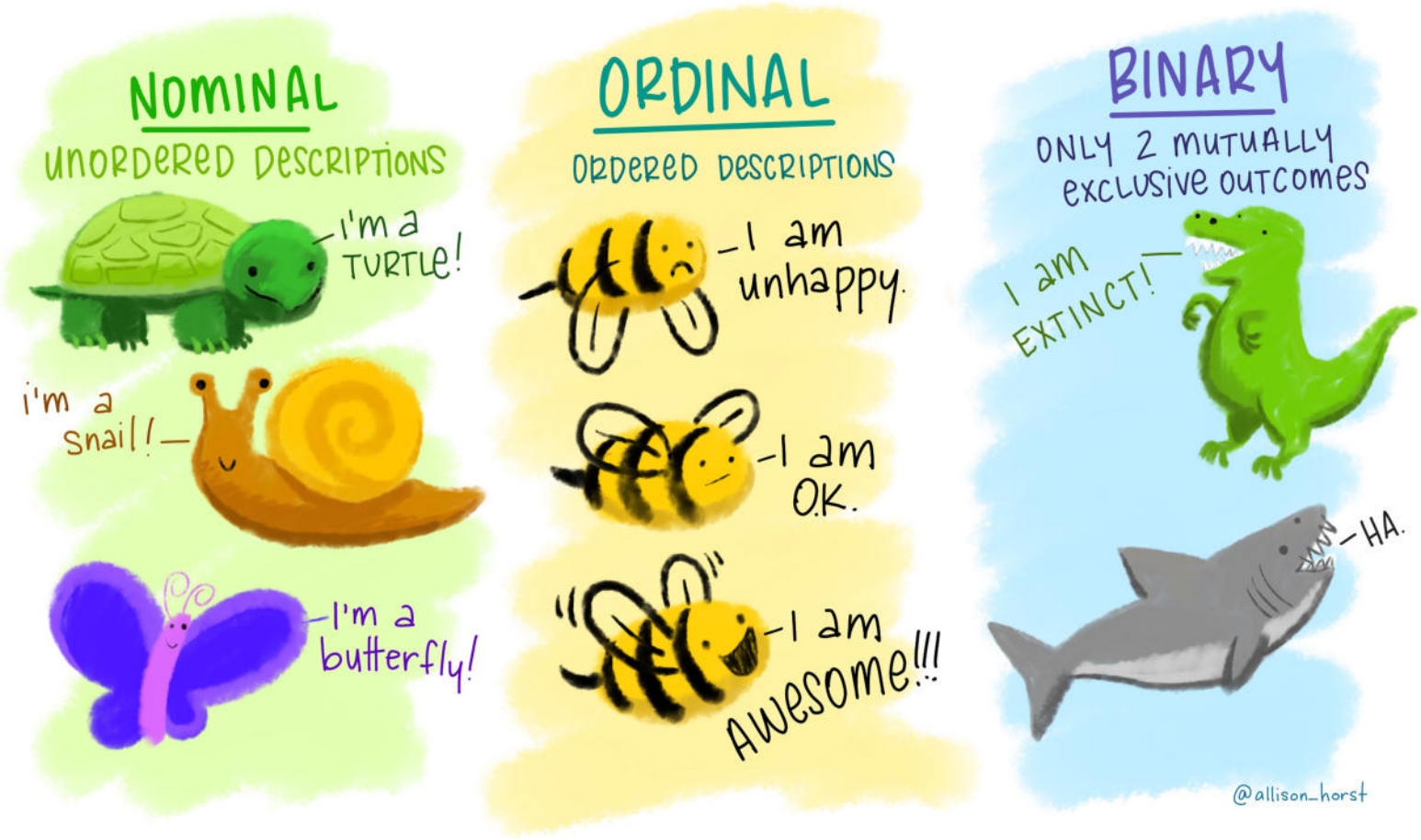


Illustration by Allison Horst

Types of Data

- **Quantitative (numerical)** versus **qualitative (categorical)** data
- **Ordered (ordinal)** versus **unordered (nominal)** data
- **Continuous** versus **discrete** data

Story

Be clear about the message of your visualization

It is hard to effectively design for others.
It is harder if you don't even care to try.

Andy Kirk

Who is my audience?

Which story is **interesting** for my audience?

What are **relevant** details to include?

Which variables are **meaningful** to my audience?

How will they **encounter** the visualization?

Who is my audience?

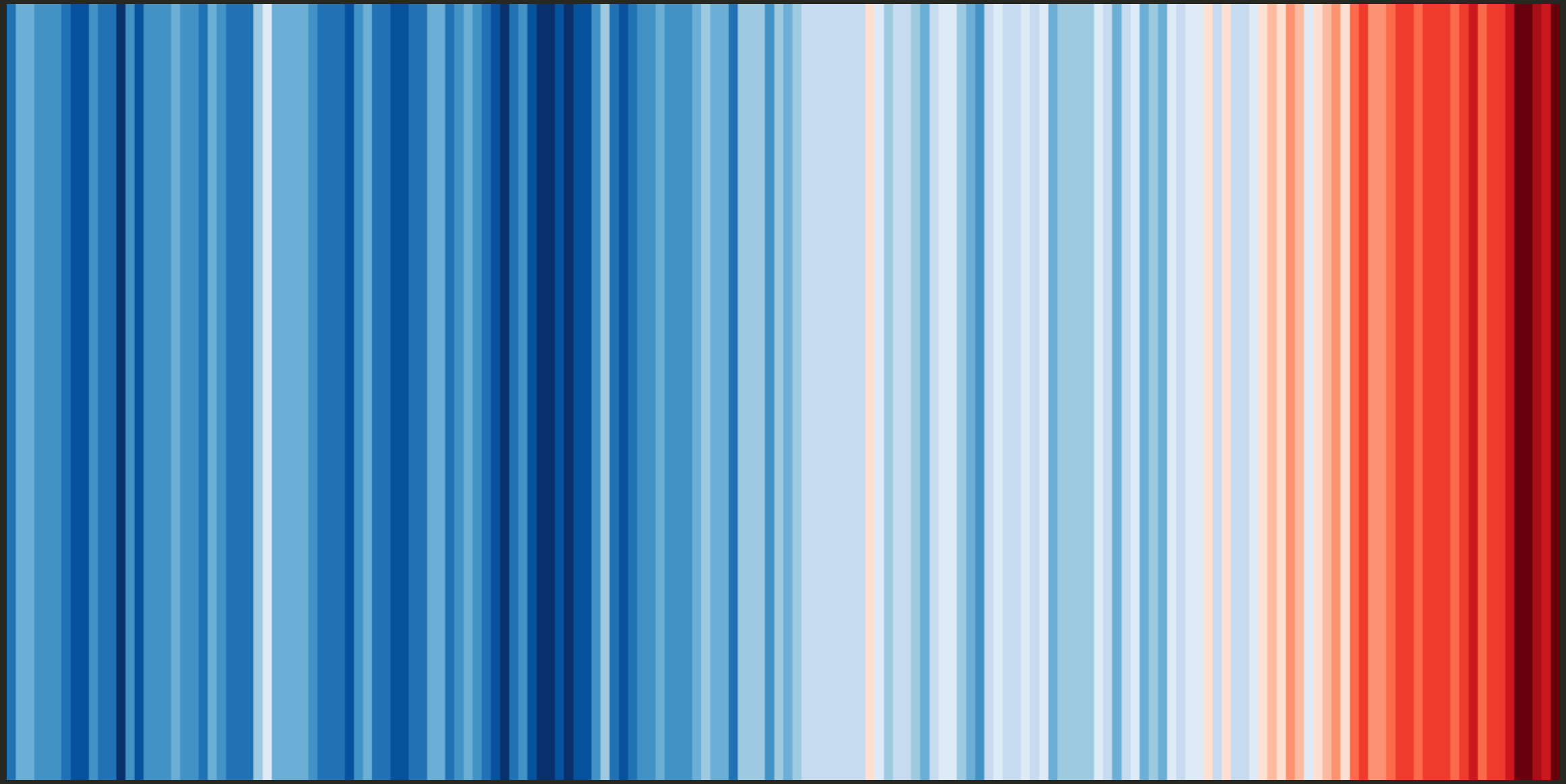
Which story is **interesting** for my audience?

What are **relevant** details to include?

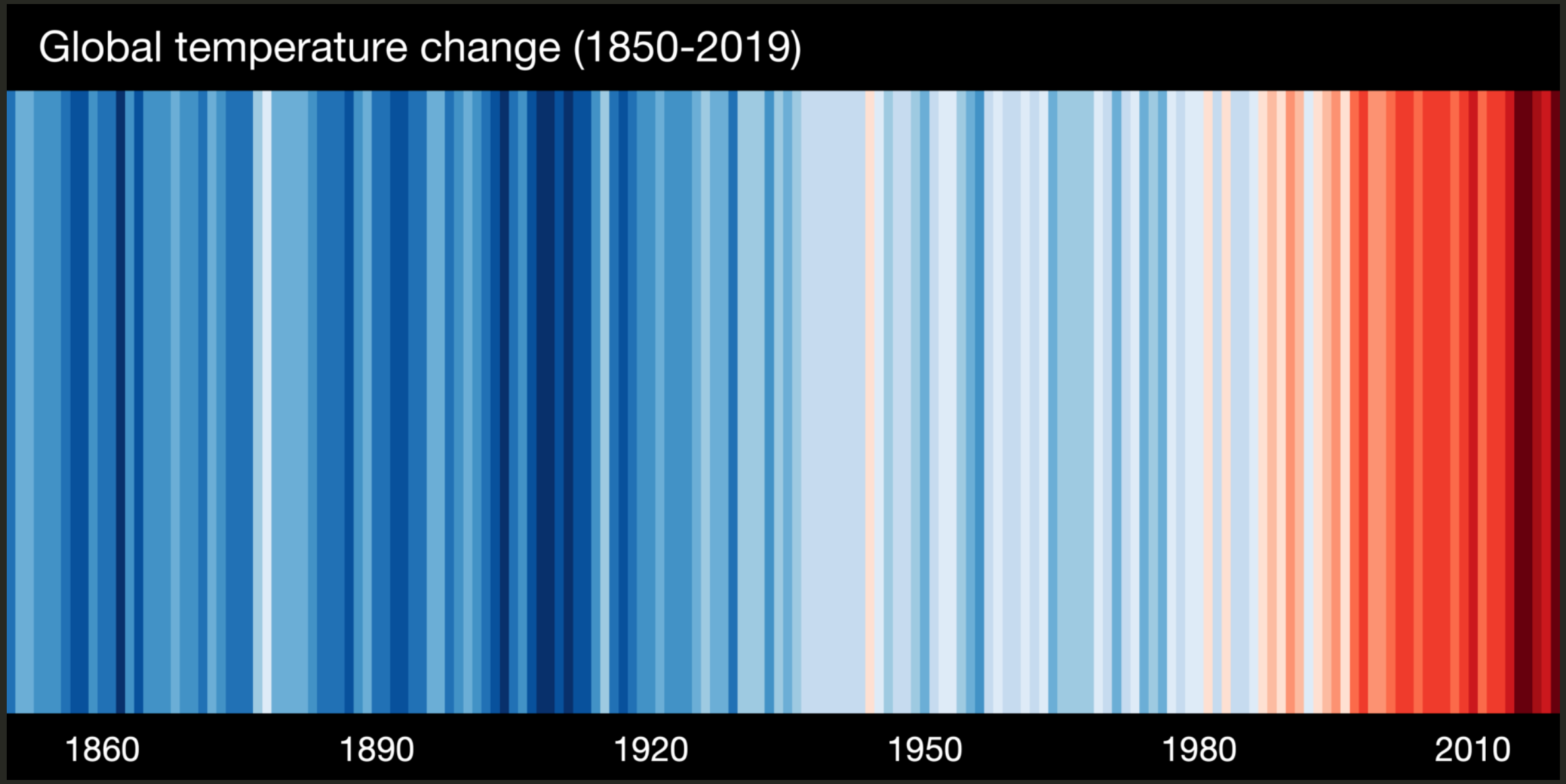
Which variables are **meaningful** to my audience?

How will they **encounter** the visualization?

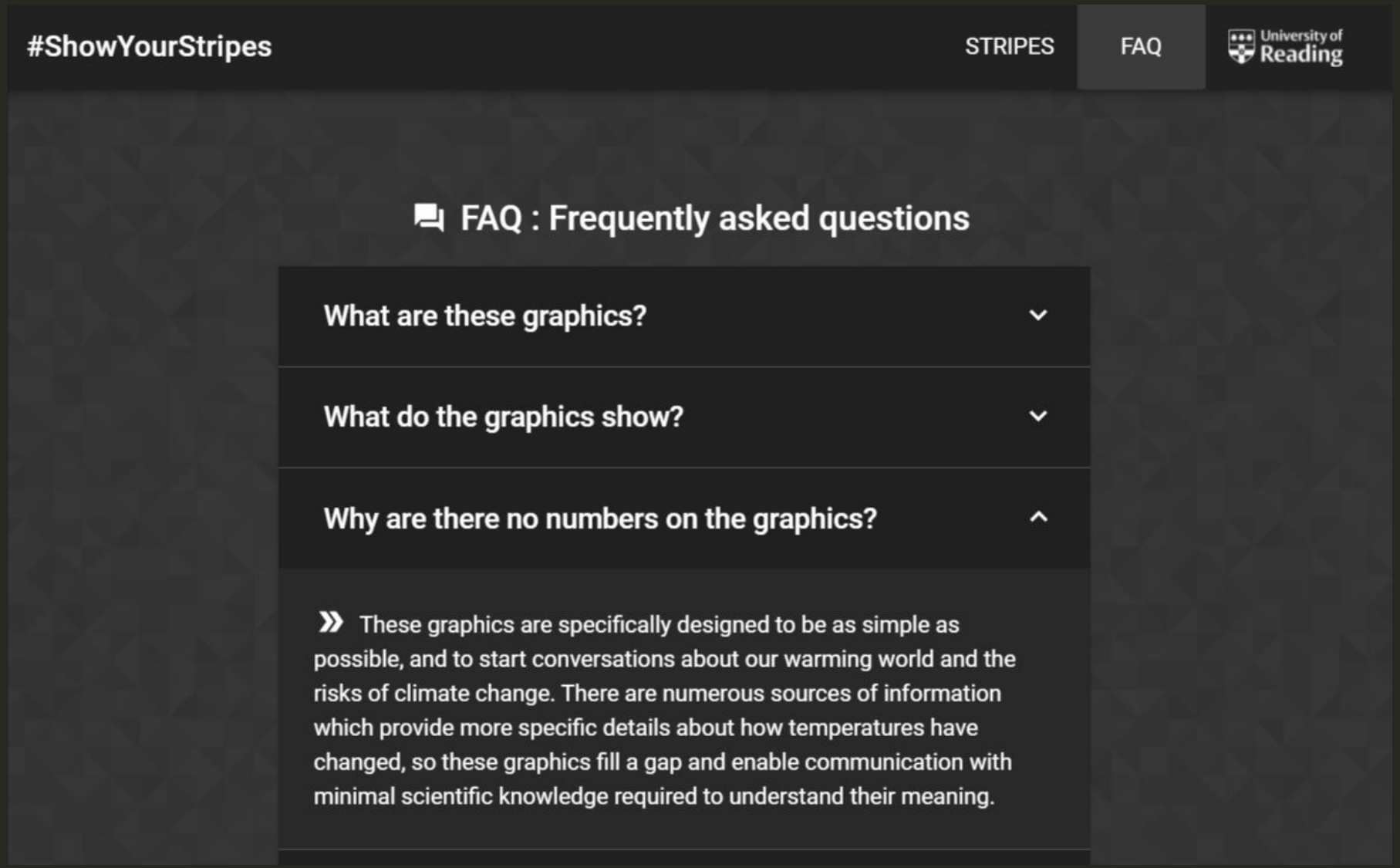
Do I need a visualization at all??



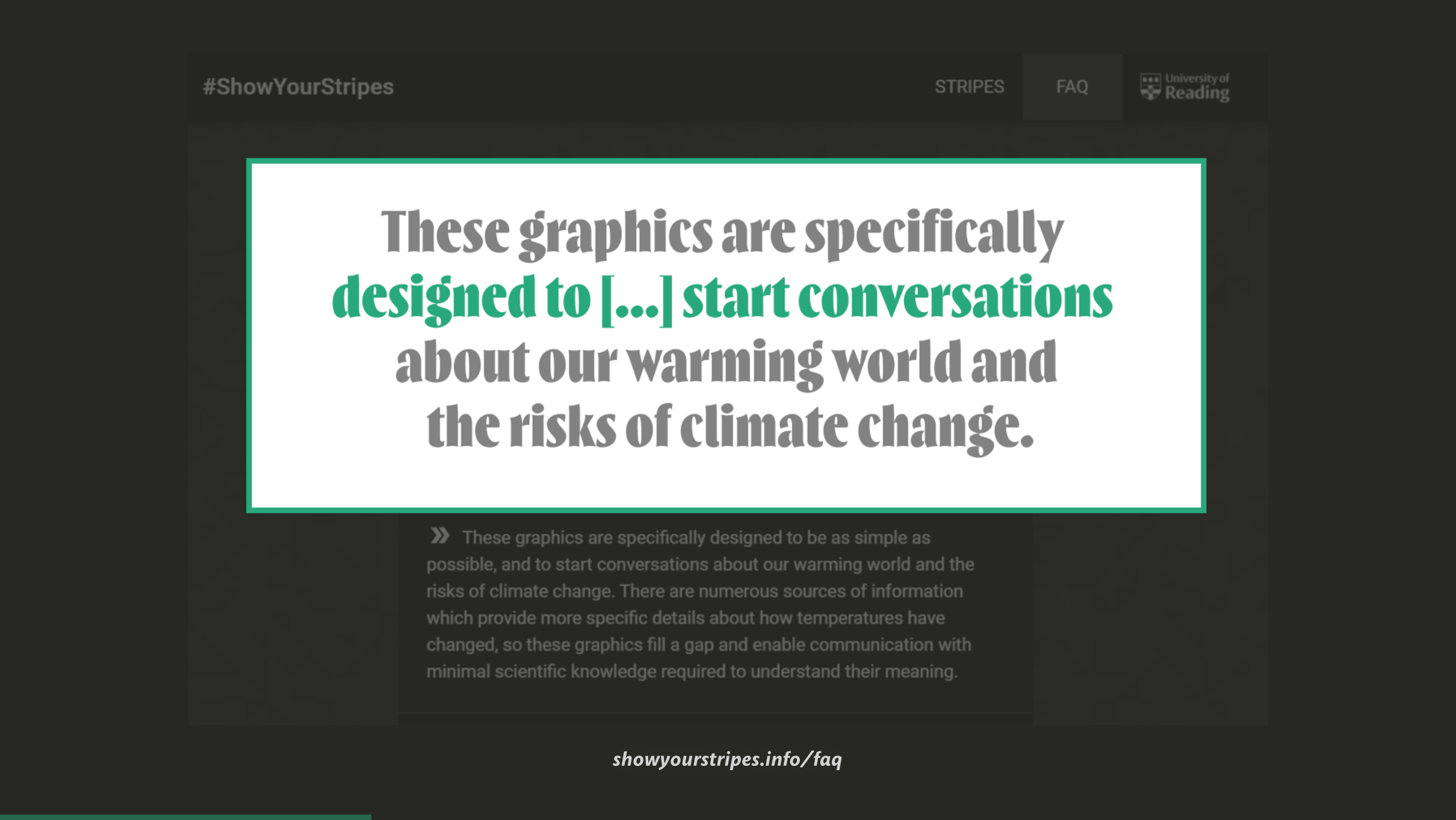
Warming Stripes by Ed Hawkins



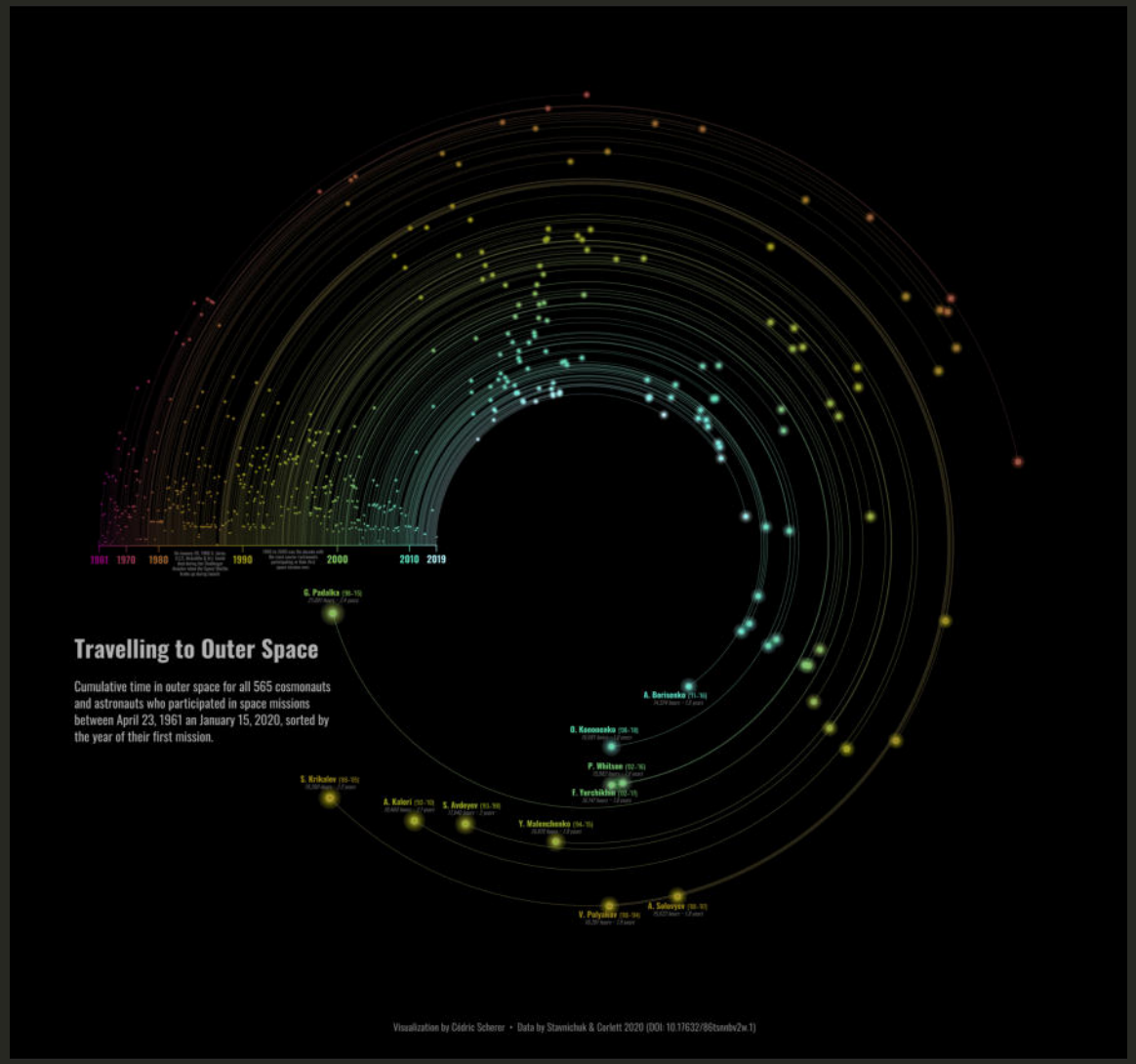
Warming Stripes by Ed Hawkins



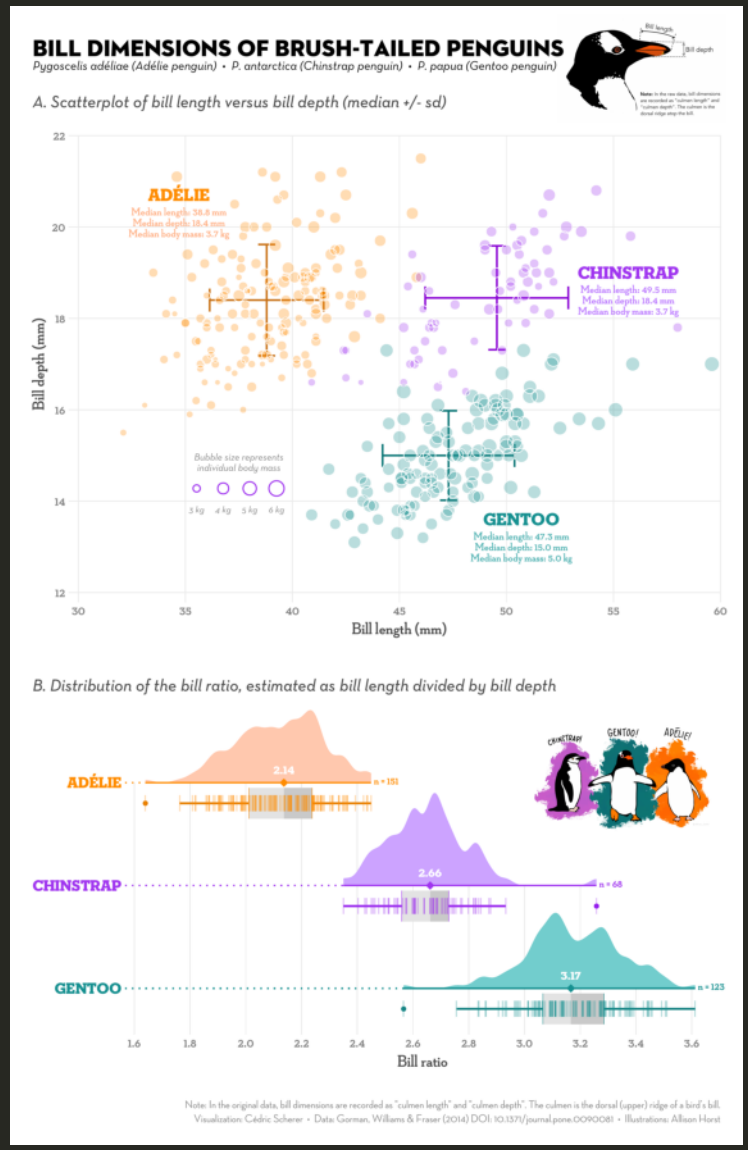
showyourstripes.info/faq



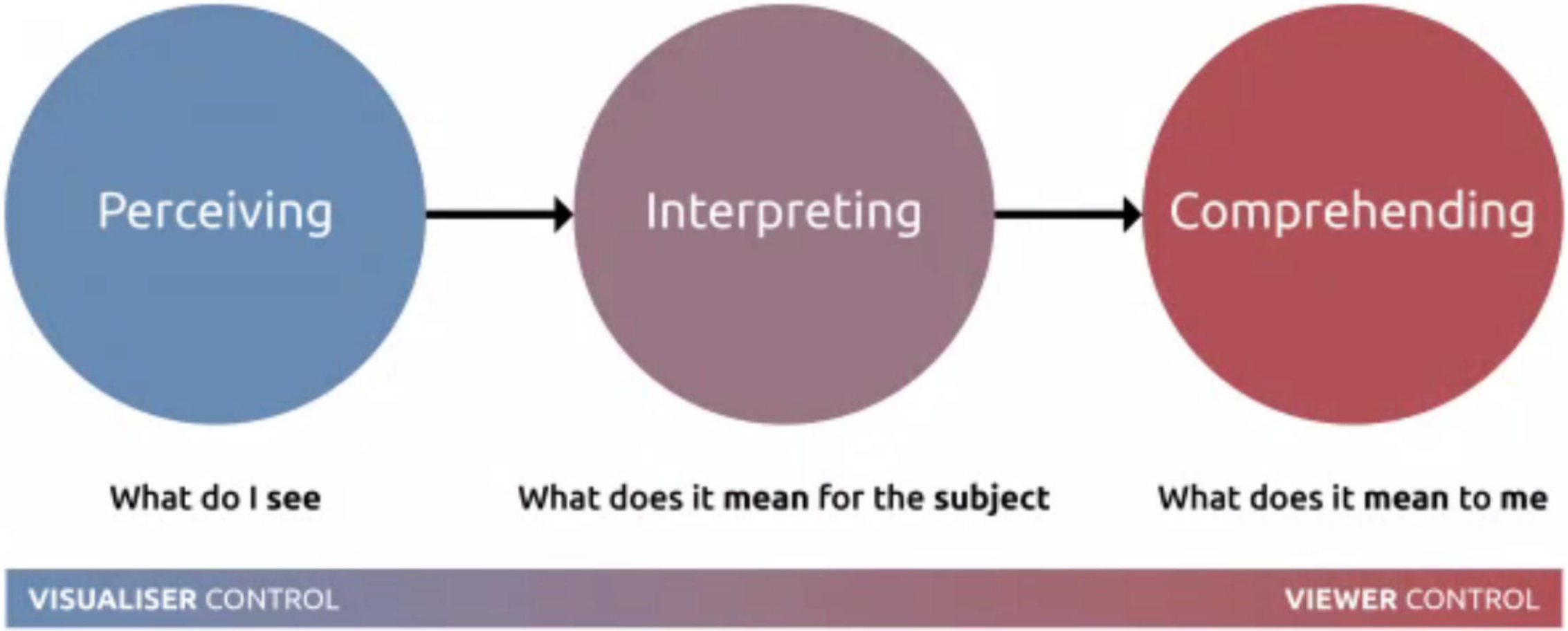
showyourstripes.info/faq



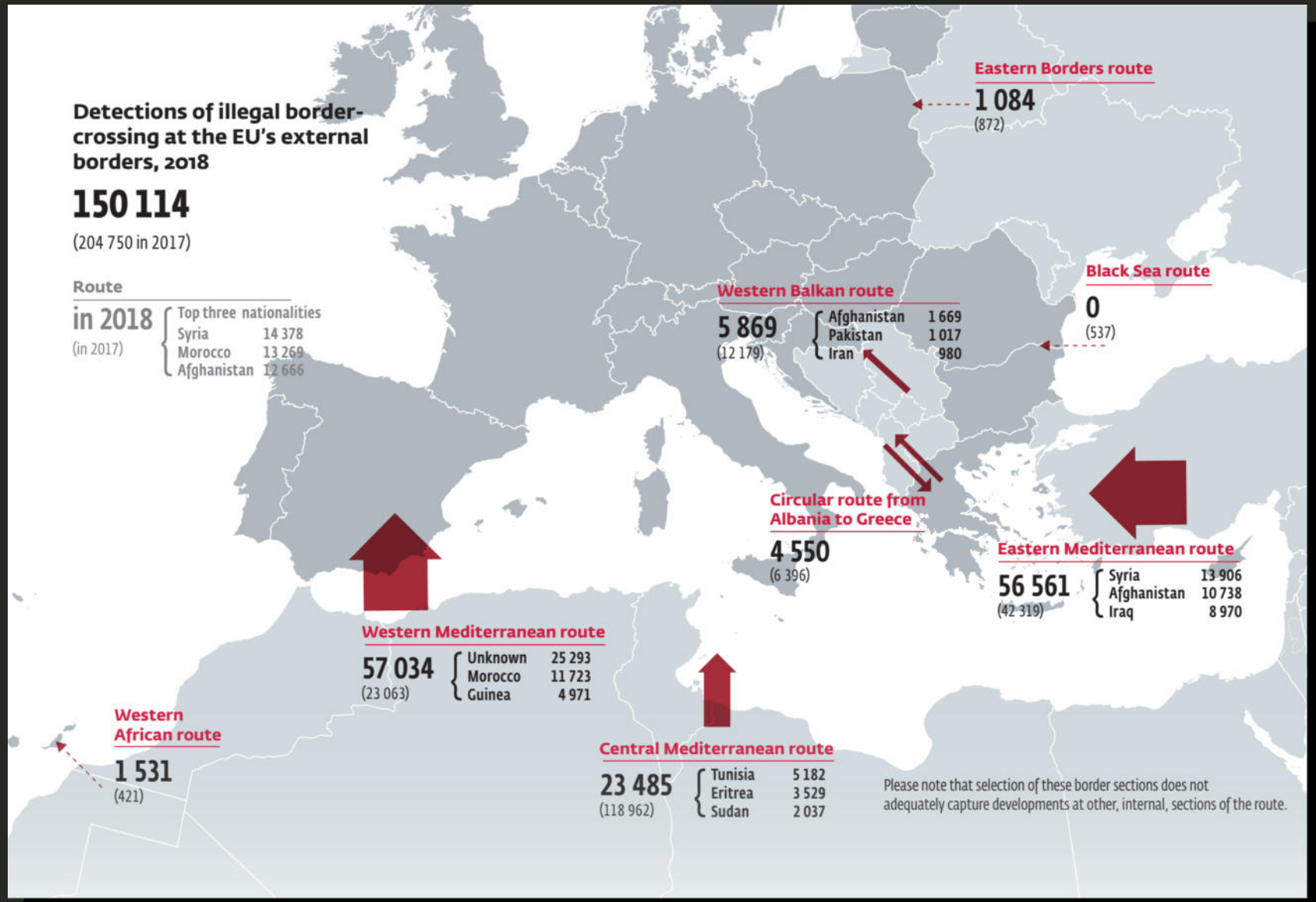
#TidyTuesday contribution Week 2020/29

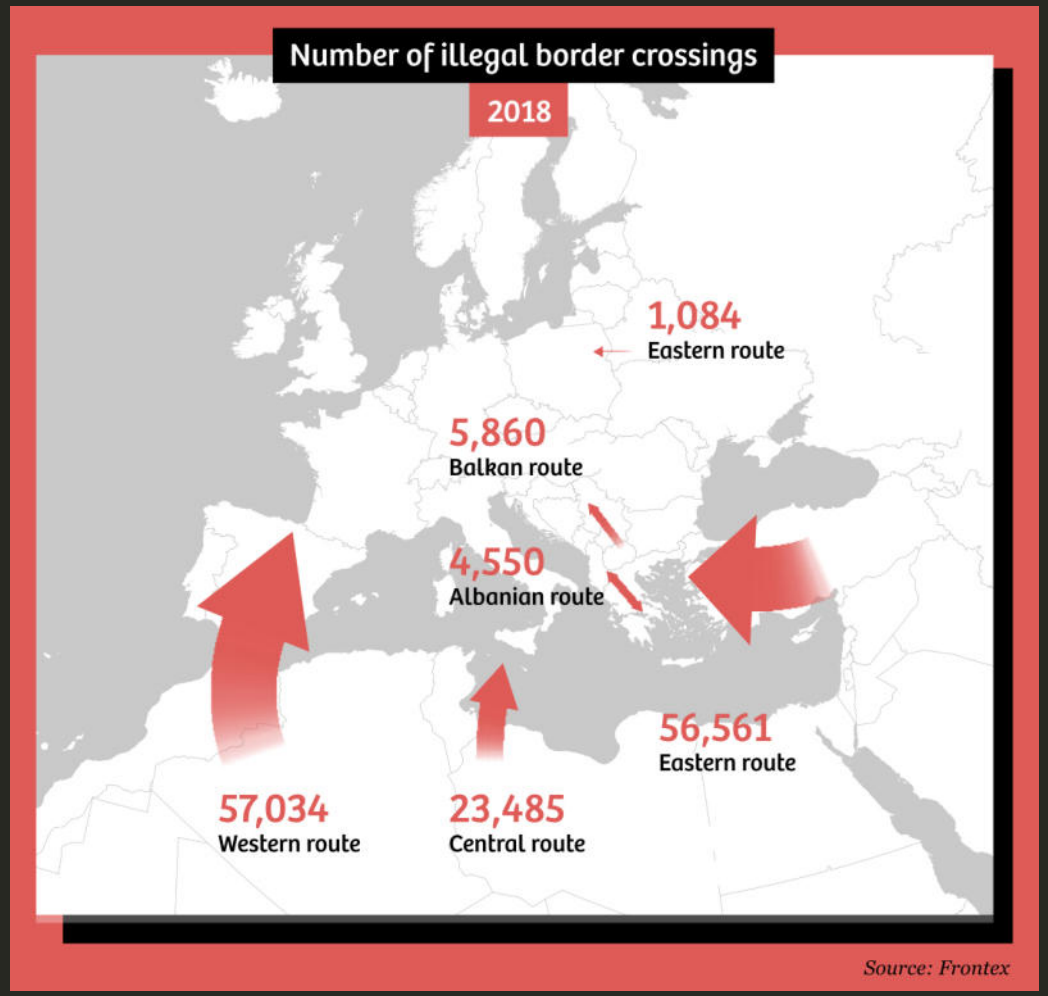


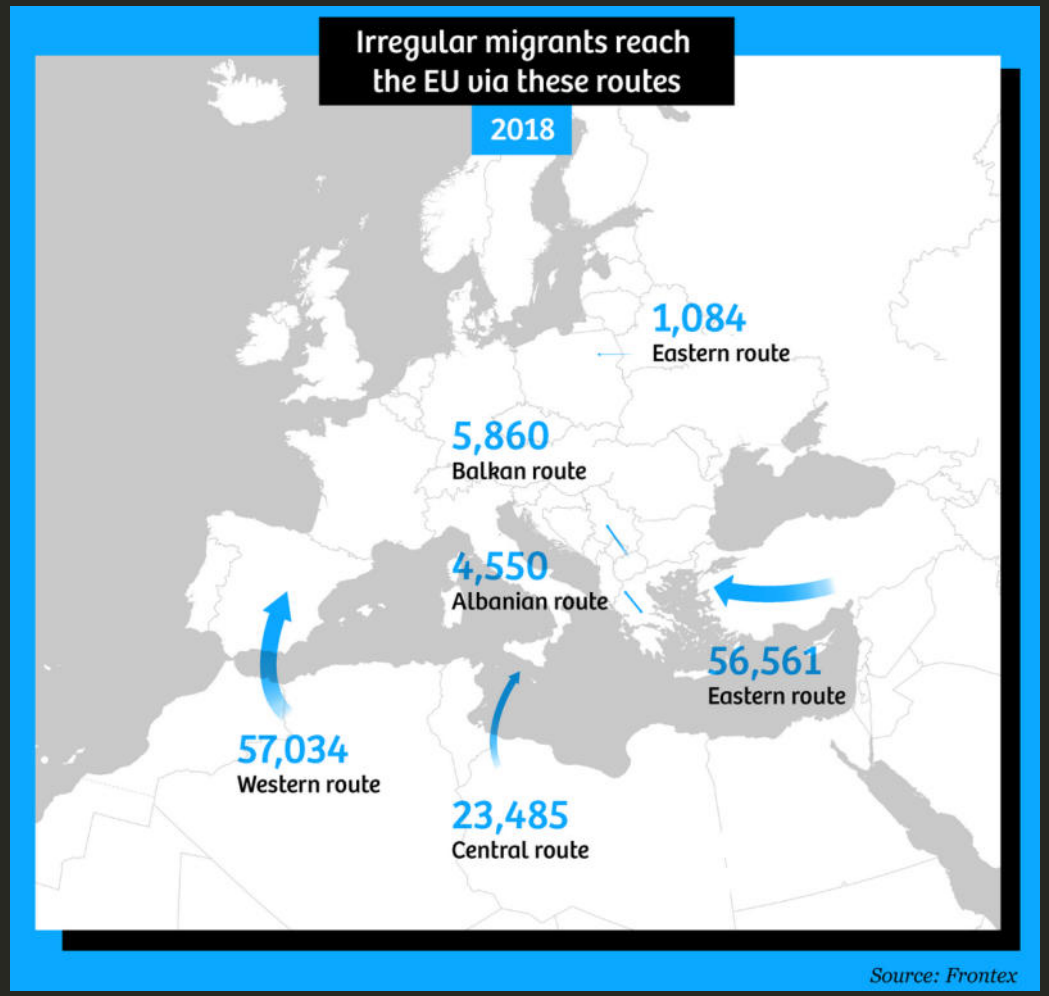
#TidyTuesday contribution Week 2020/31



Source: Andy Kirk, S-H-O-W Feb 2021

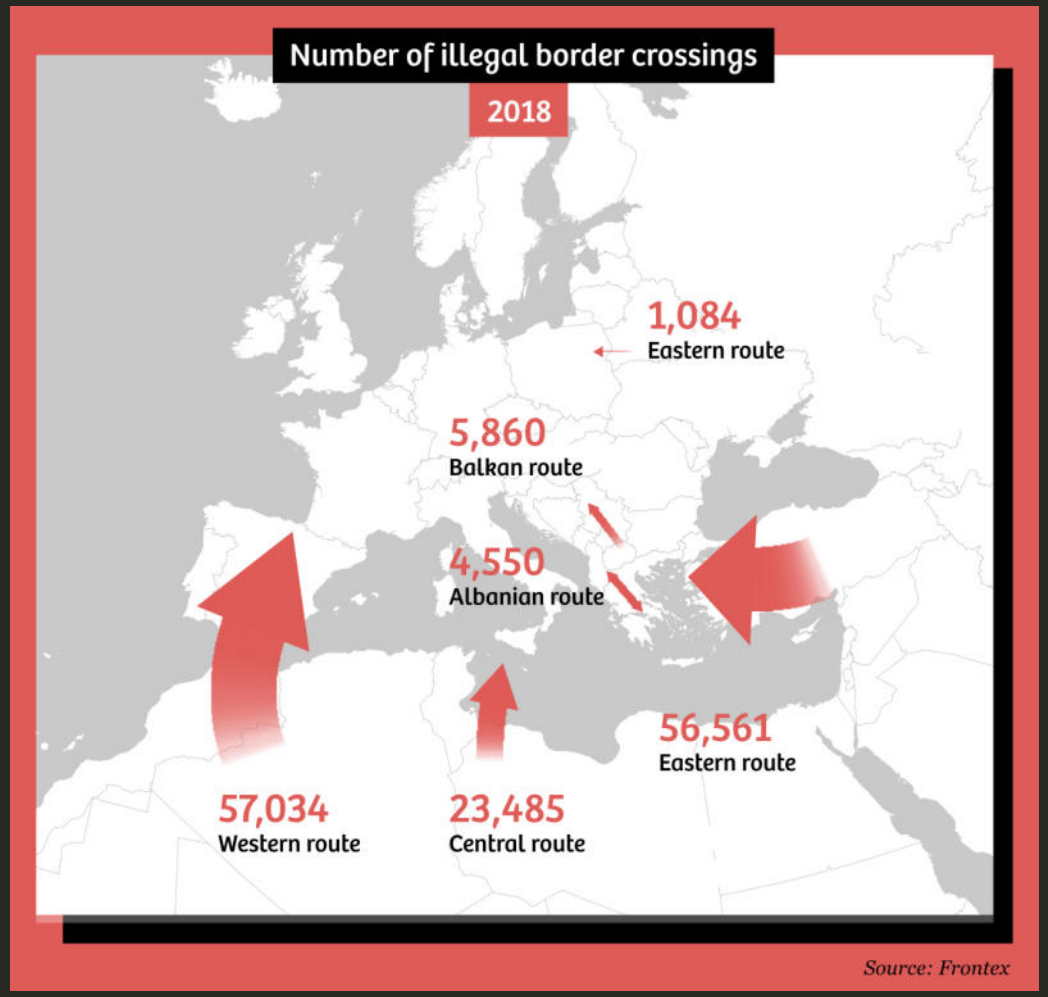

Detections of illegal border-crossing at the EU's external borders, 2018
150 114
(204 750 in 2017)
Route
in 2018
(in 2017)
Top three nationalities
Syria 14 378
Morocco 13 269
Afghanistan 12 666
Eastern Borders route
1 084
(872)
Western Balkan route
5 869
(12 179)
Afghanistan 1 669
Pakistan 1 017
Iran 980
Black Sea route
0
(537)
Circular route from Albania to Greece
4 550
(6 396)
Eastern Mediterranean route
56 561
(42 319)
Syria 13 906
Afghanistan 10 738
Iraq 8 970
Western Mediterranean route
57 034
(23 063)
Unknown 25 293
Morocco 11 723
Guinea 4 971
Western African route
1 531
(421)
Central Mediterranean route
23 485
(118 962)
Tunisia 5 182
Eritrea 3 529
Sudan 2 037
Please note that selection of these border sections does not adequately capture developments at other, internal, sections of the route.

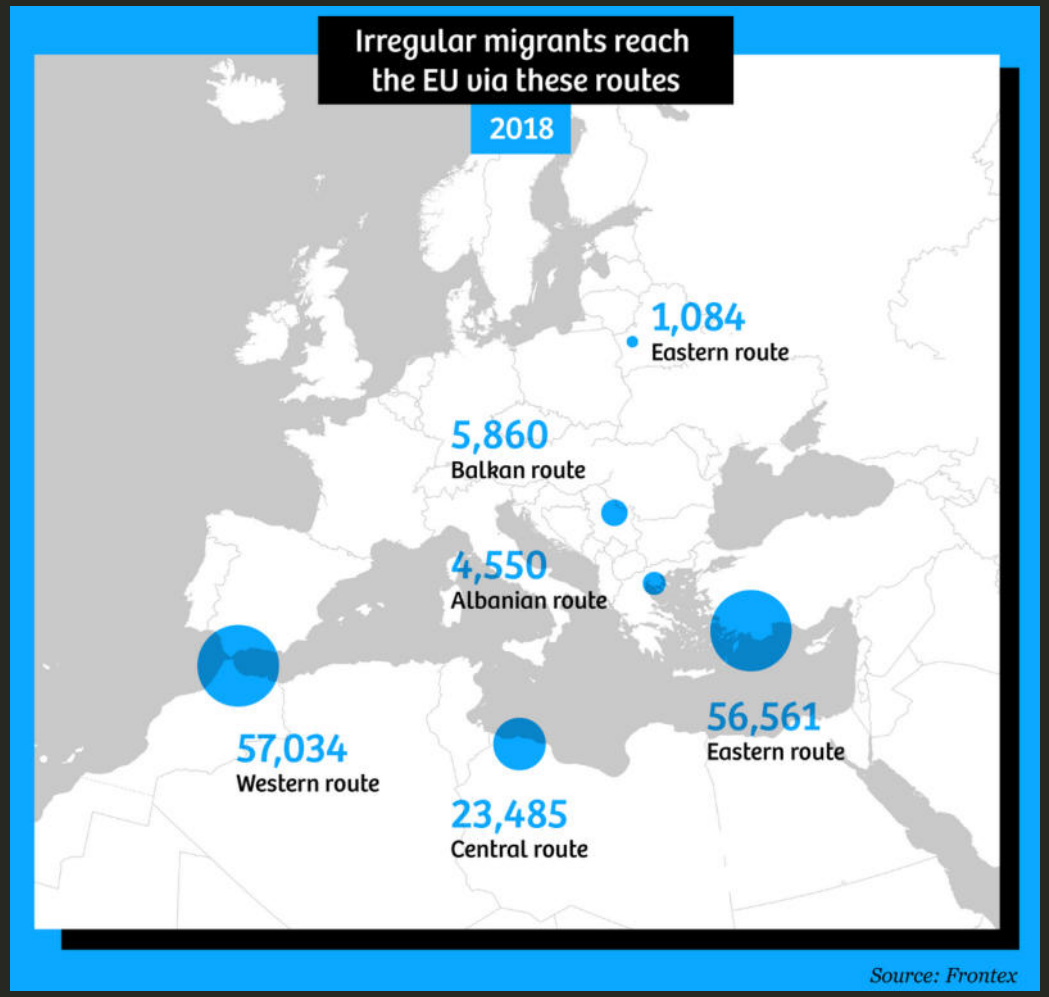




How maps in the media make us more negative about migrants

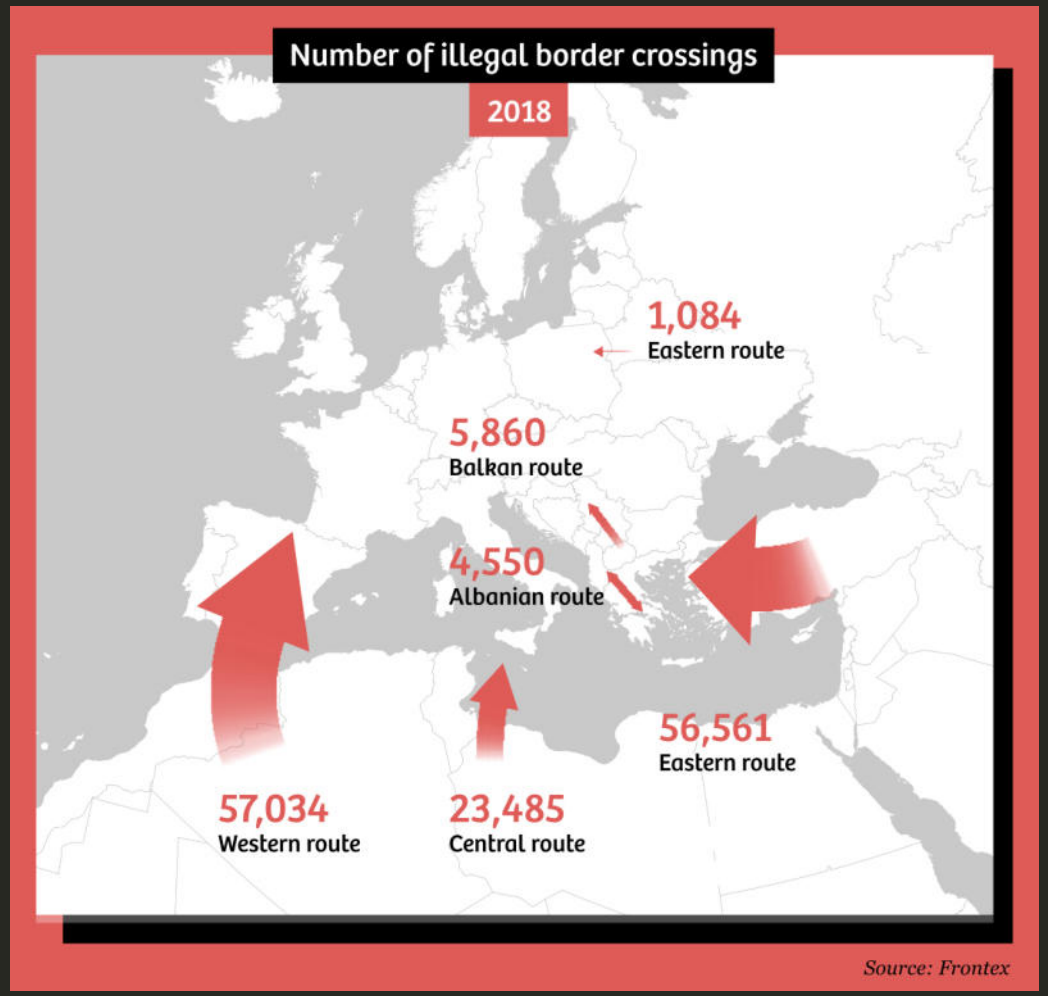
by Maite Vermeulen, Leon de Korte & Henk van Houtum

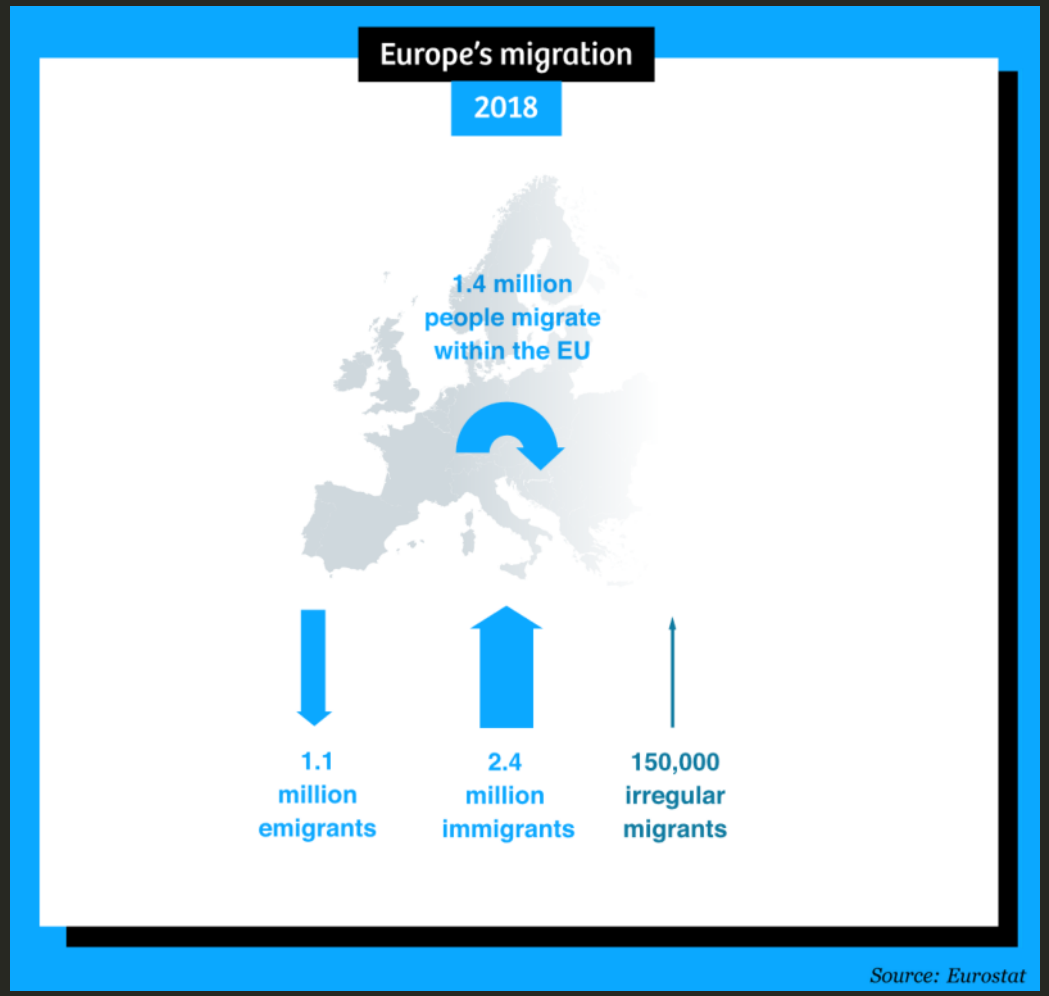




How maps in the media make us more negative about migrants

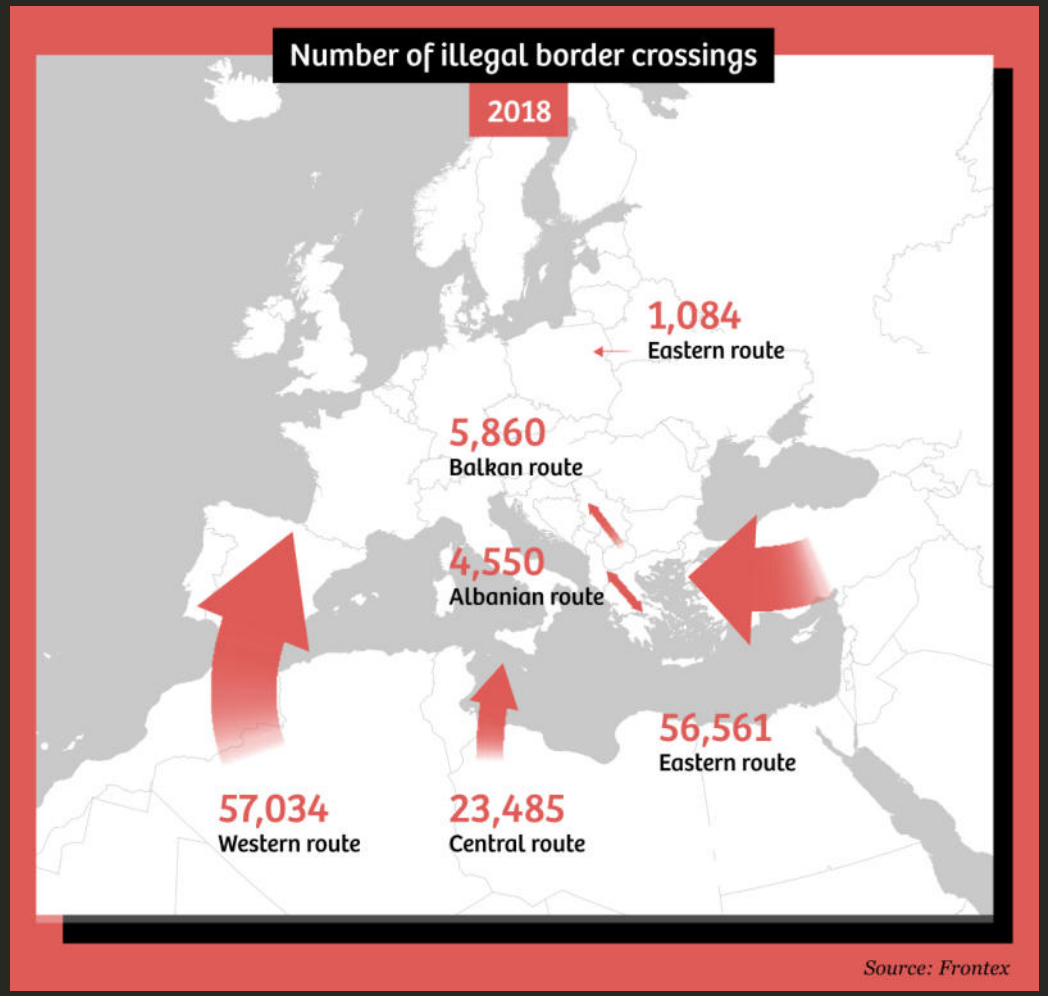
by Maite Vermeulen, Leon de Korte & Henk van Houtum

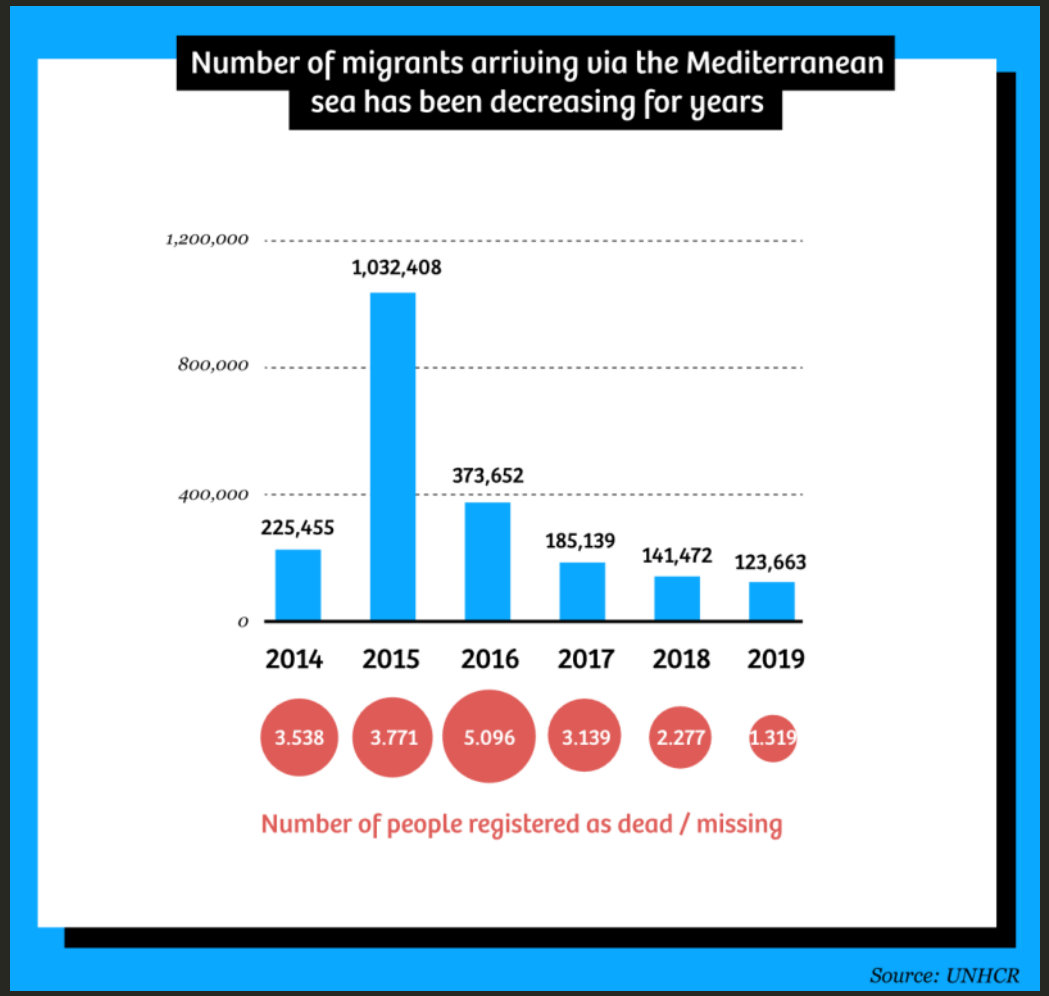




How maps in the media make us more negative about migrants

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

Goal

Select charts that successfully transport your story

Typology of Information Graphics

by Scott Berinato, "Good Charts" (2016), pp. 54–63

Is the information conceptual or data-driven?

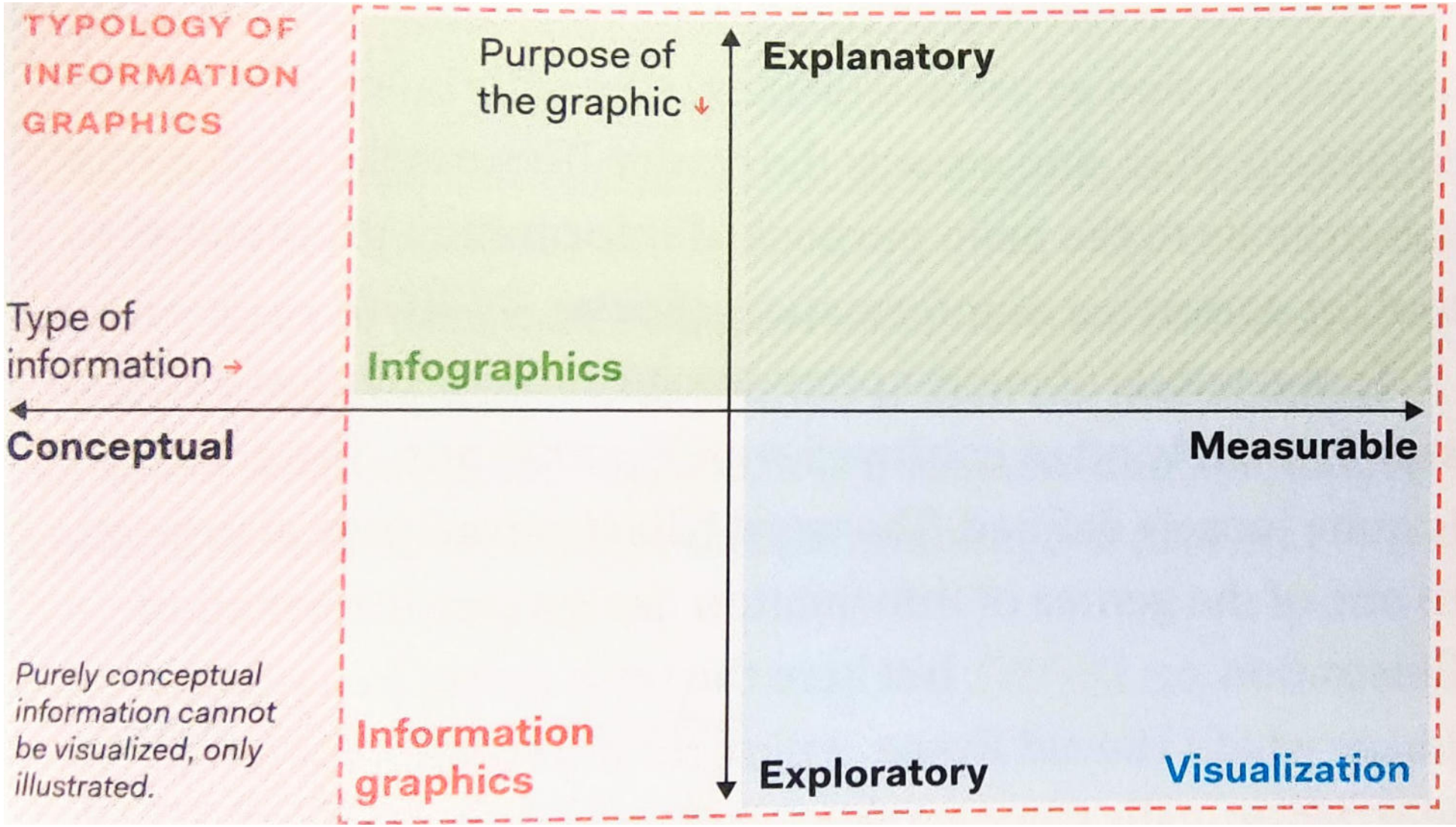
Is the purpose to declare or to explore the information?

Typology of Information Graphics

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

Is the information conceptual or measurable?

Is the purpose to explain or to explore the information?



Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 27 (Aalto University Press)

Typology of Information Graphics

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

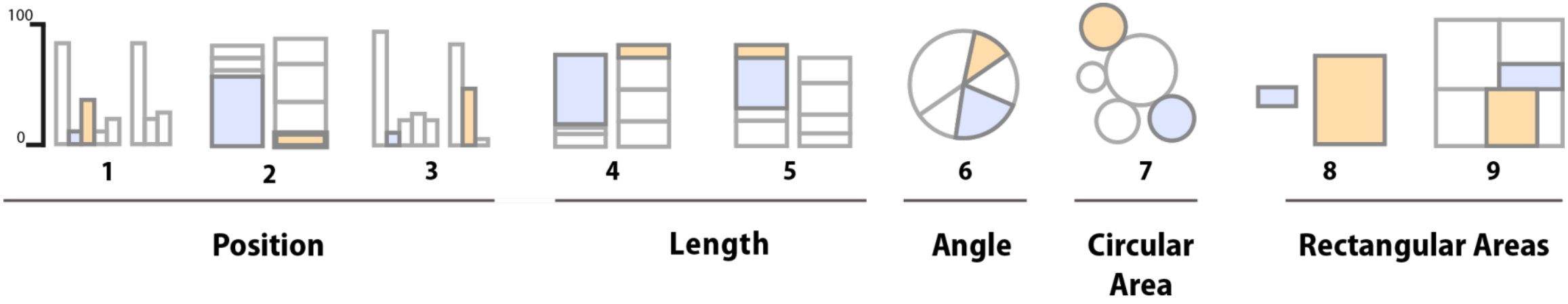
Is the information conceptual or measurable?

→ **Type of Information** — depict conceptual information *versus* convert information into visual forms

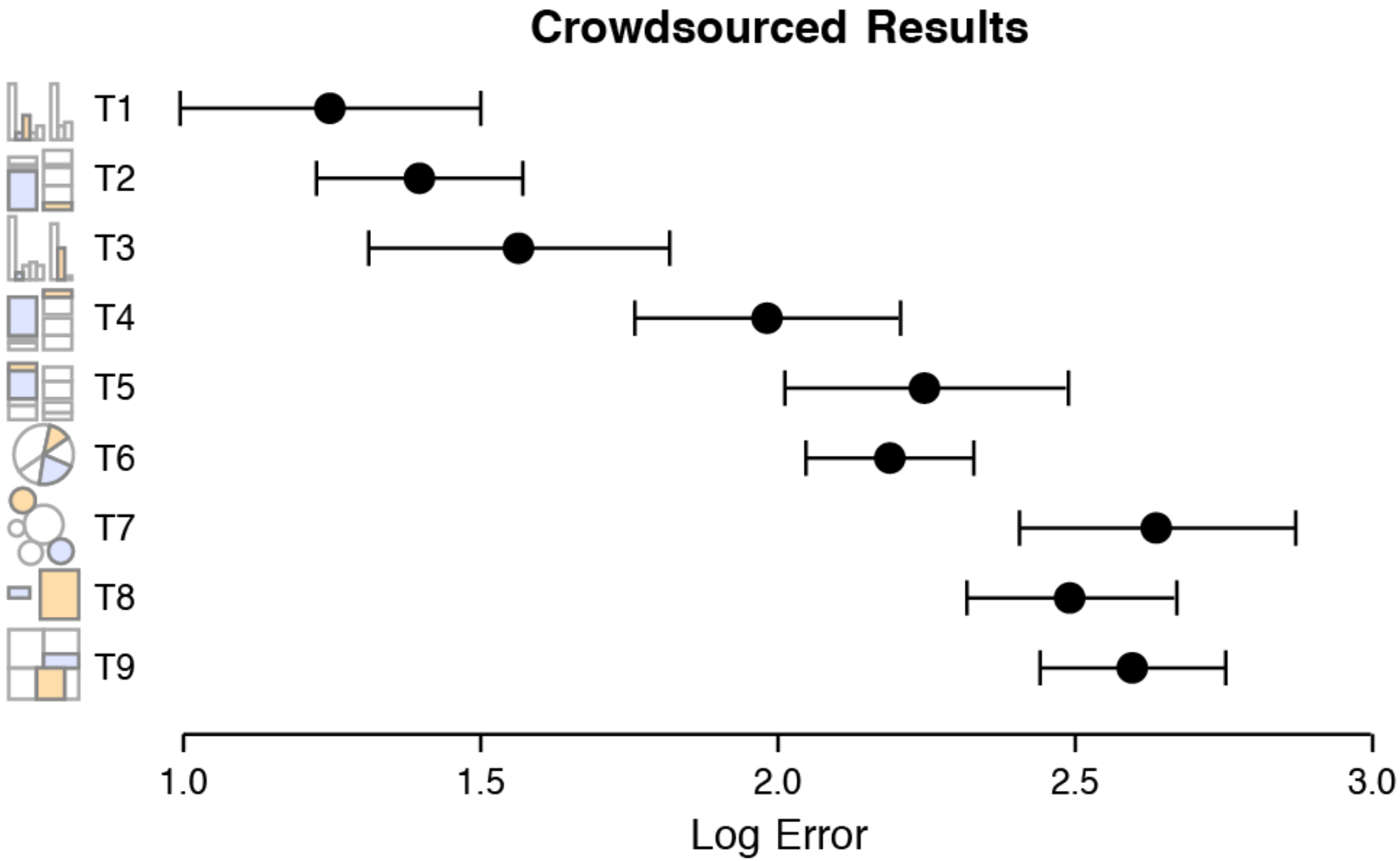
Is the purpose to explain or to explore the information?

→ **Purpose of the Graphic** — communicate information *versus* facilitate discovery

Data visualizations convert information into visual forms as quantifiable features

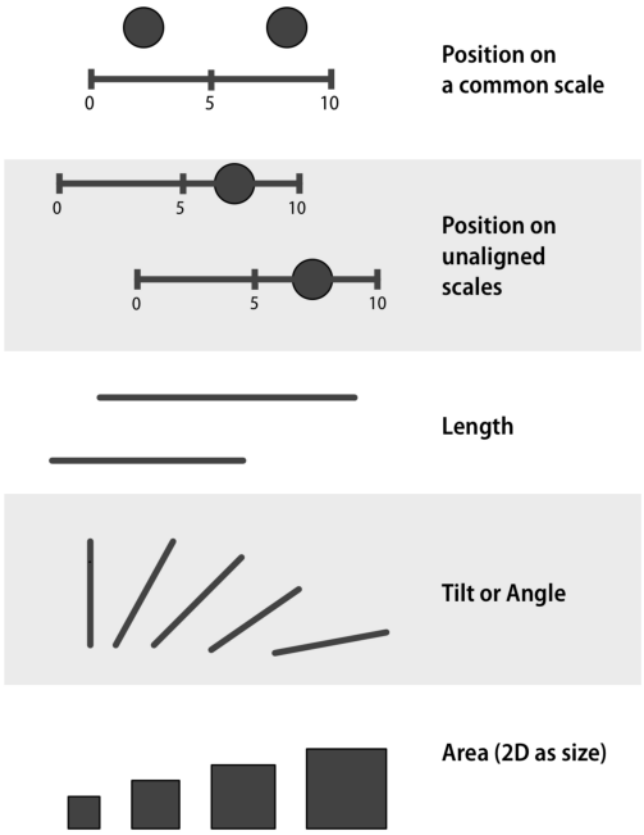


Kieran Healy based on Heer and Bostock, following Cleveland and McGill

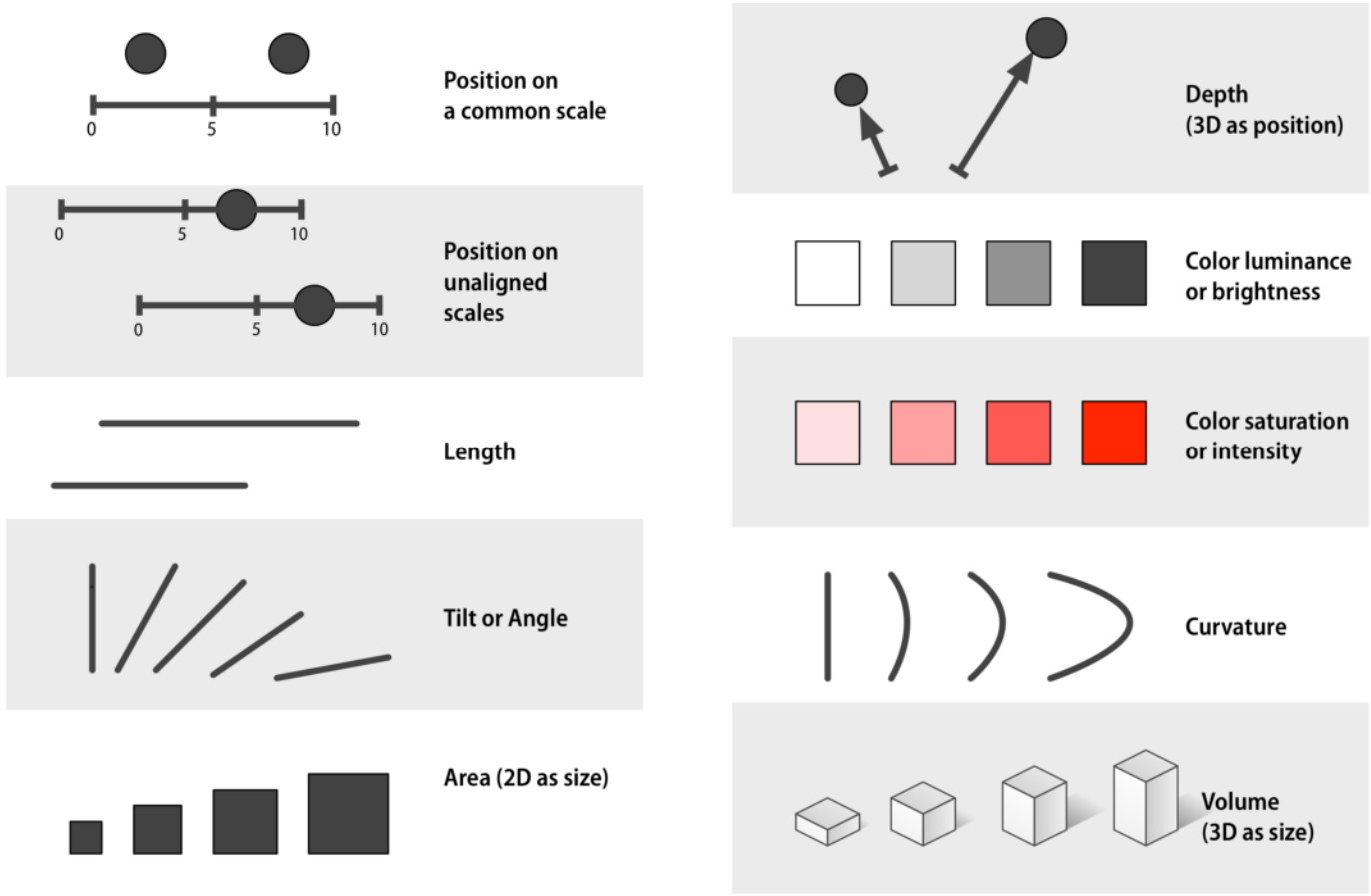


Kieran Healy based on Heer and Bostock, following Cleveland and McGill

Data visualizations convert information into visual forms as quantifiable features

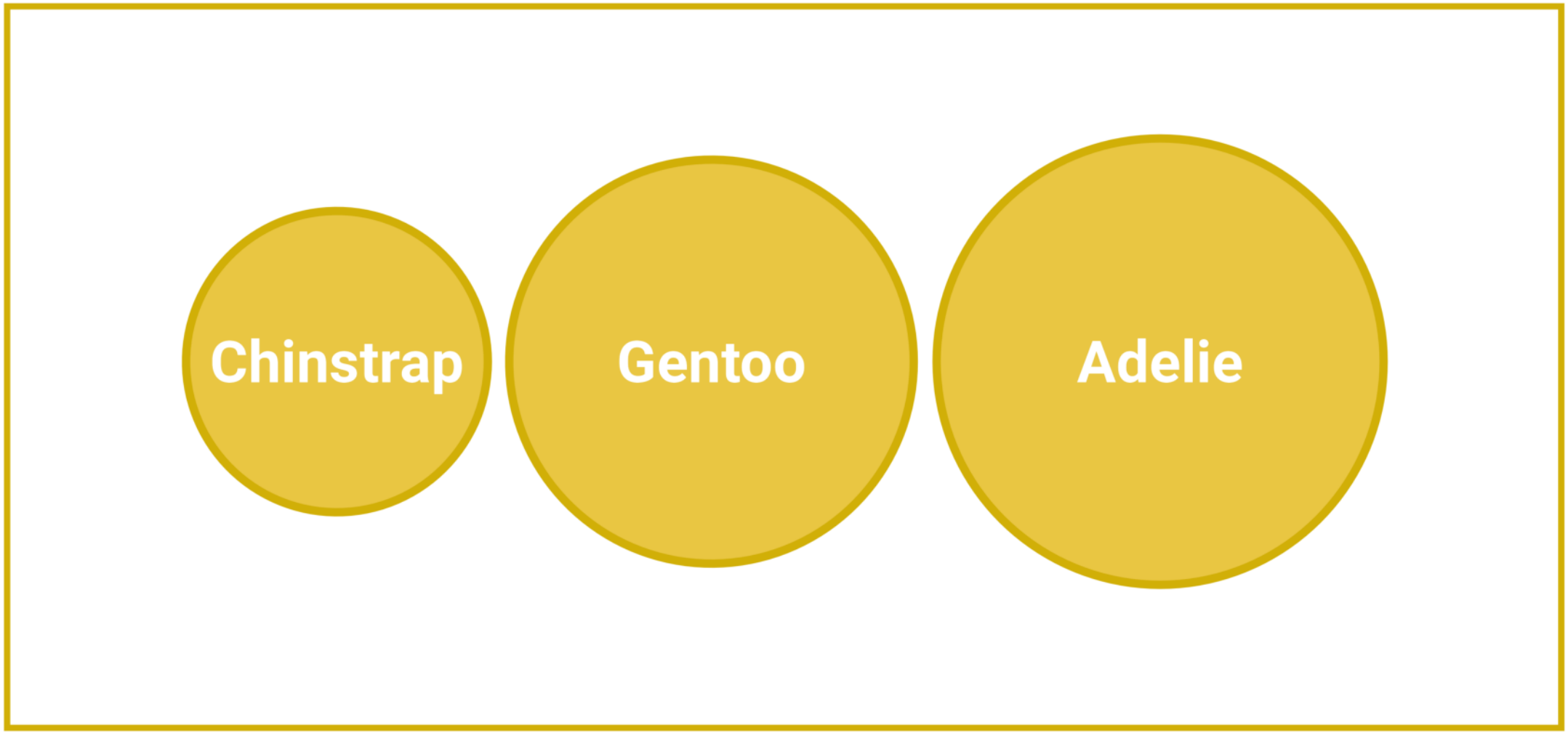


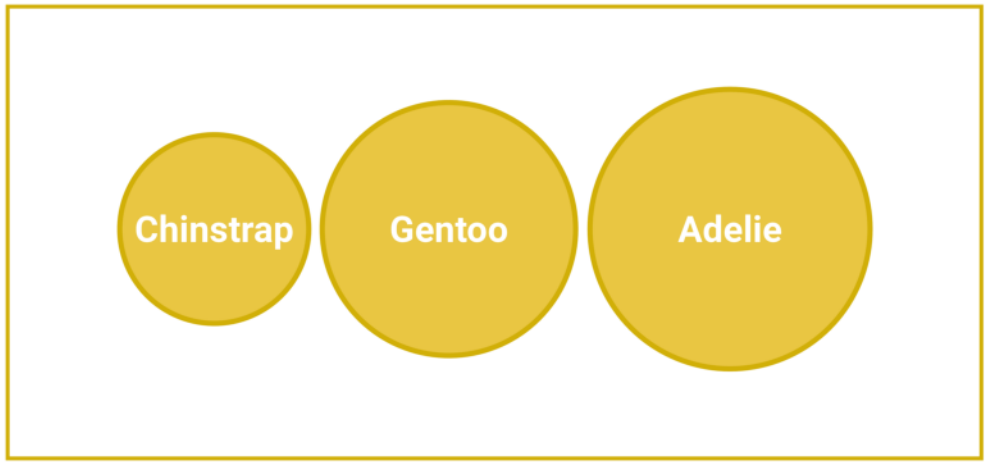
Data visualizations convert information into visual forms as quantifiable features

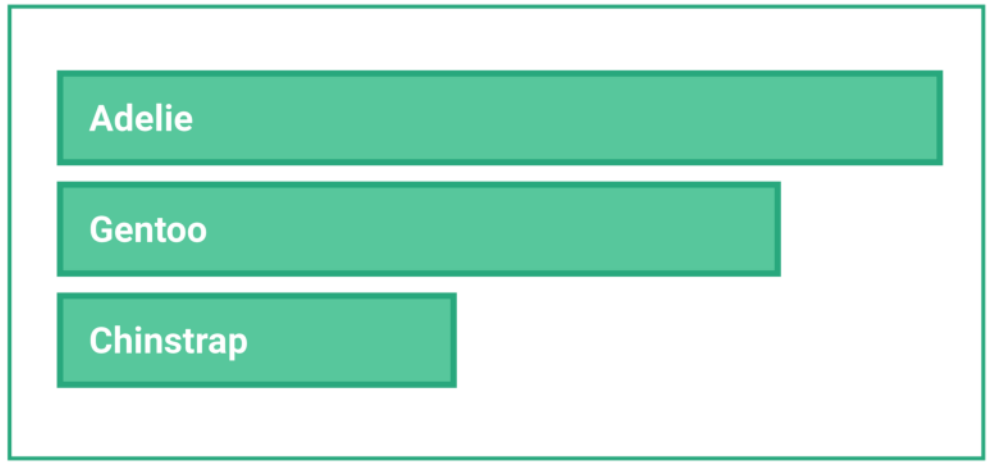


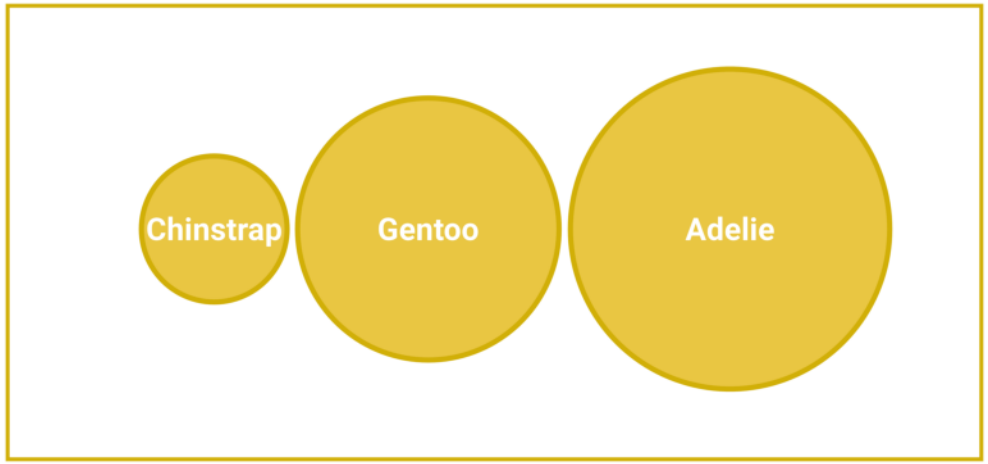
Qualitative **Nominal**	**Qualitative** **Ordinal**	**Quantitative** **Interval/Ratio**
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol

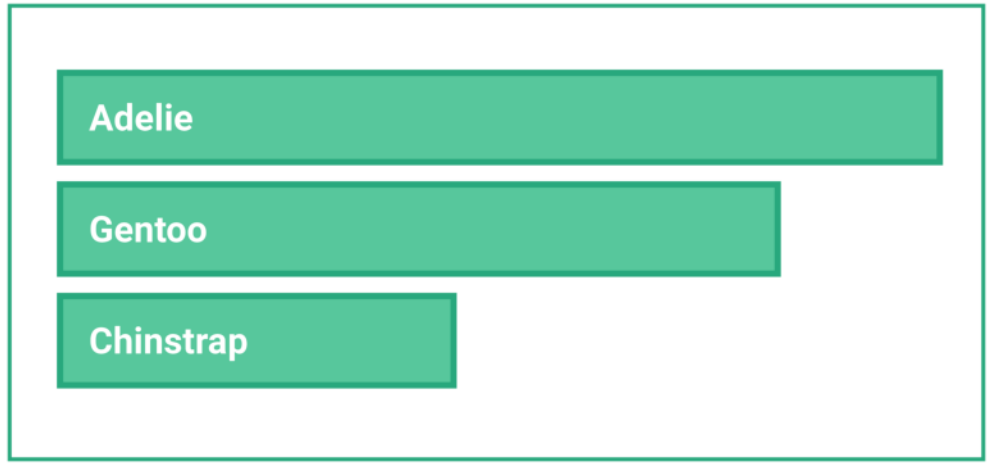
Qualitative Nominal	Qualitative Ordinal	Quantitative Interval/Ratio
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol


Chinstrap
Gentoo
Adelie

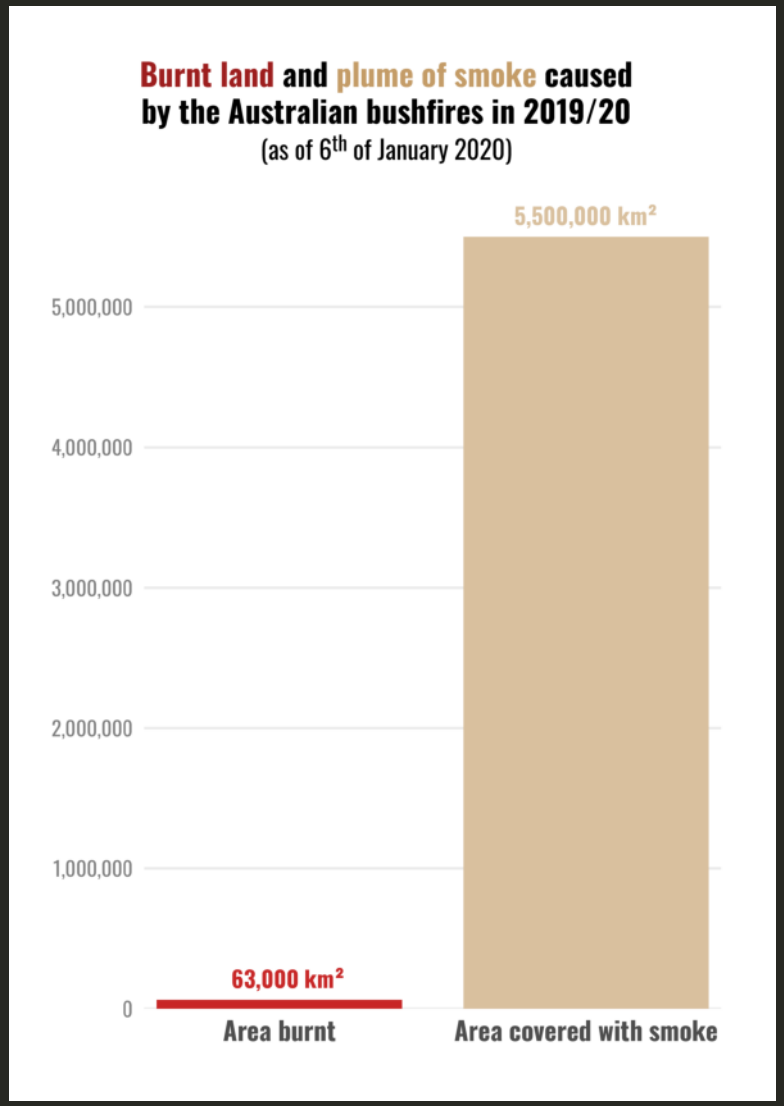




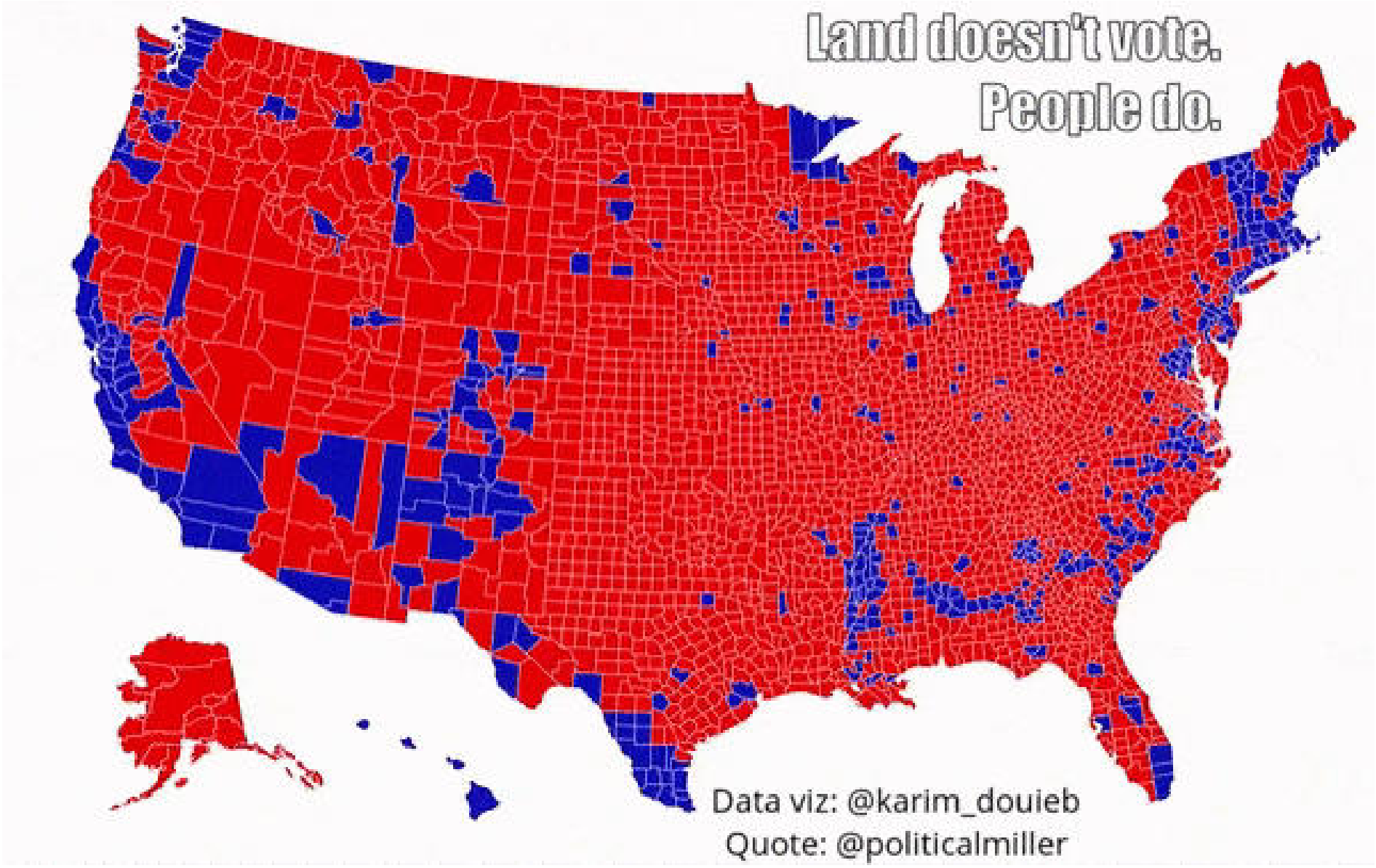




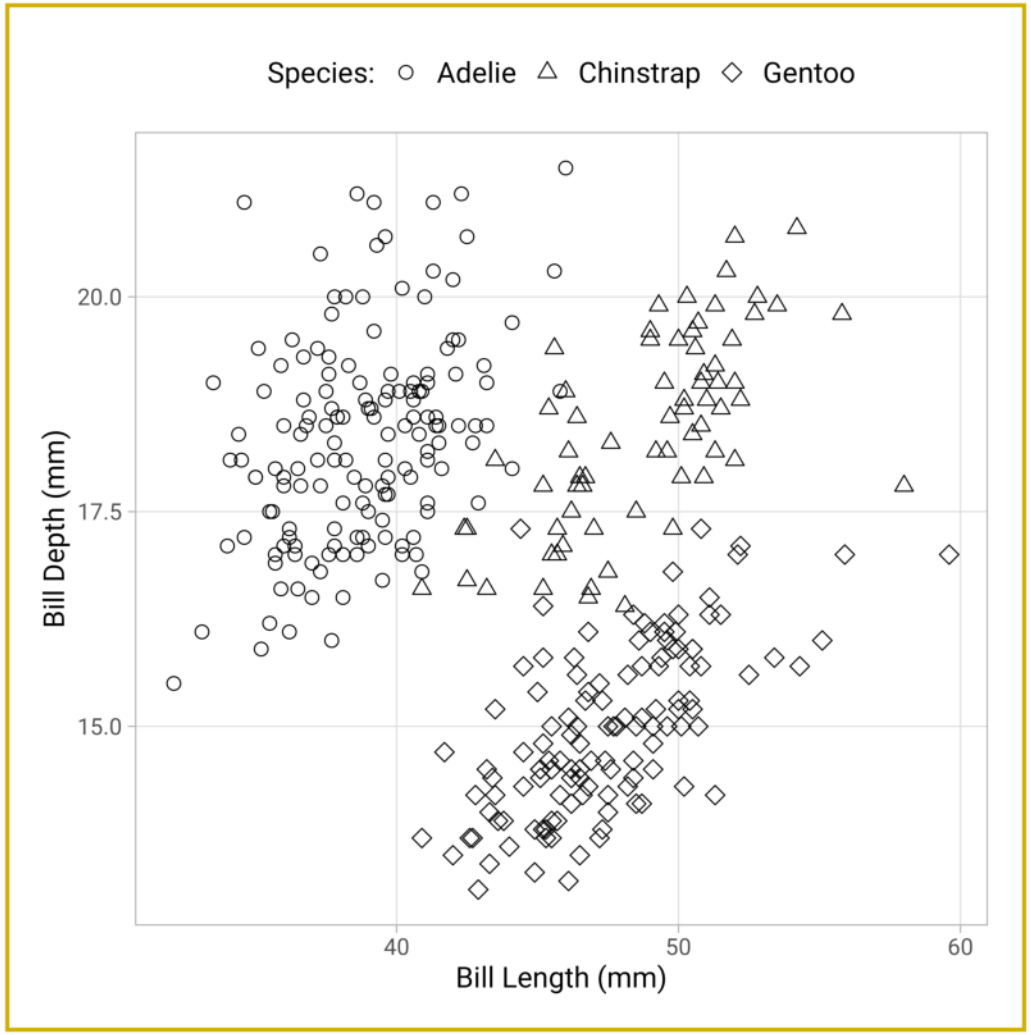
Always use area. Never use radius!

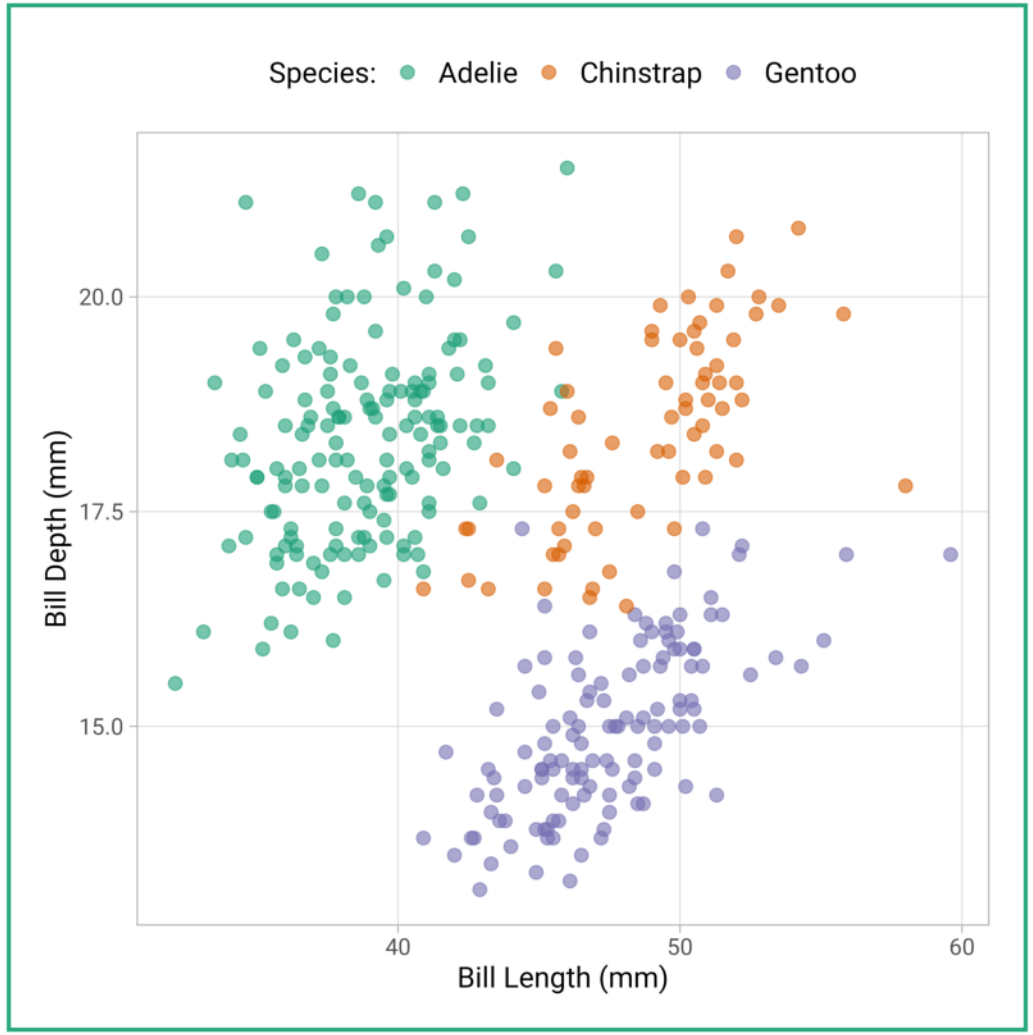


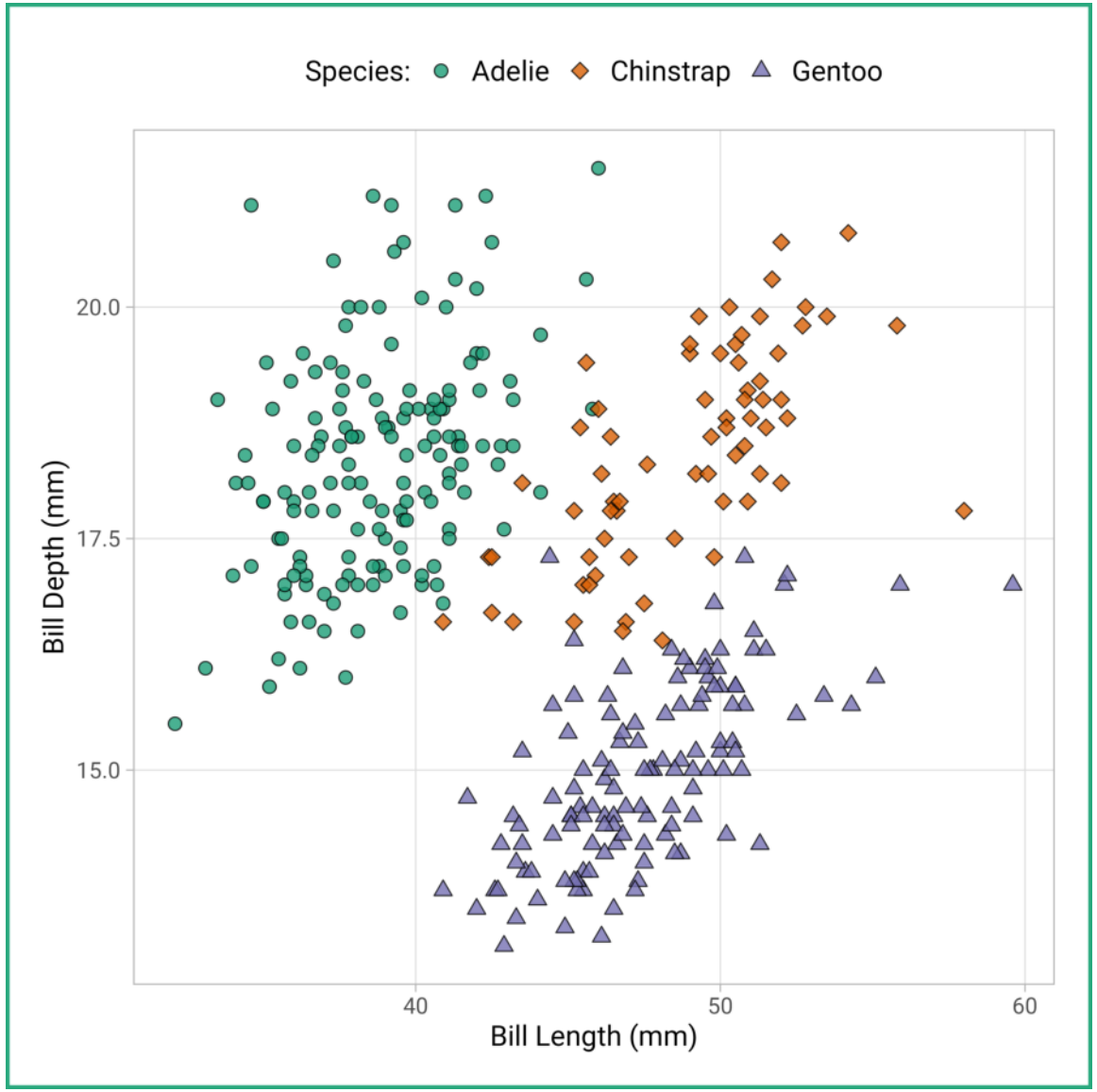
#TidyTuesday contribution Week 2020/02


Land doesn't vote.
People do.
Data viz: @karim_douieb
Quote: @politicalmiller

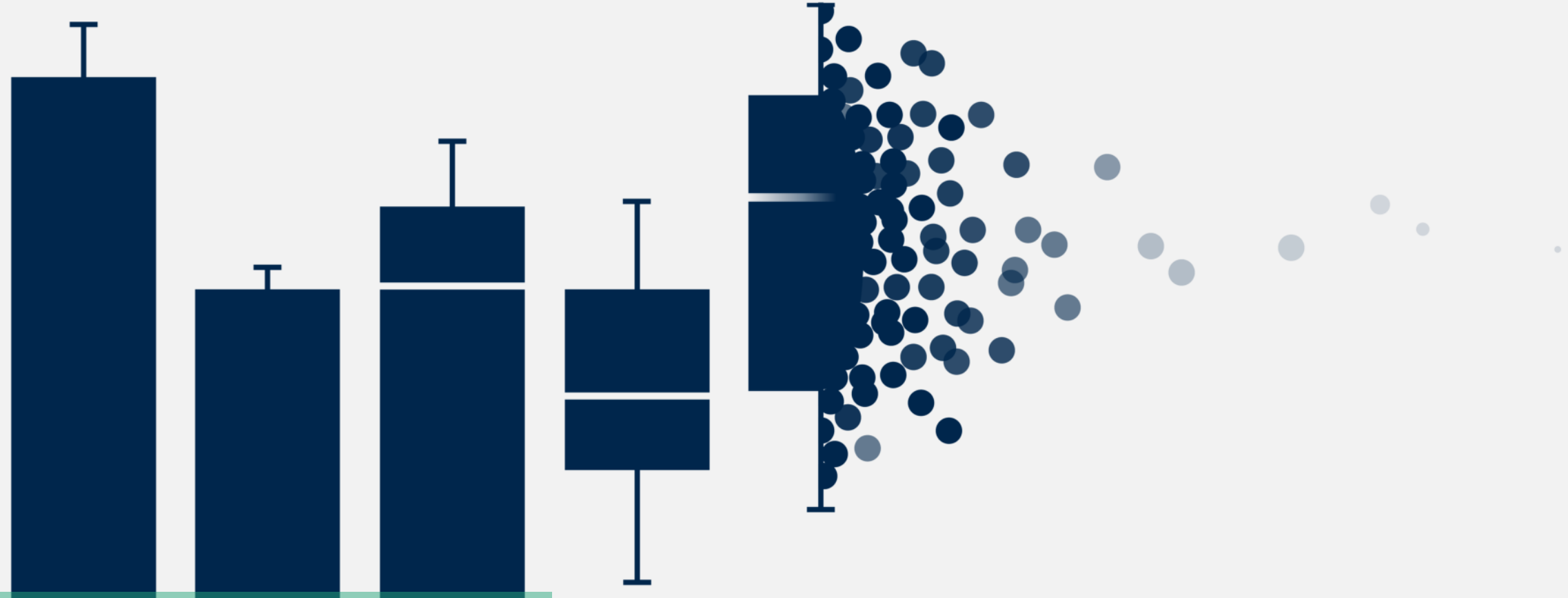
Qualitative **Nominal**	**Qualitative** **Ordinal**	**Quantitative** **Interval/Ratio**
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol

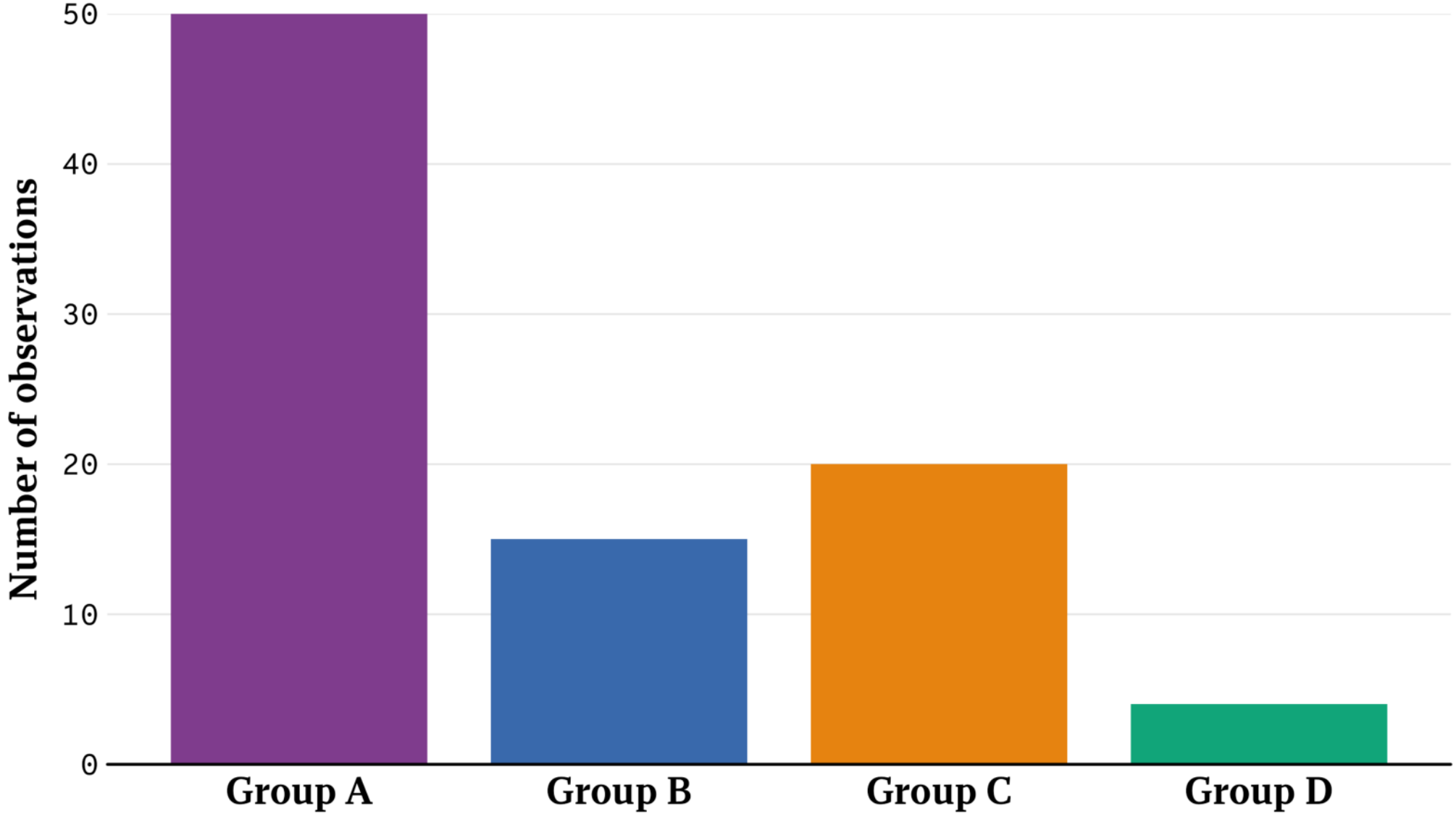

Species: Adelie Chinstrap Gentoo
20.0
17.5
15.0
Bill Depth (mm)
40
50
60
Bill Length (mm)

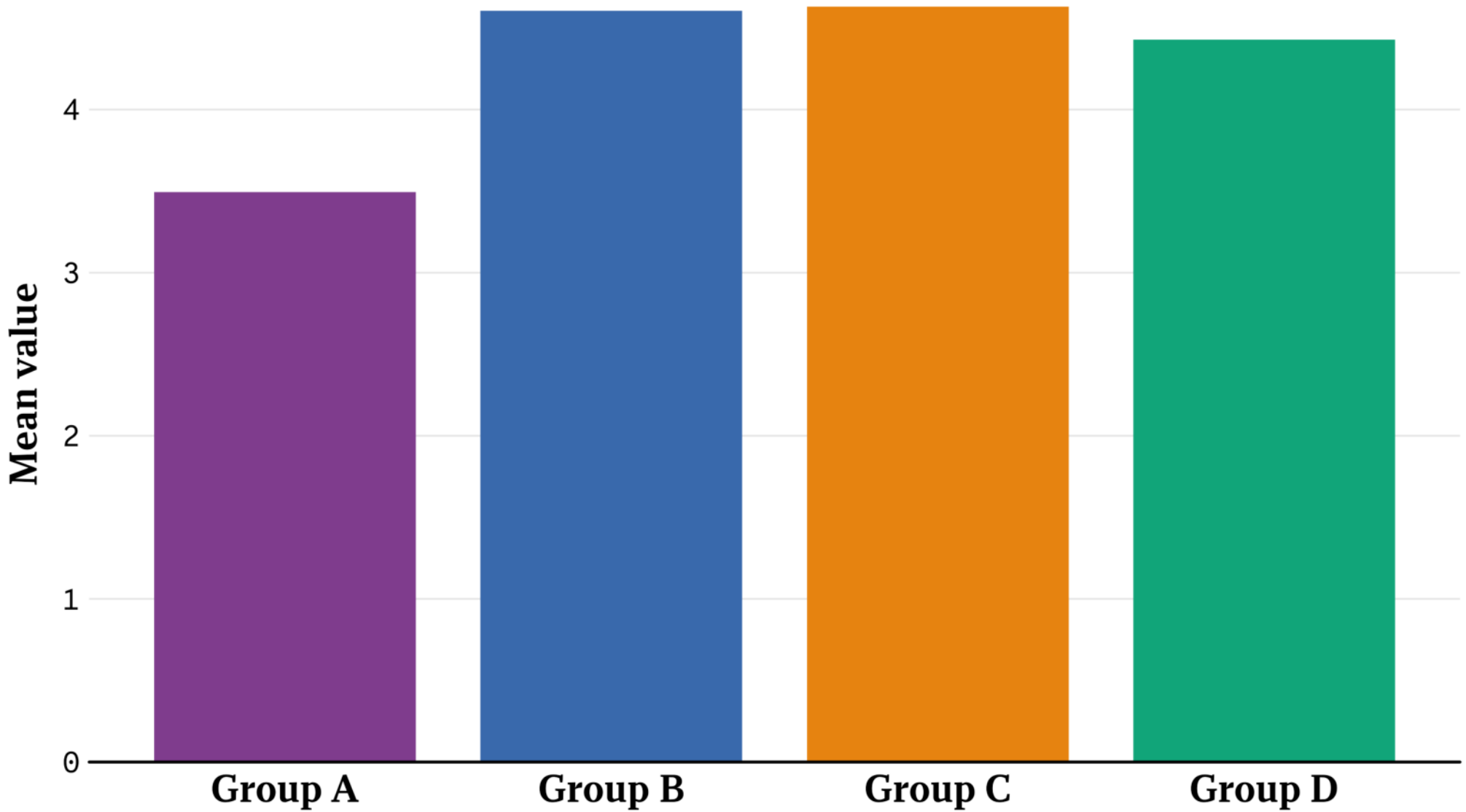

Species: Adelie Chinstrap Gentoo
20.0
17.5
15.0
Bill Depth (mm)
40
50
60
Bill Length (mm)

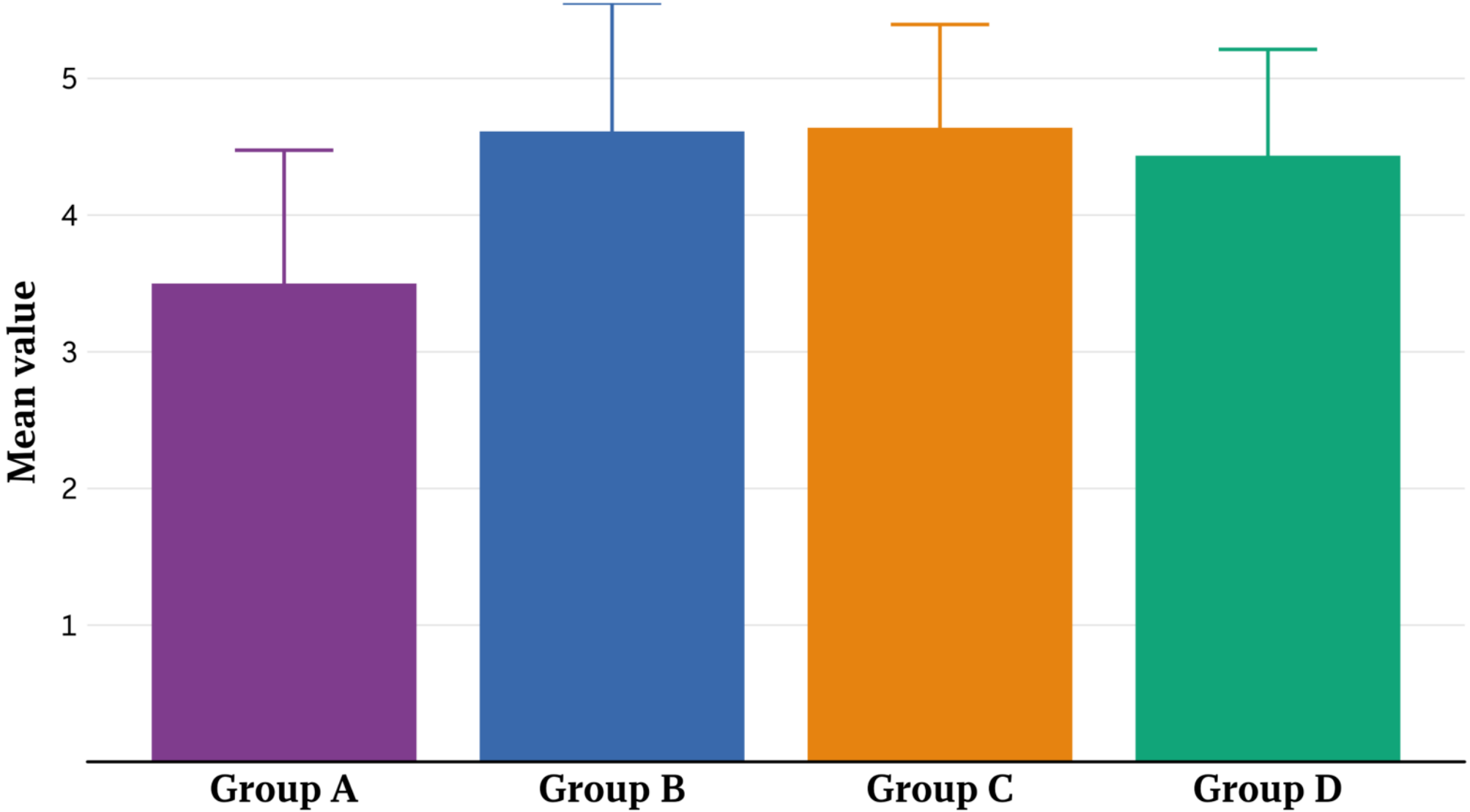

Species: Adelie Chinstrap Gentoo
Bill Depth (mm)
20.0
17.5
15.0
40
50
60
Bill Length (mm)

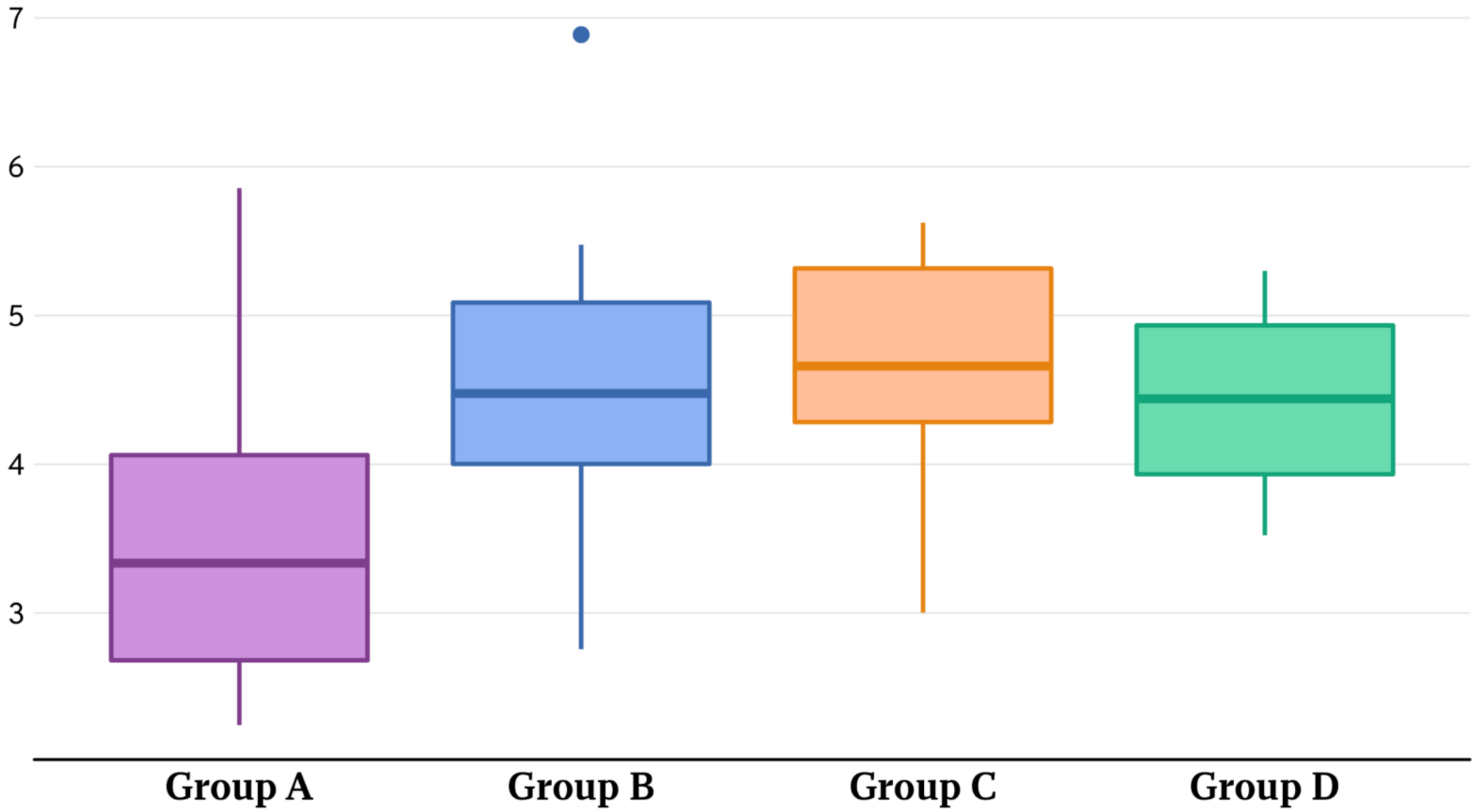
Beyond Bar and Box Plots

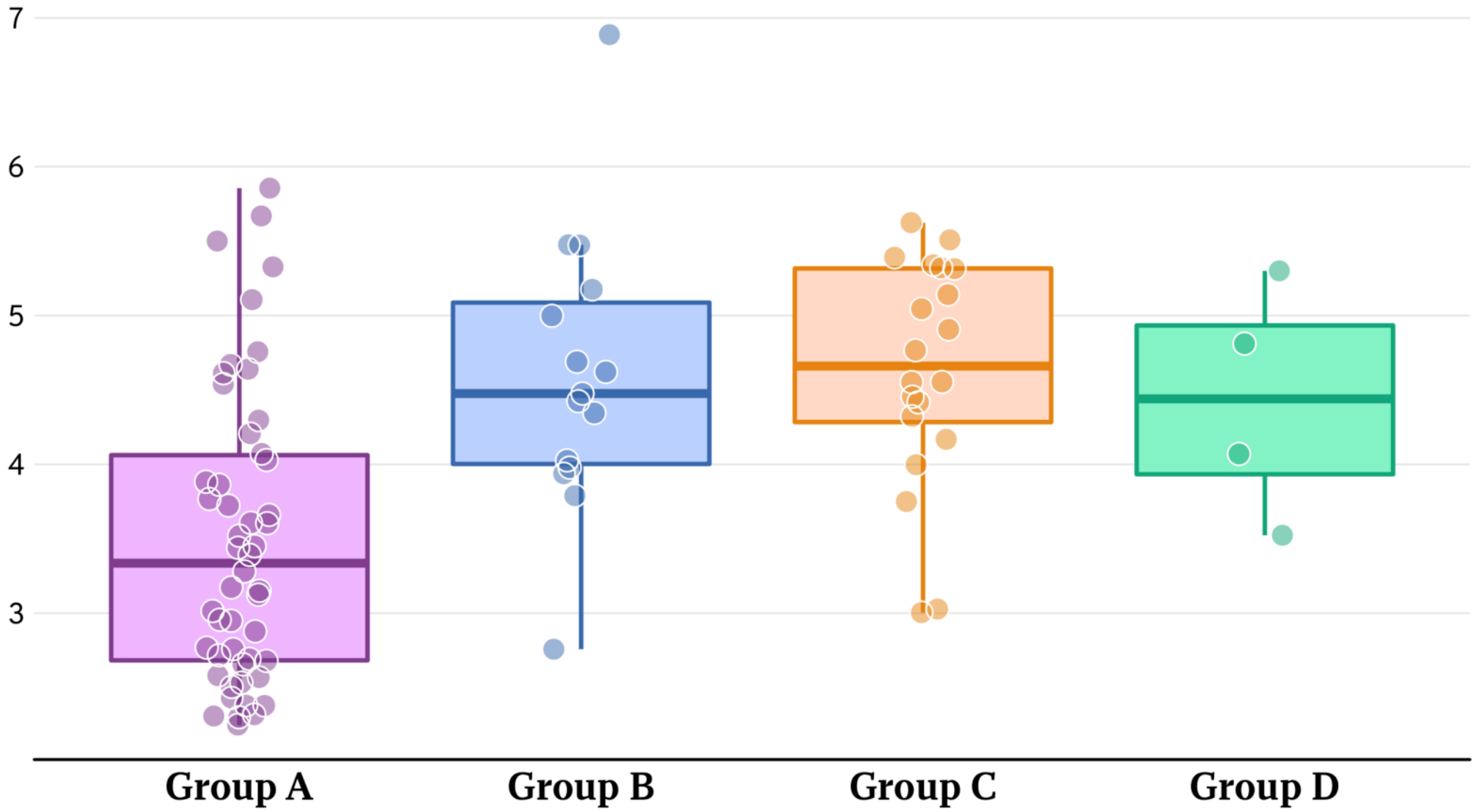


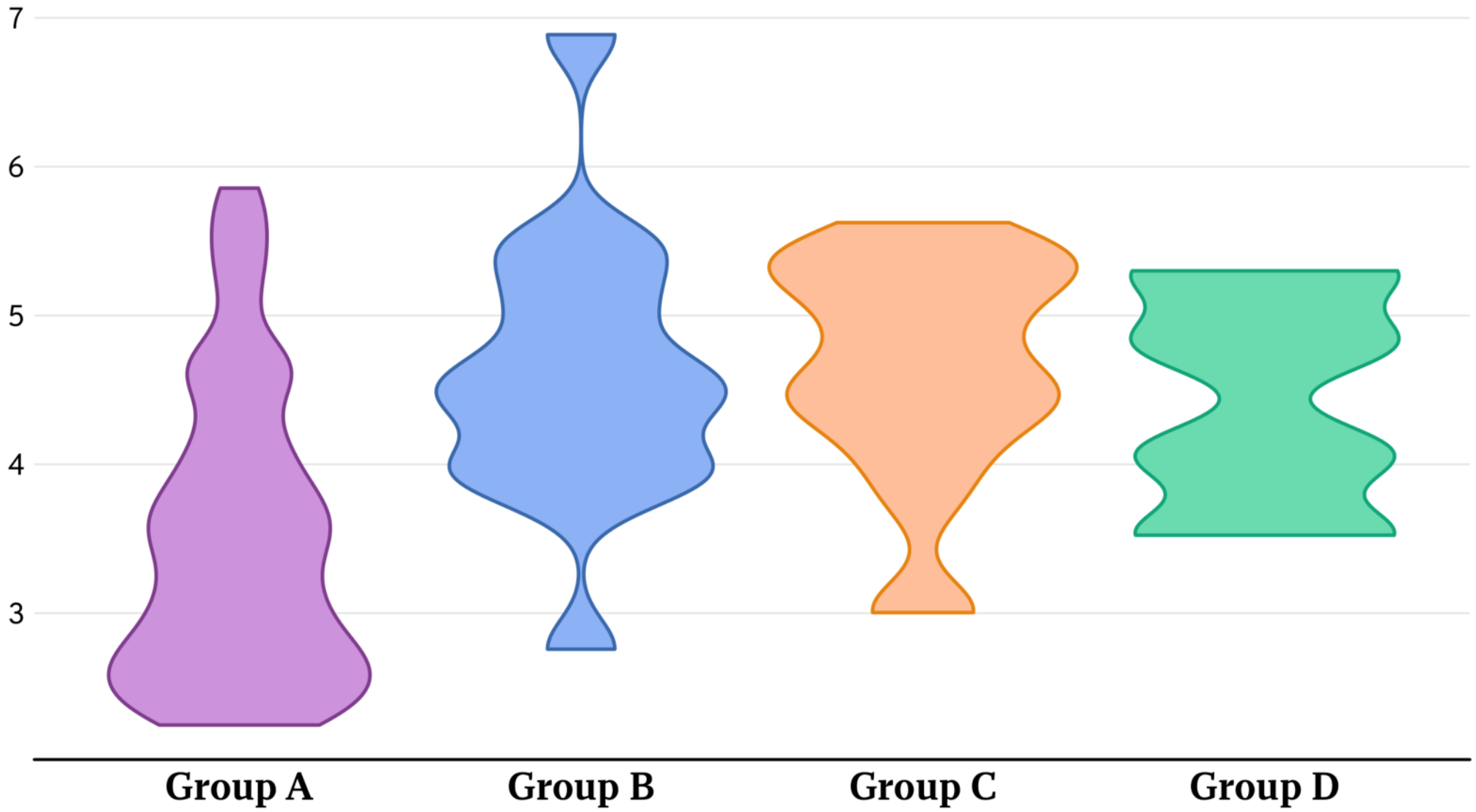

Number of observations
50
40
30
20
10
0
Group A
Group B
Group C
Group D

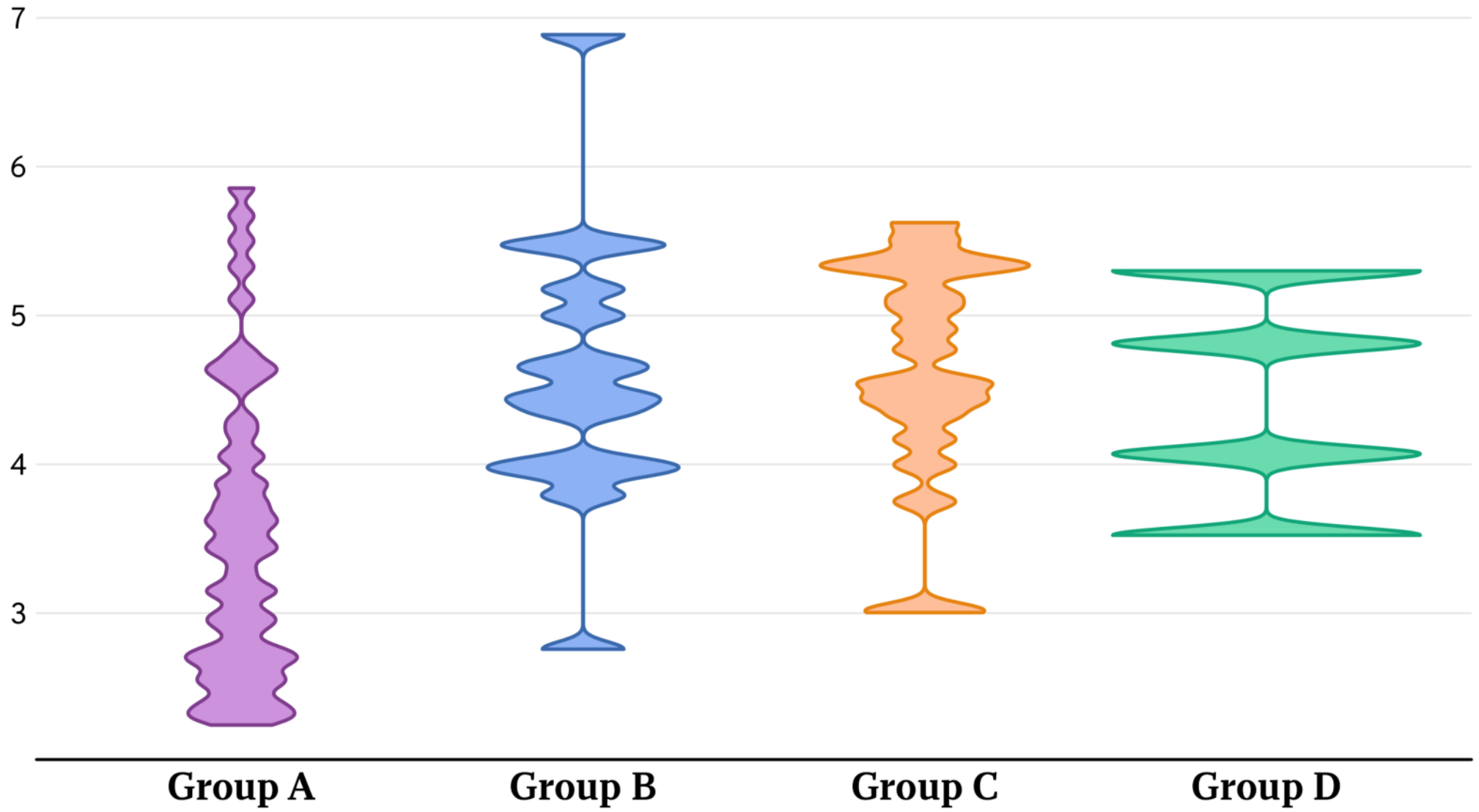

Mean value
0
1
2
3
4
Group A
Group B
Group C
Group D

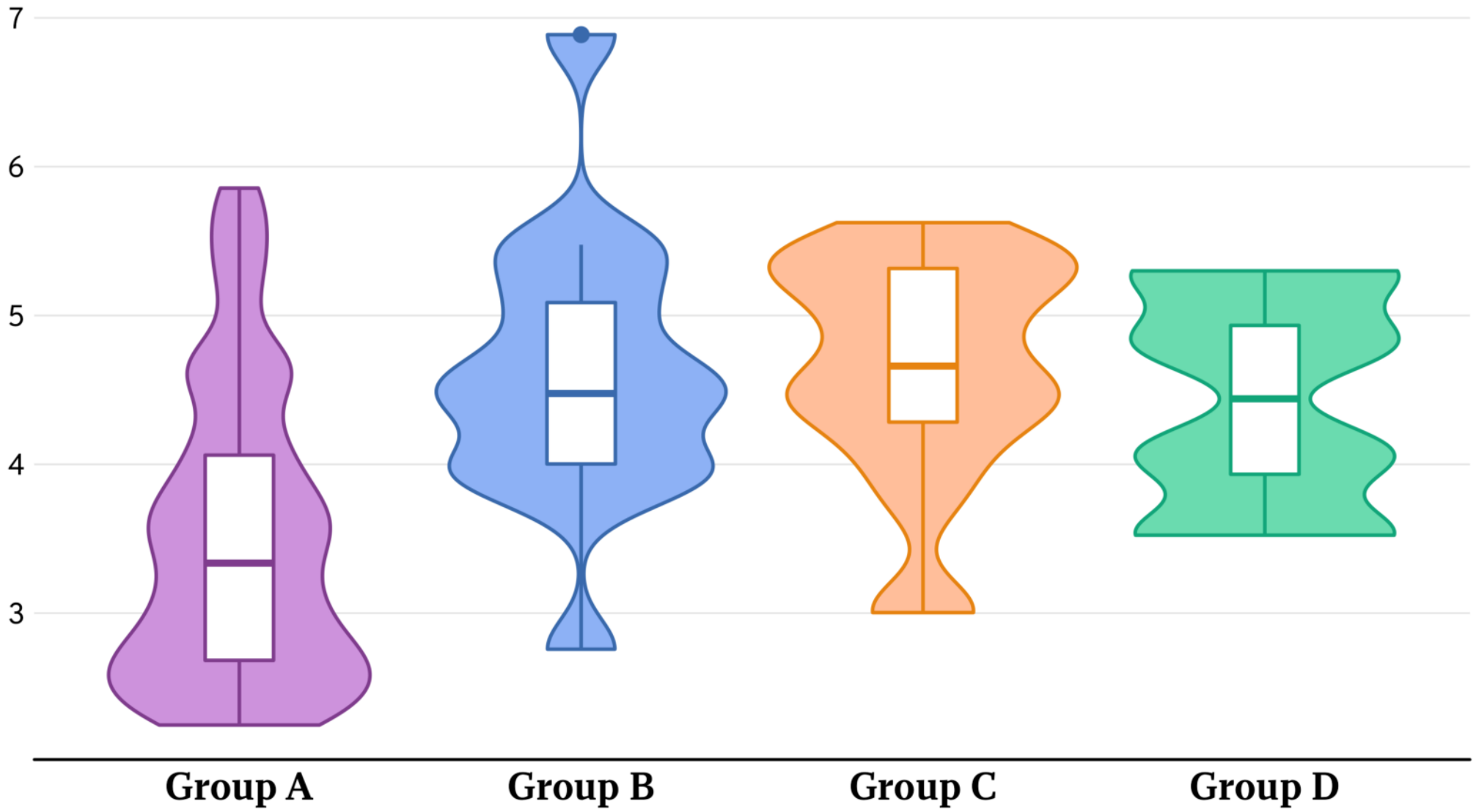

Mean value
5
4
3
2
1
Group A
Group B
Group C
Group D

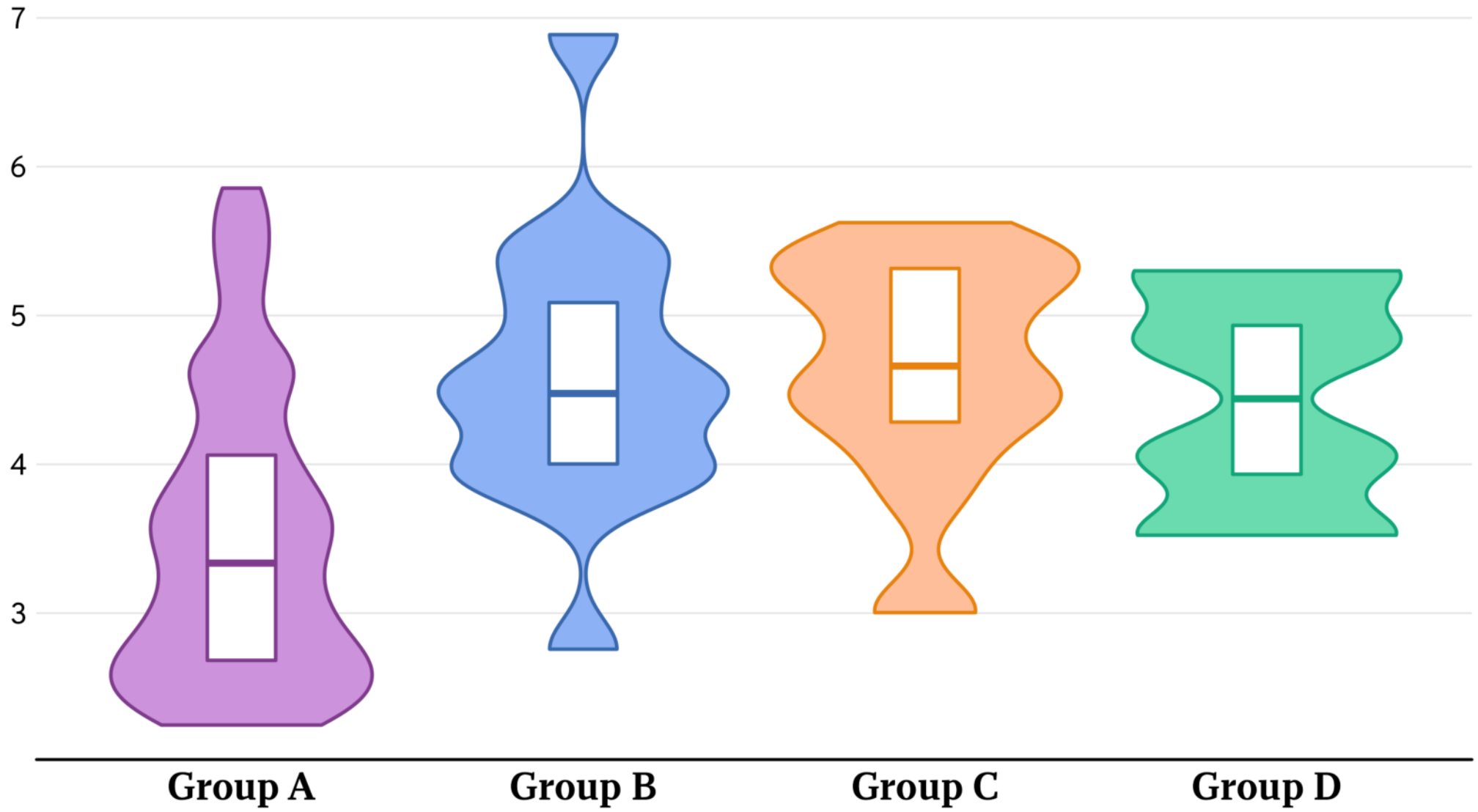

7
6
5
4
3
Group A
Group B
Group C
Group D

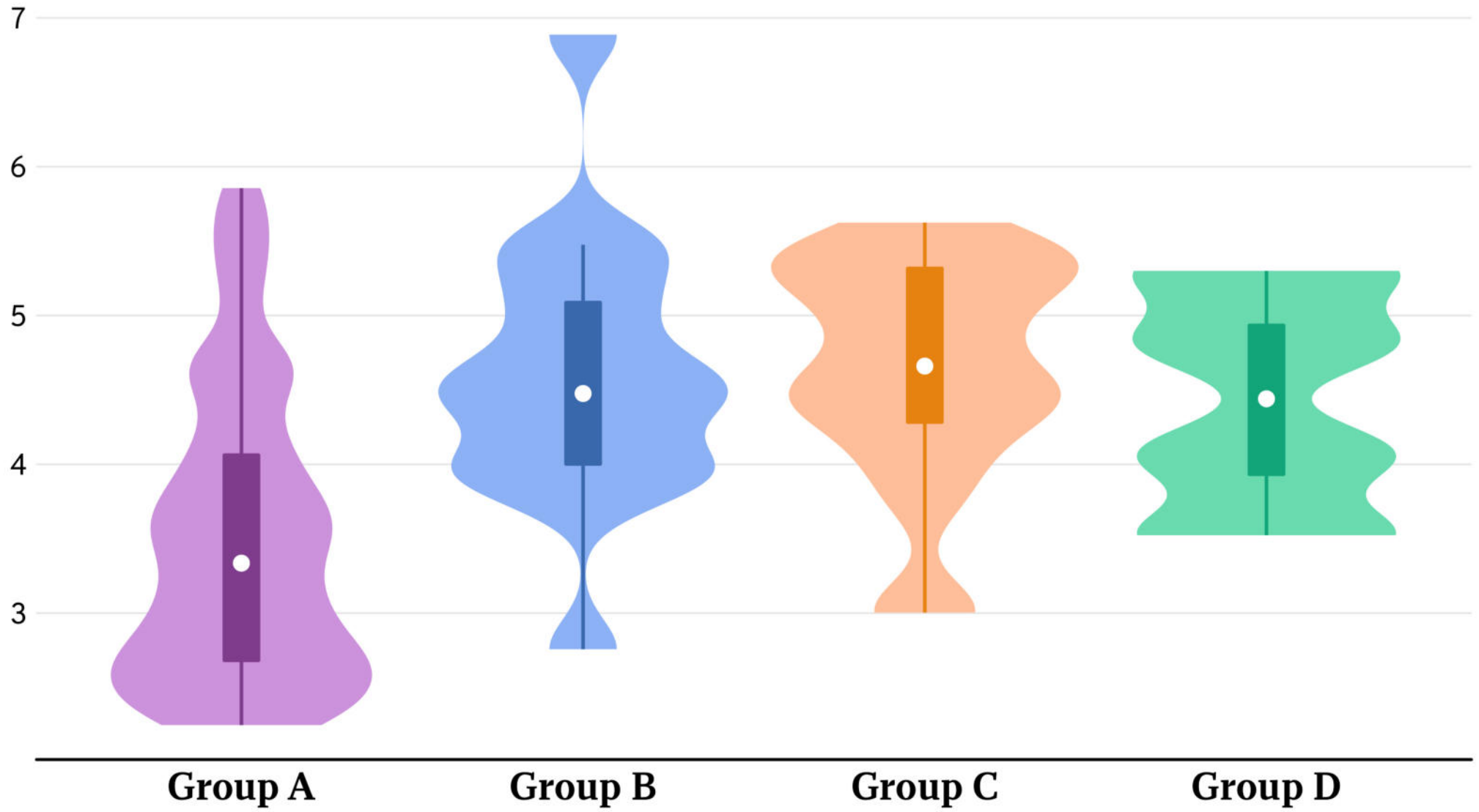

7
6
5
4
3
Group A
Group B
Group C
Group D

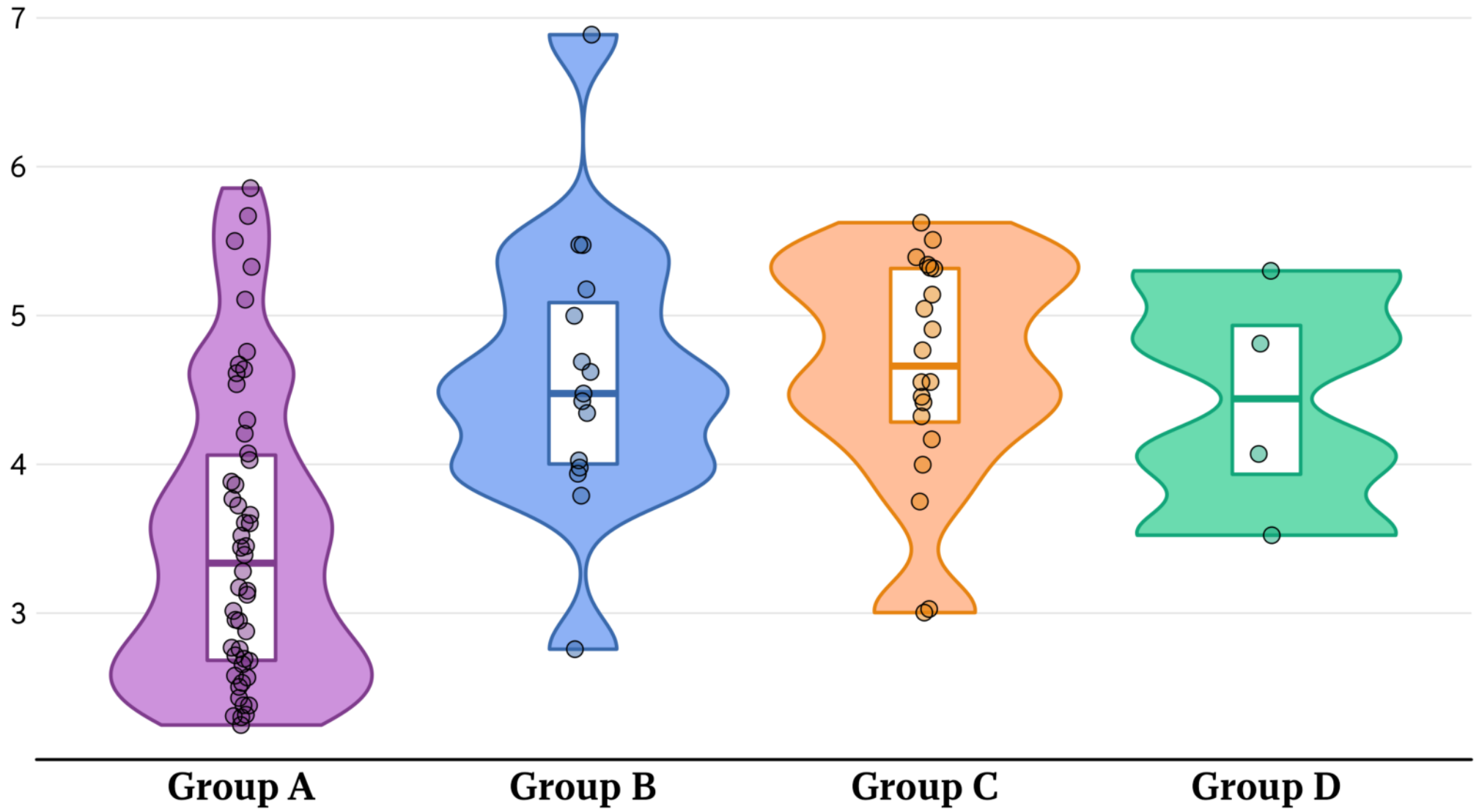

7
6
5
4
3
Group A
Group B
Group C
Group D

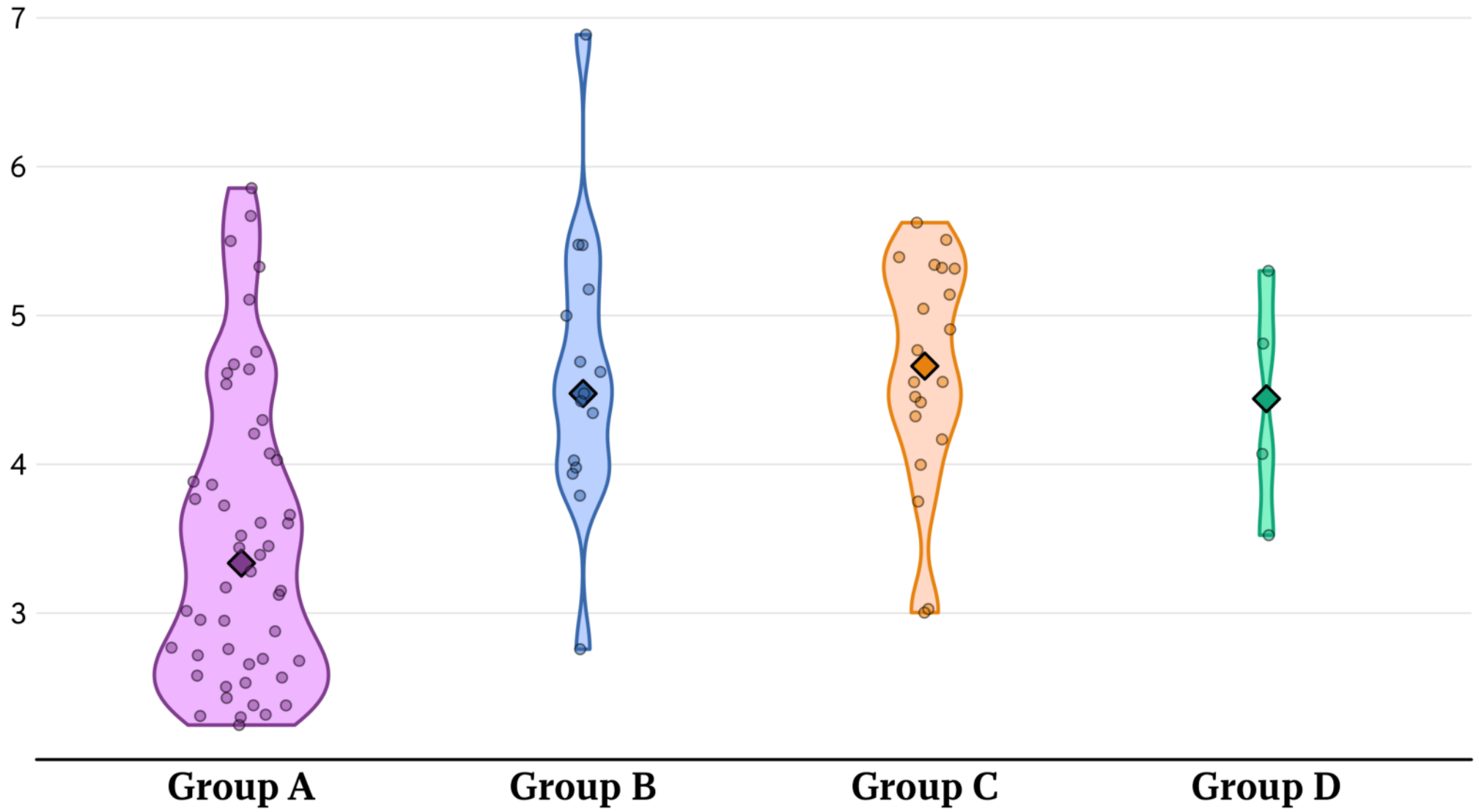

7
6
5
4
3
Group A
Group B
Group C
Group D

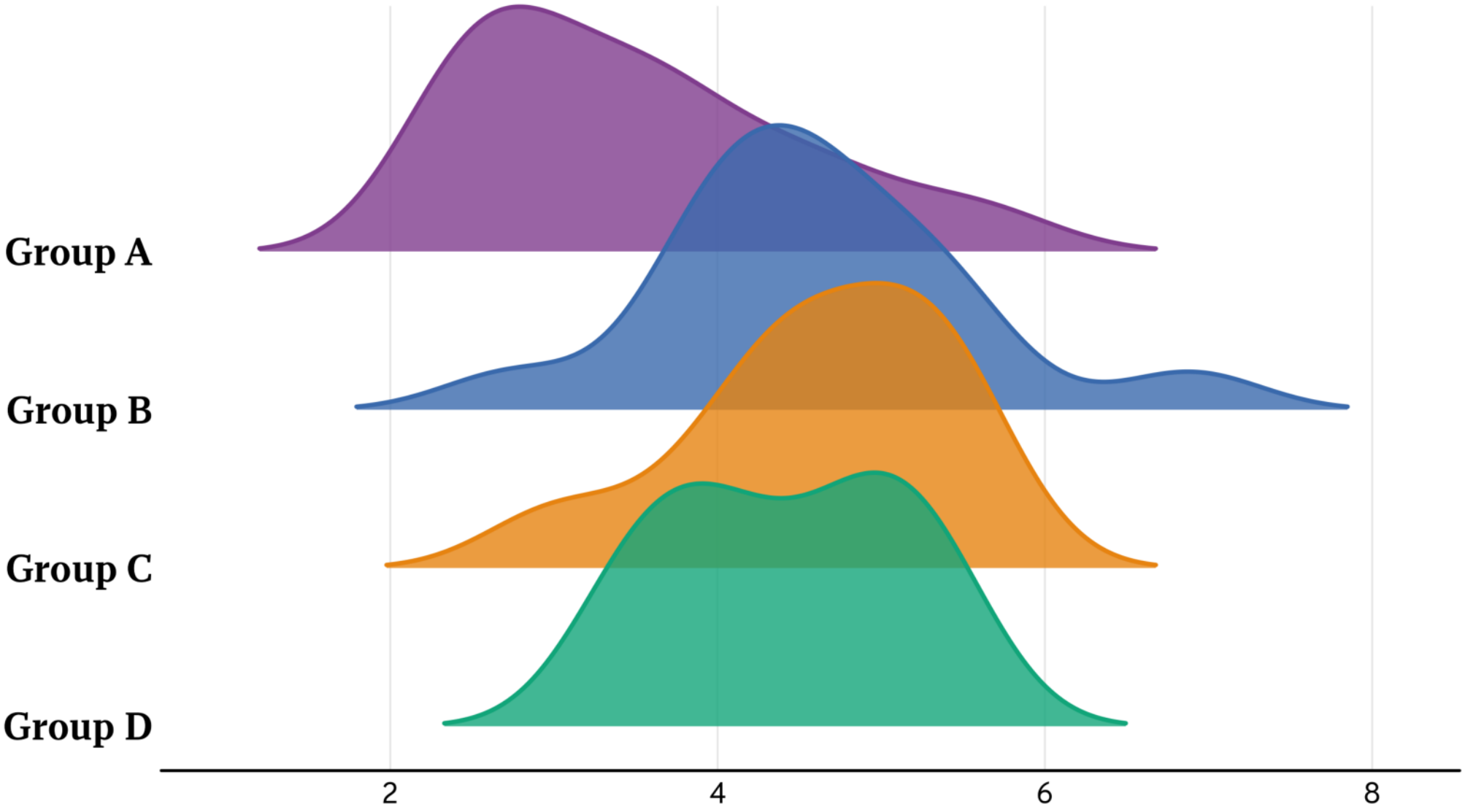

7
6
5
4
3
Group A
Group B
Group C
Group D

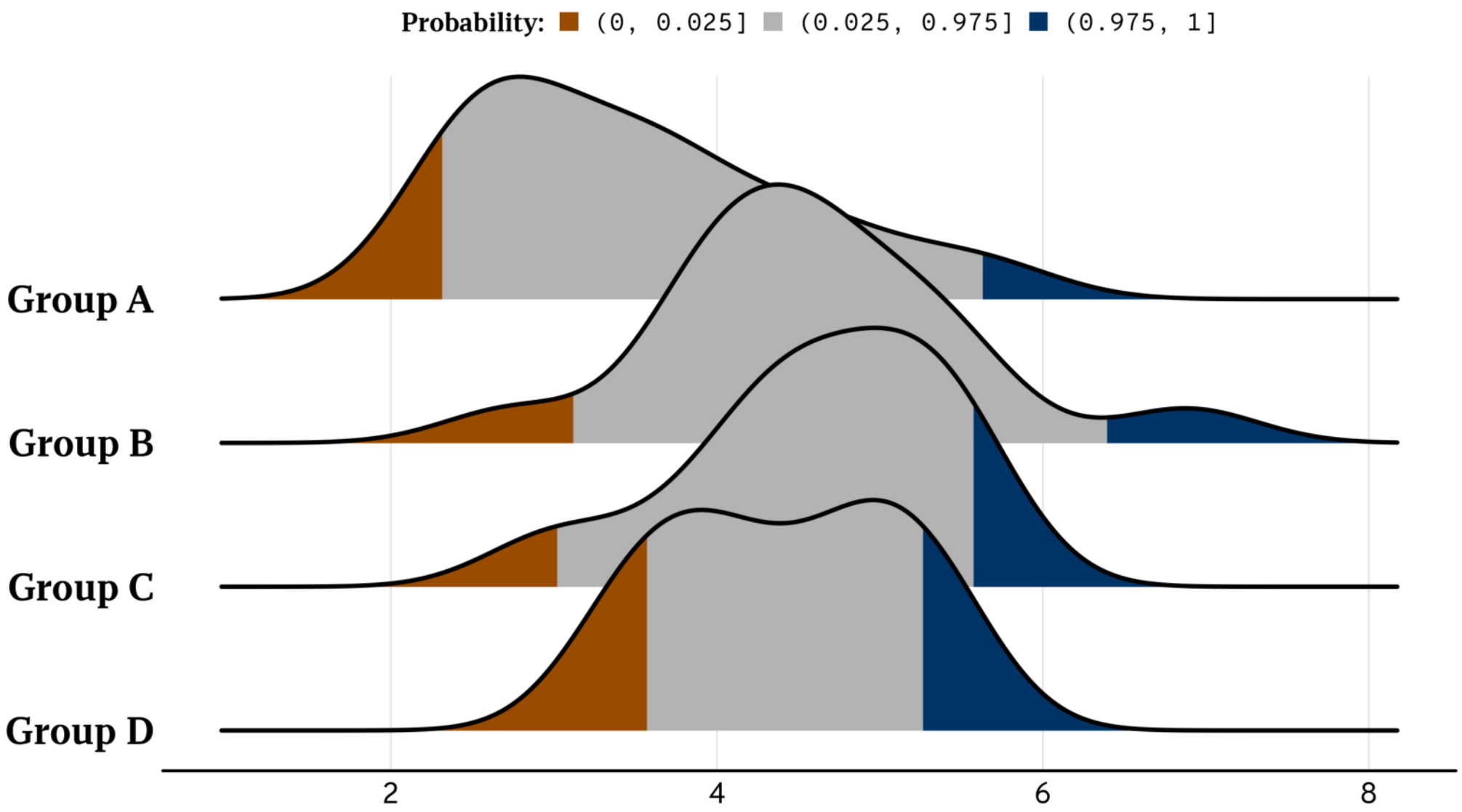

7
6
5
4
3
Group A
Group B
Group C
Group D

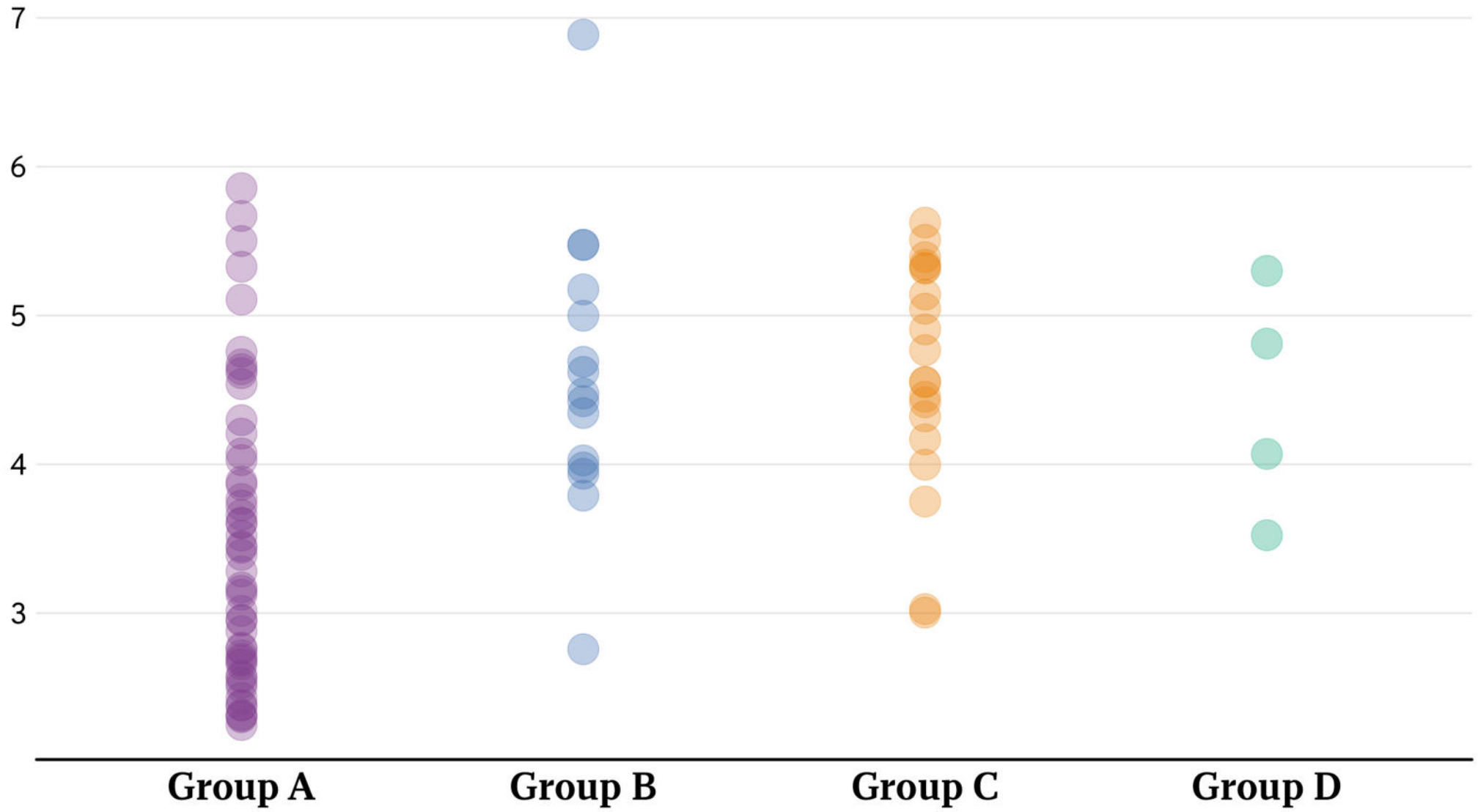

7
6
5
4
3
Group A
Group B
Group C
Group D

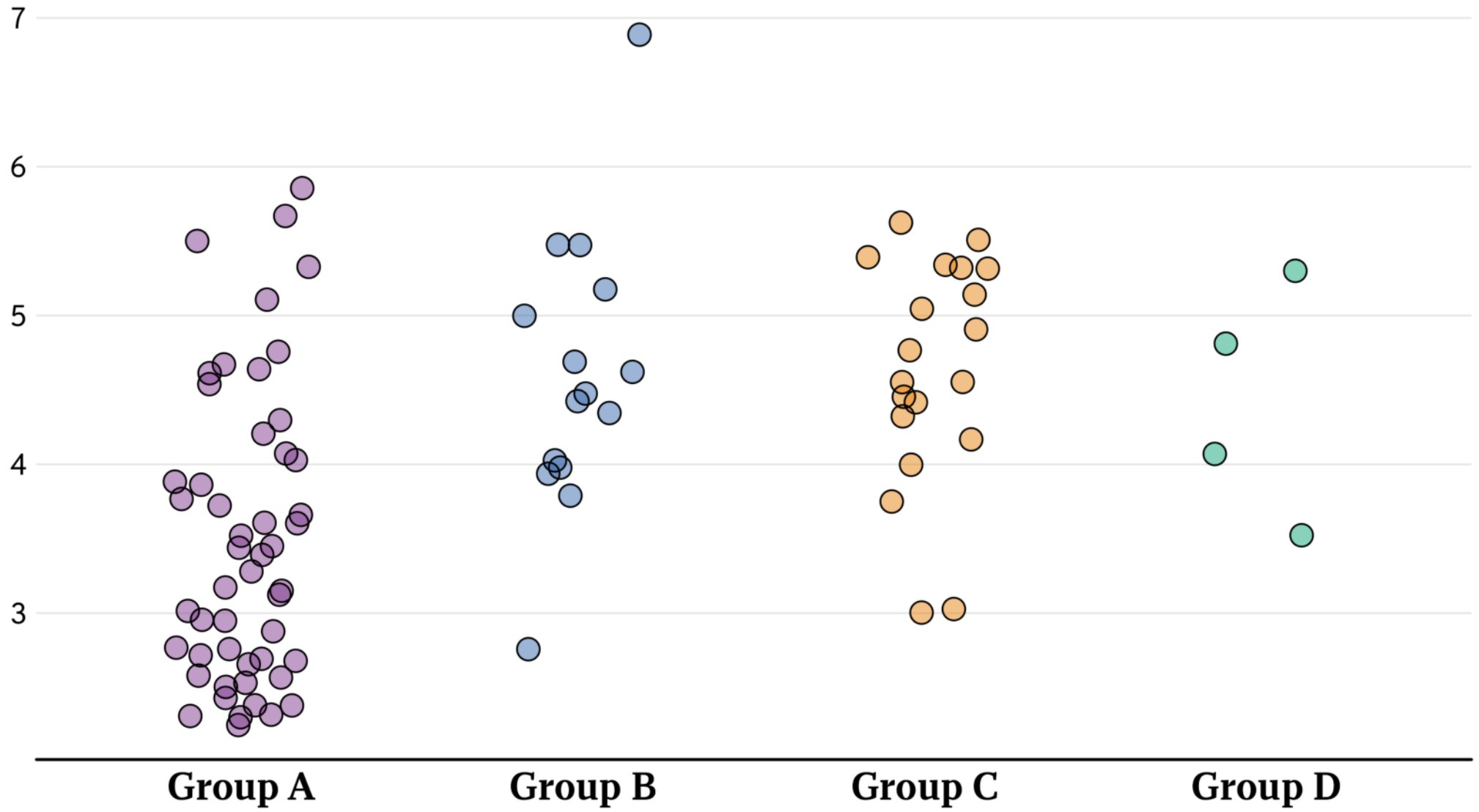

7
6
5
4
3
Group A
Group B
Group C
Group D

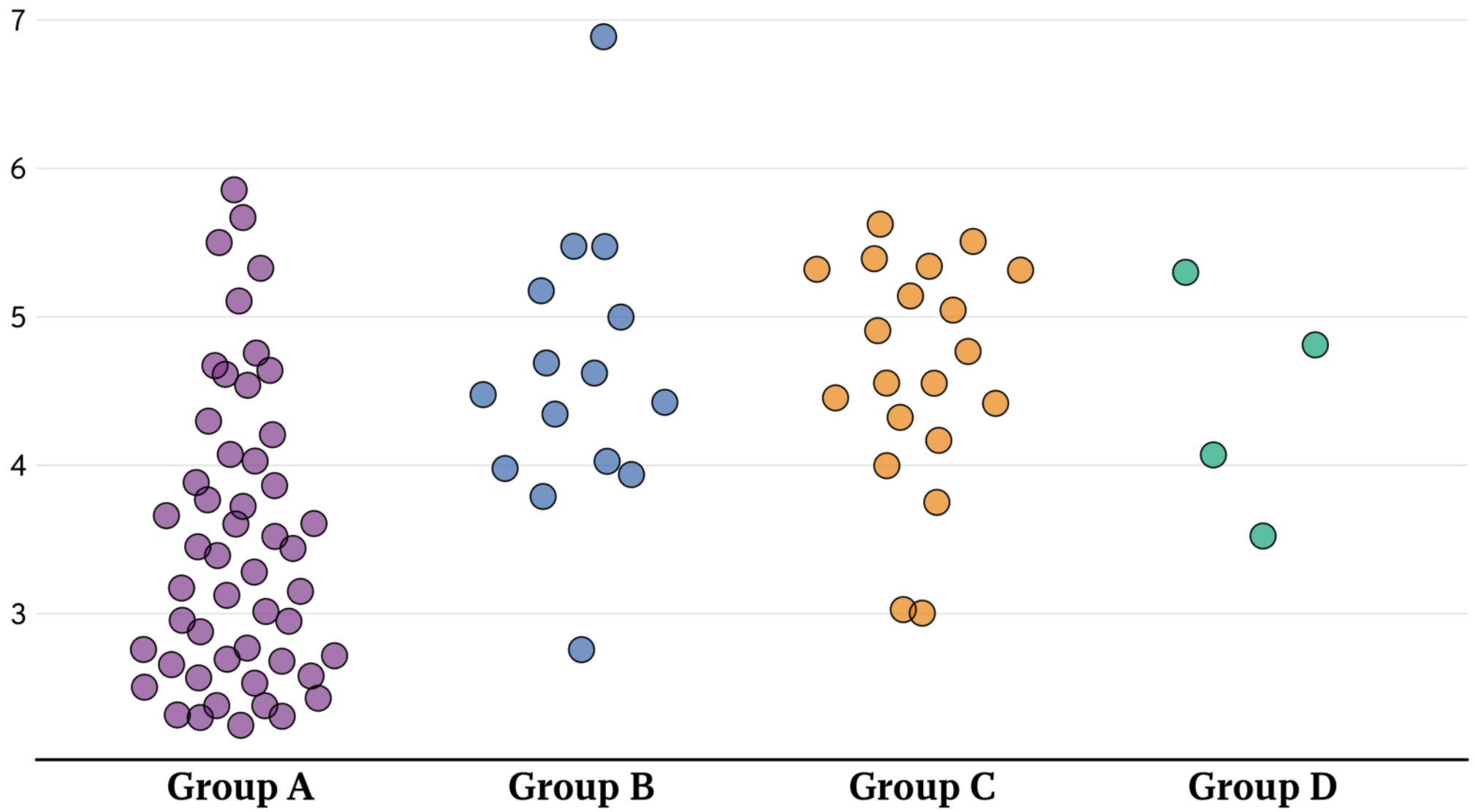

7
6
5
4
3
Group A
Group B
Group C
Group D

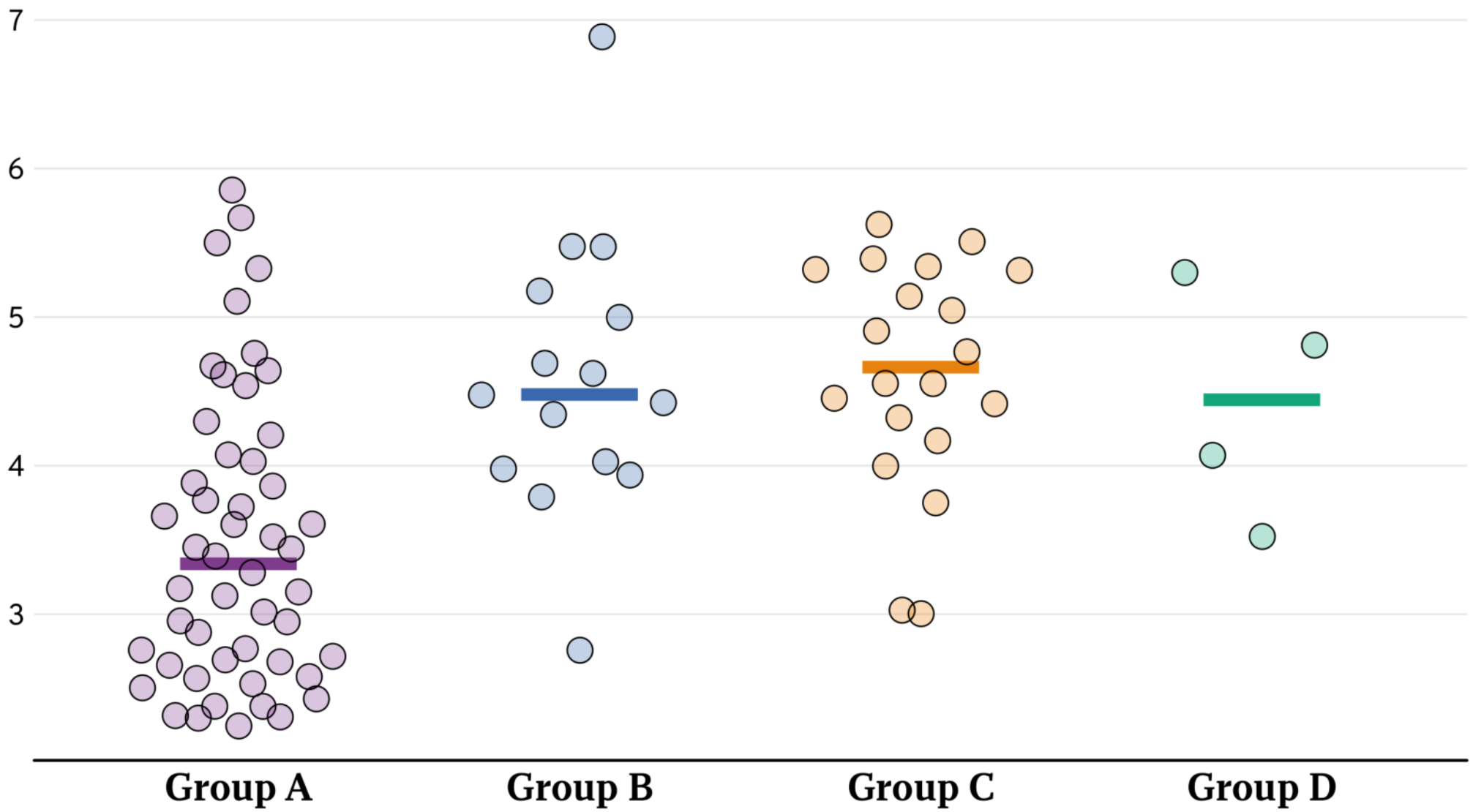

Group A
Group B
Group C
Group D
2
4
6
8

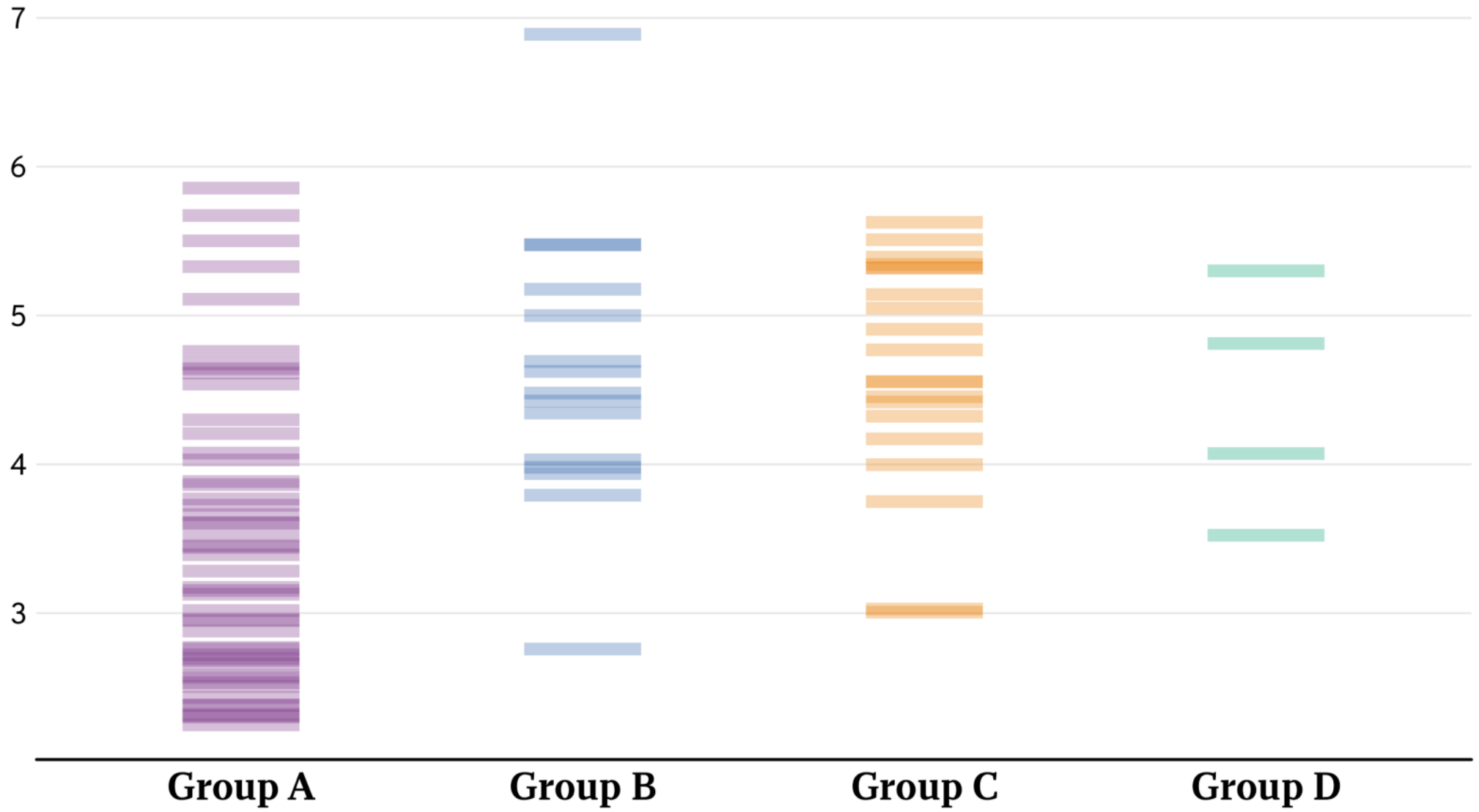

Probability: (0, 0.025] (0.025, 0.975] (0.975, 1]
Group A
Group B
Group C
Group D
2
4
6
8

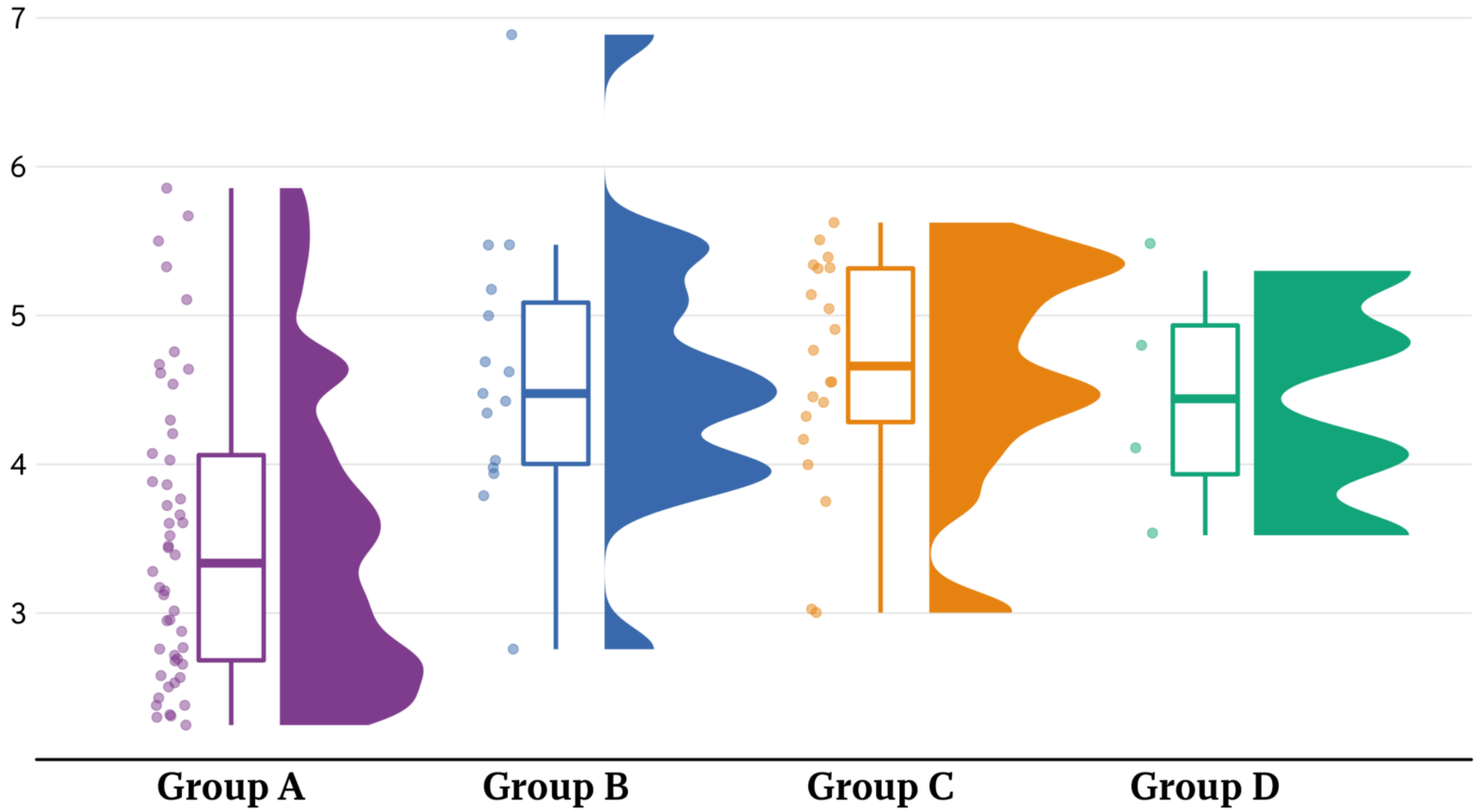

7
6
5
4
3
Group A
Group B
Group C
Group D

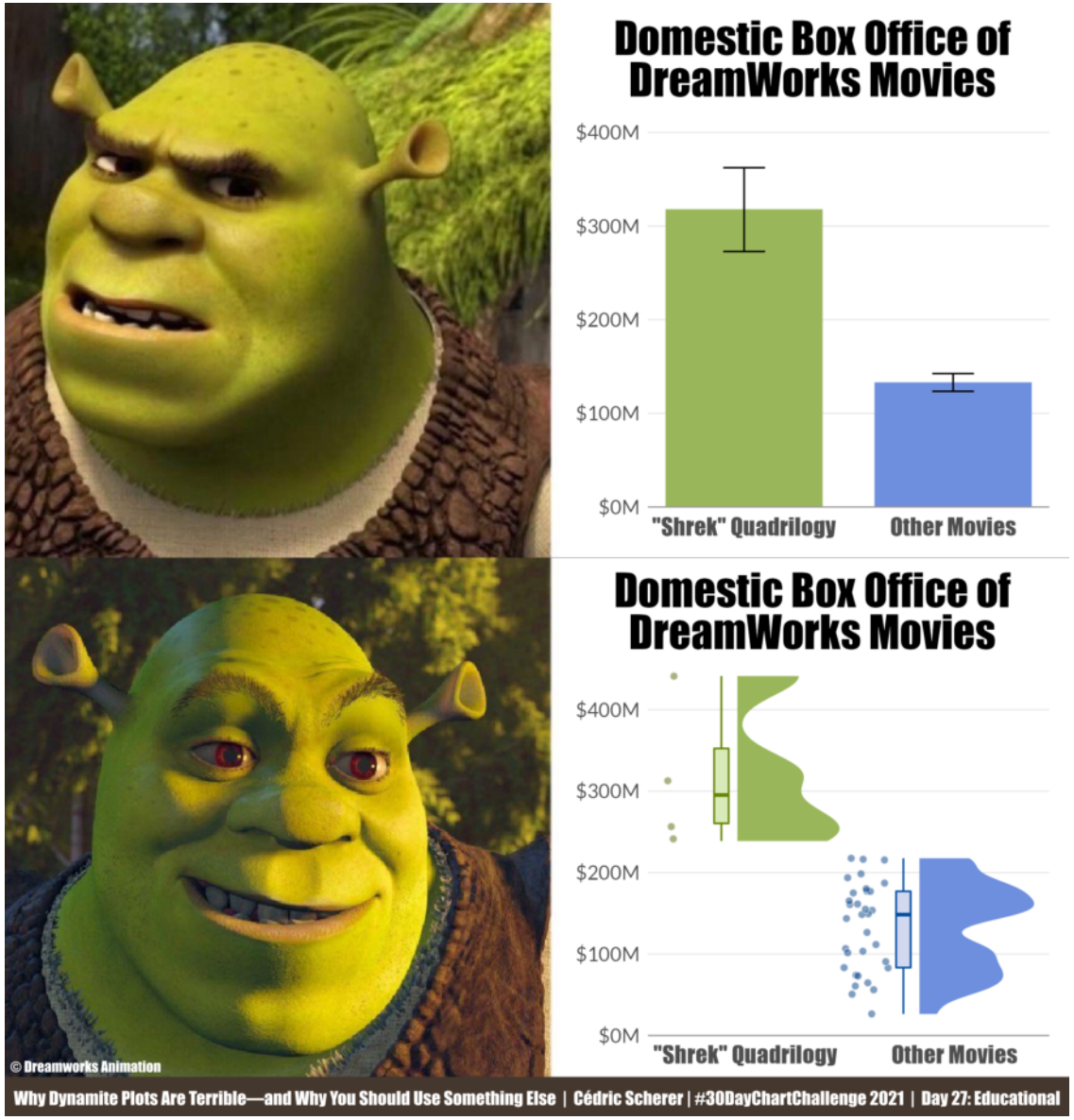

7
6
5
4
3
Group A
Group B
Group C
Group D

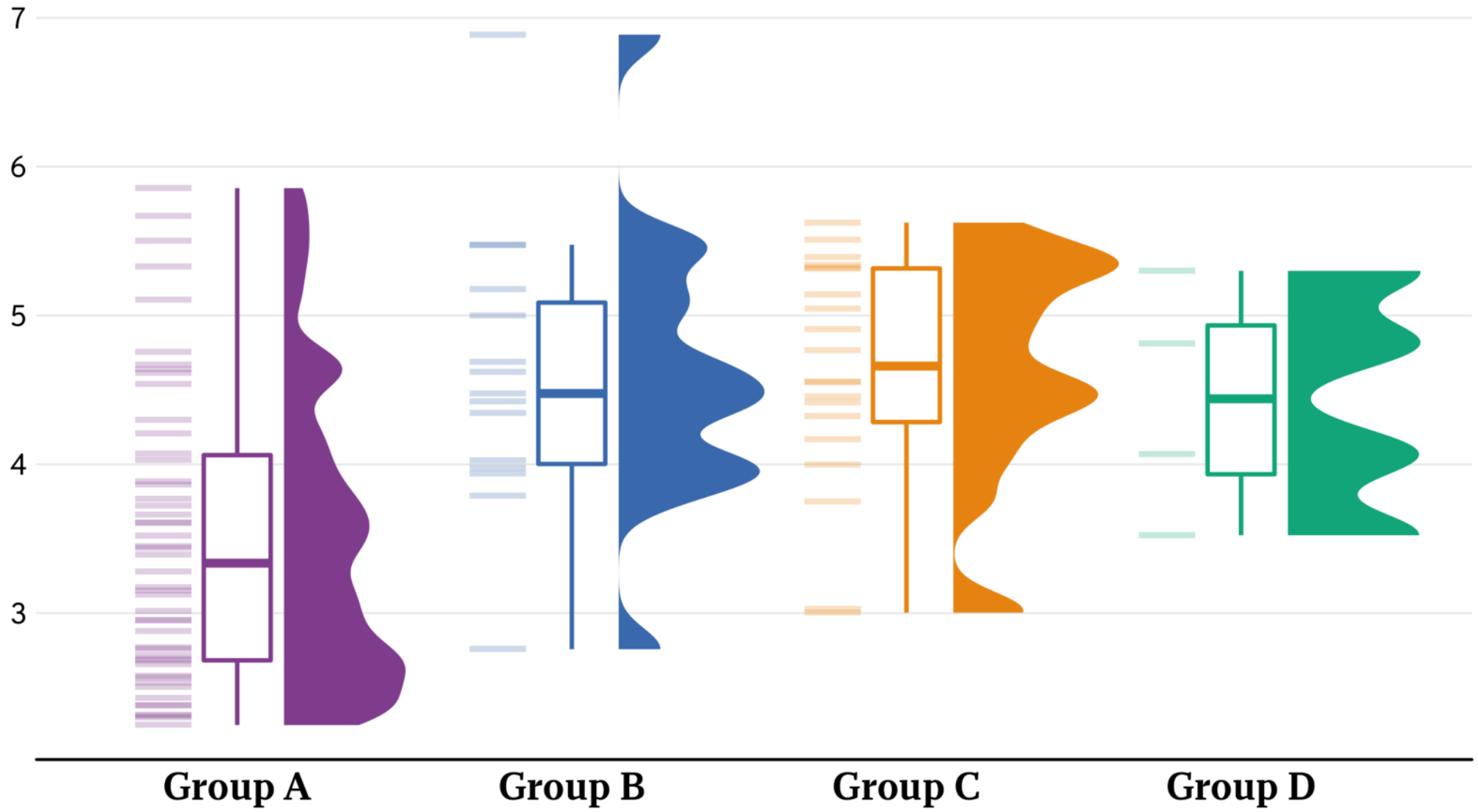

7
6
5
4
3
Group A
Group B
Group C
Group D


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

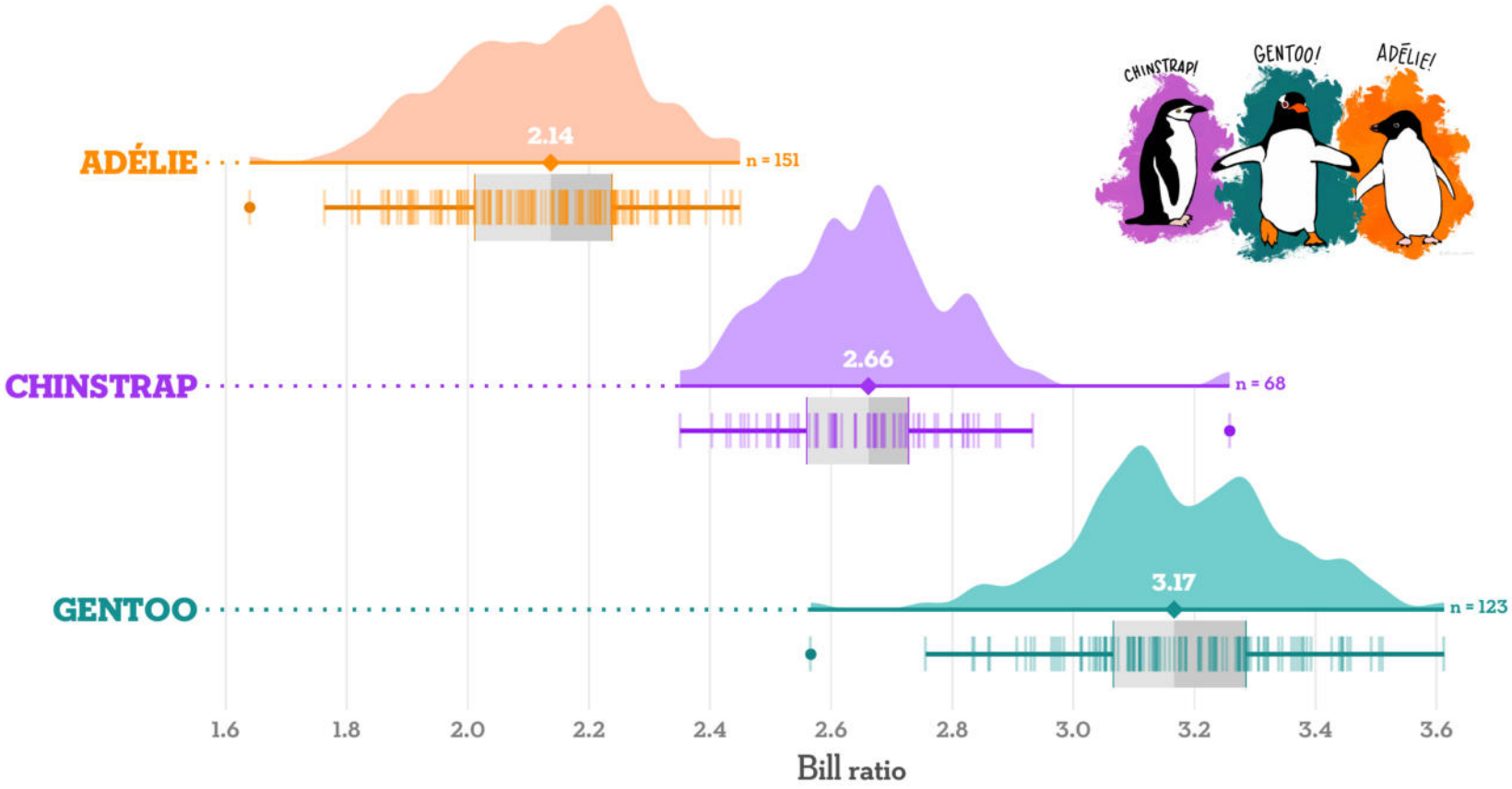

7
6
5
4
3
Group A
Group B
Group C
Group D


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


Domestic Box Office of DreamWorks Movies
$400M
$300M
$200M
$100M
$0M
"Shrek" Quadrilogy
Other Movies
Domestic Box Office of DreamWorks Movies
$400M
$300M
$200M
$100M
$0M
"Shrek" Quadrilogy
Other Movies
© Dreamworks Animation
Why Dynamite Plots Are Terrible—and Why You Should Use Something Else | Cédric Scherer | #30DayChartChallenge 2021 | Day 27: Educational

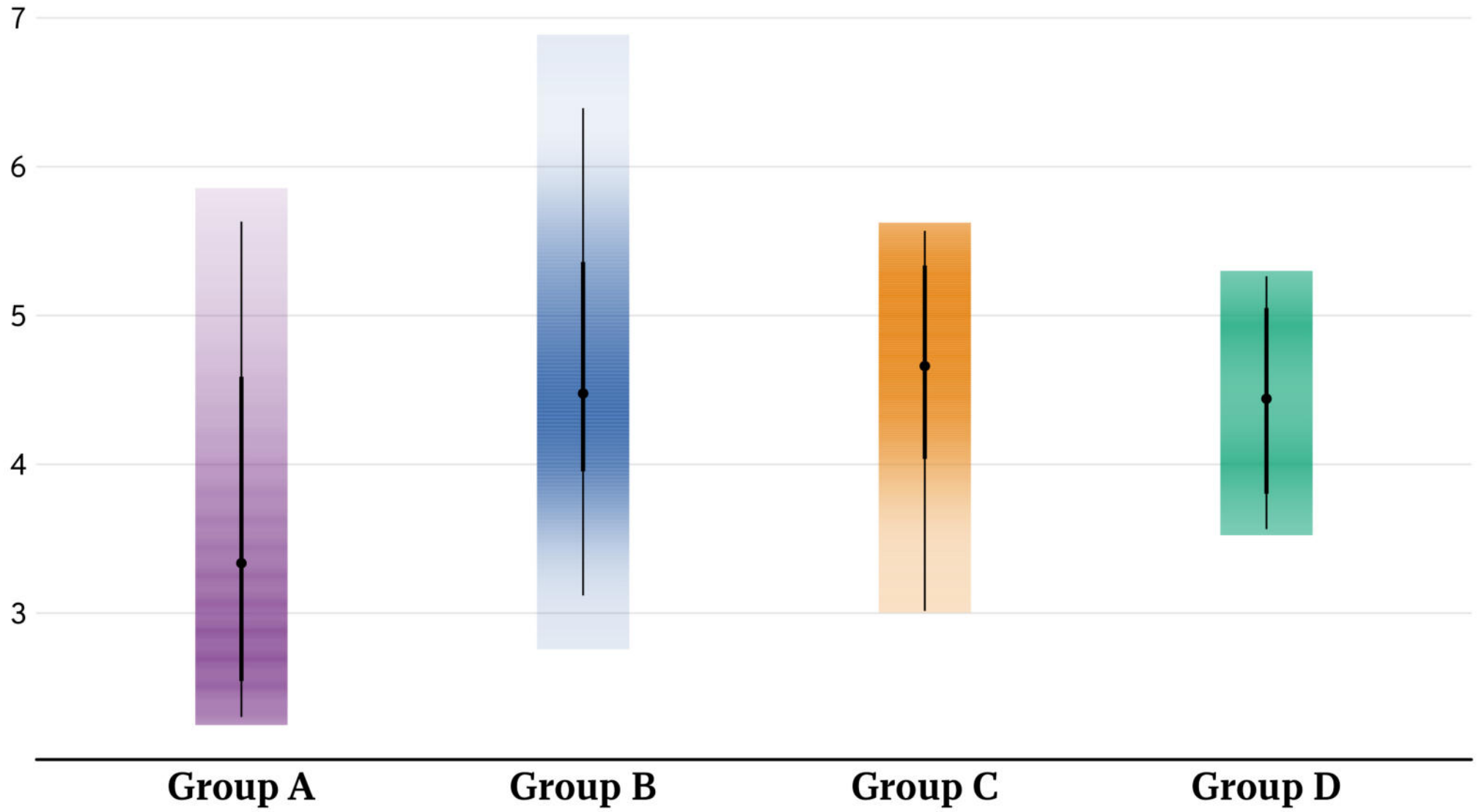

7
6
5
4
3
Group A
Group B
Group C
Group D

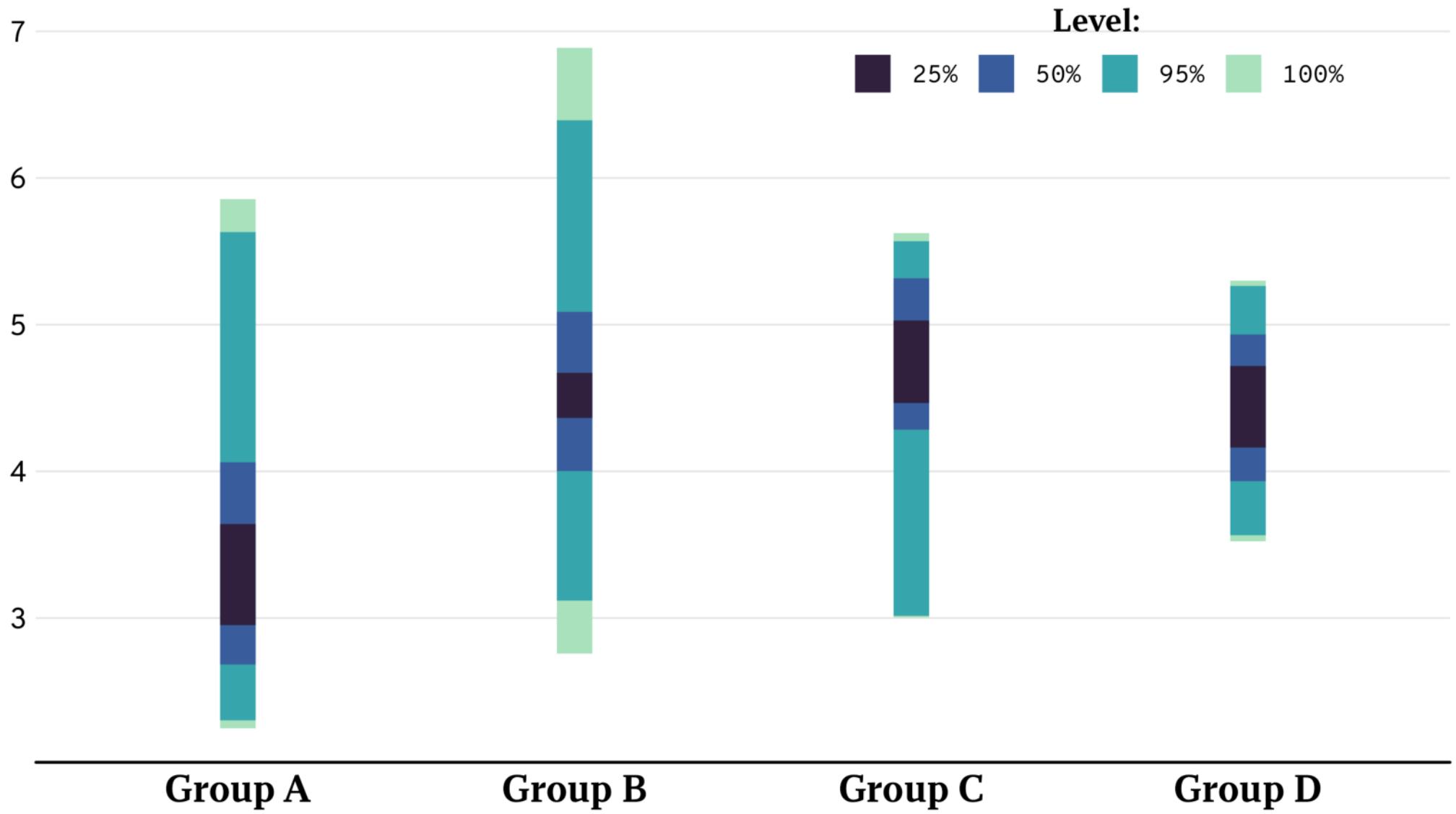
B. Distribution of the bill ratio, estimated as bill length divided by bill depth

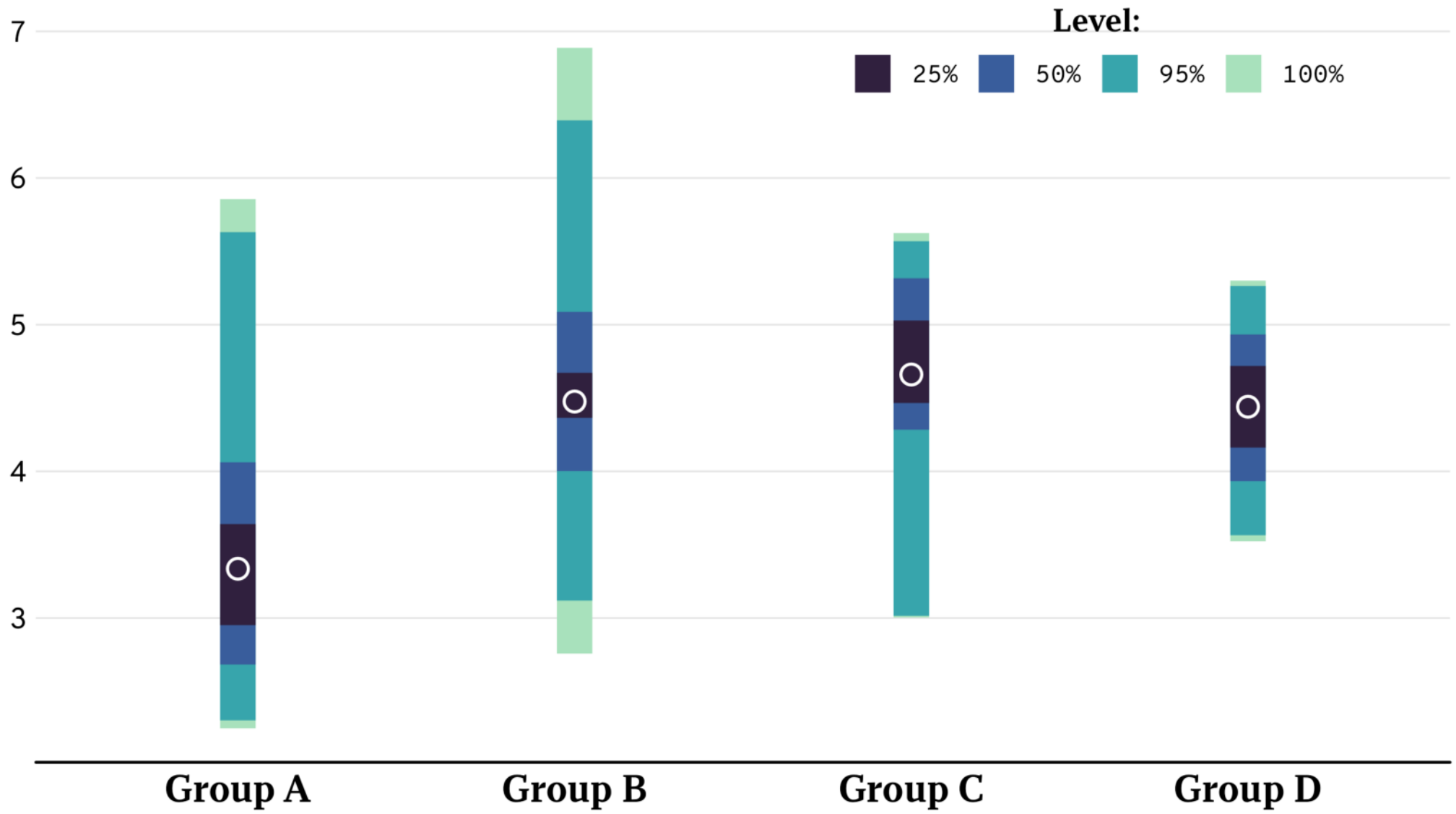


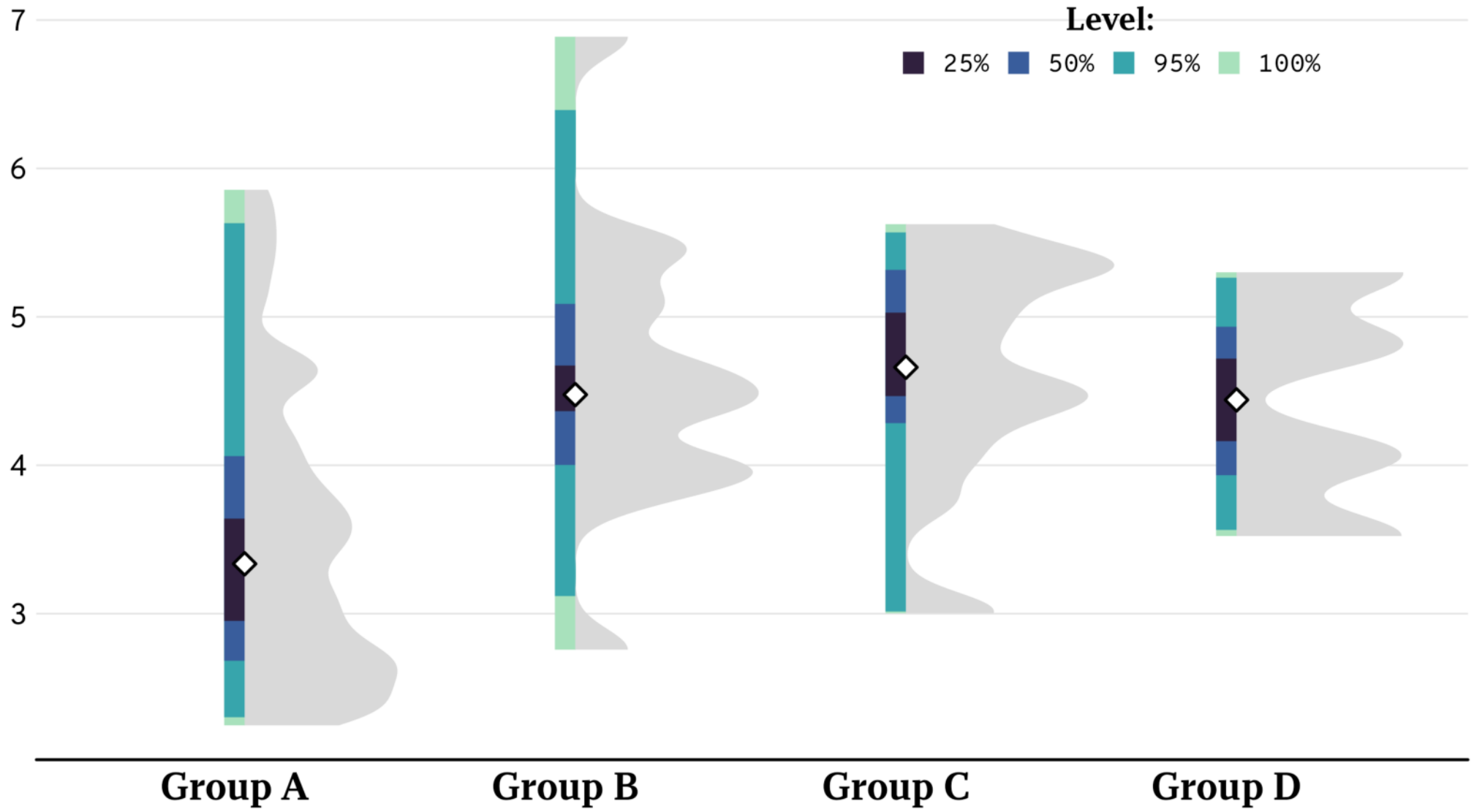
Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.

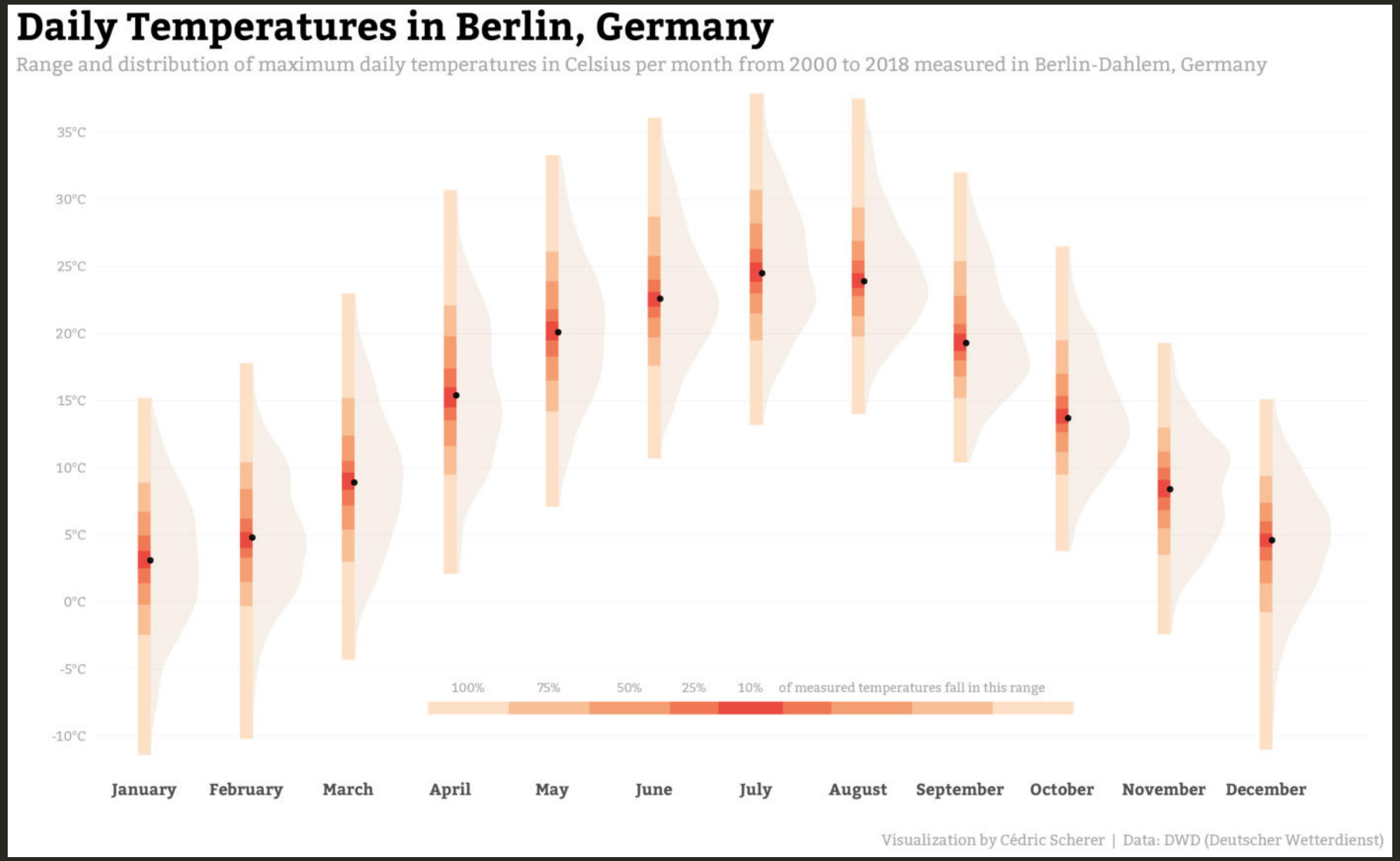
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0090081 • Illustrations: Allison Horst


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

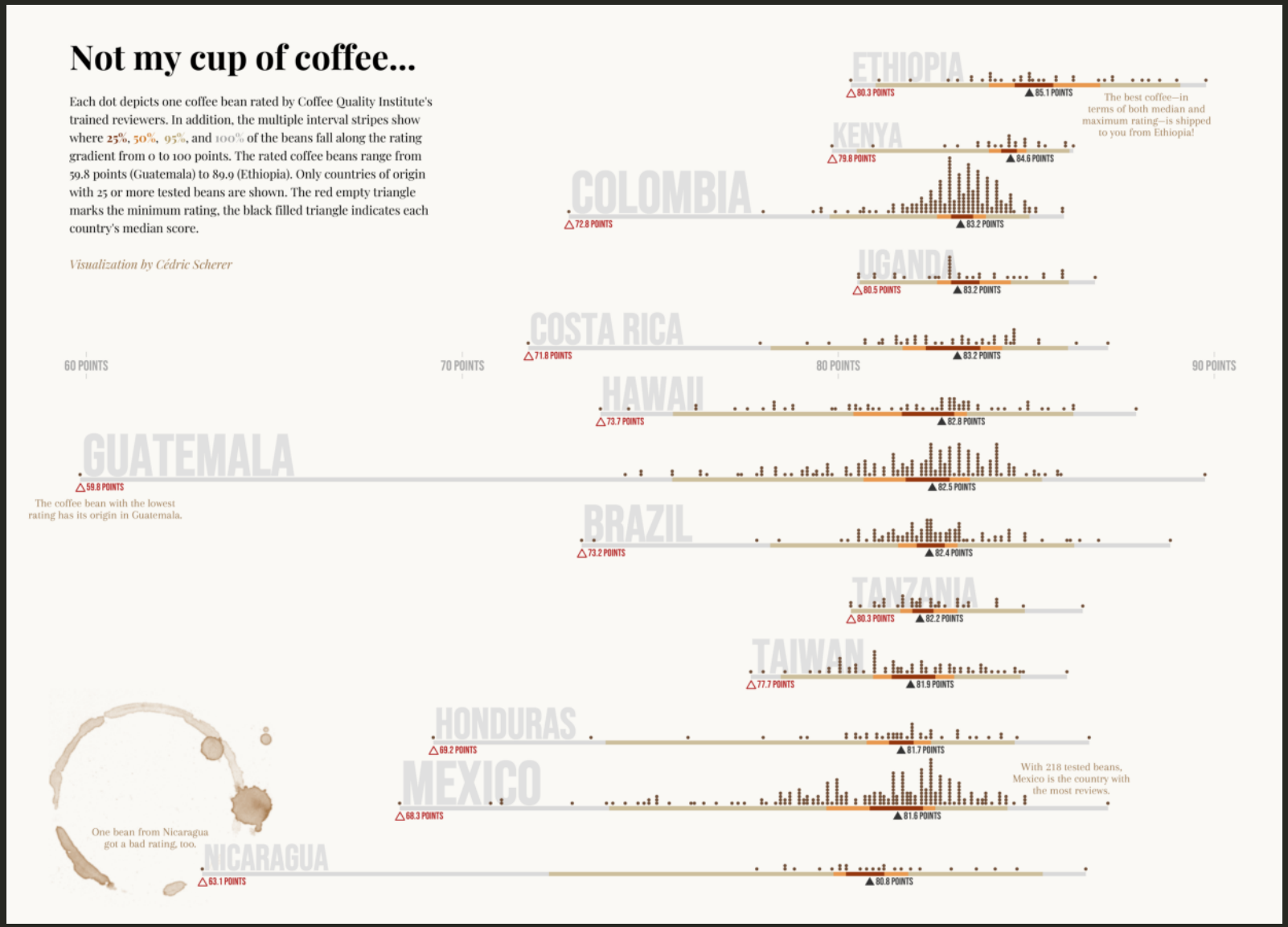

Level:
25%
50%
95%
100%
7
6
5
4
3
Group A
Group B
Group C
Group D


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

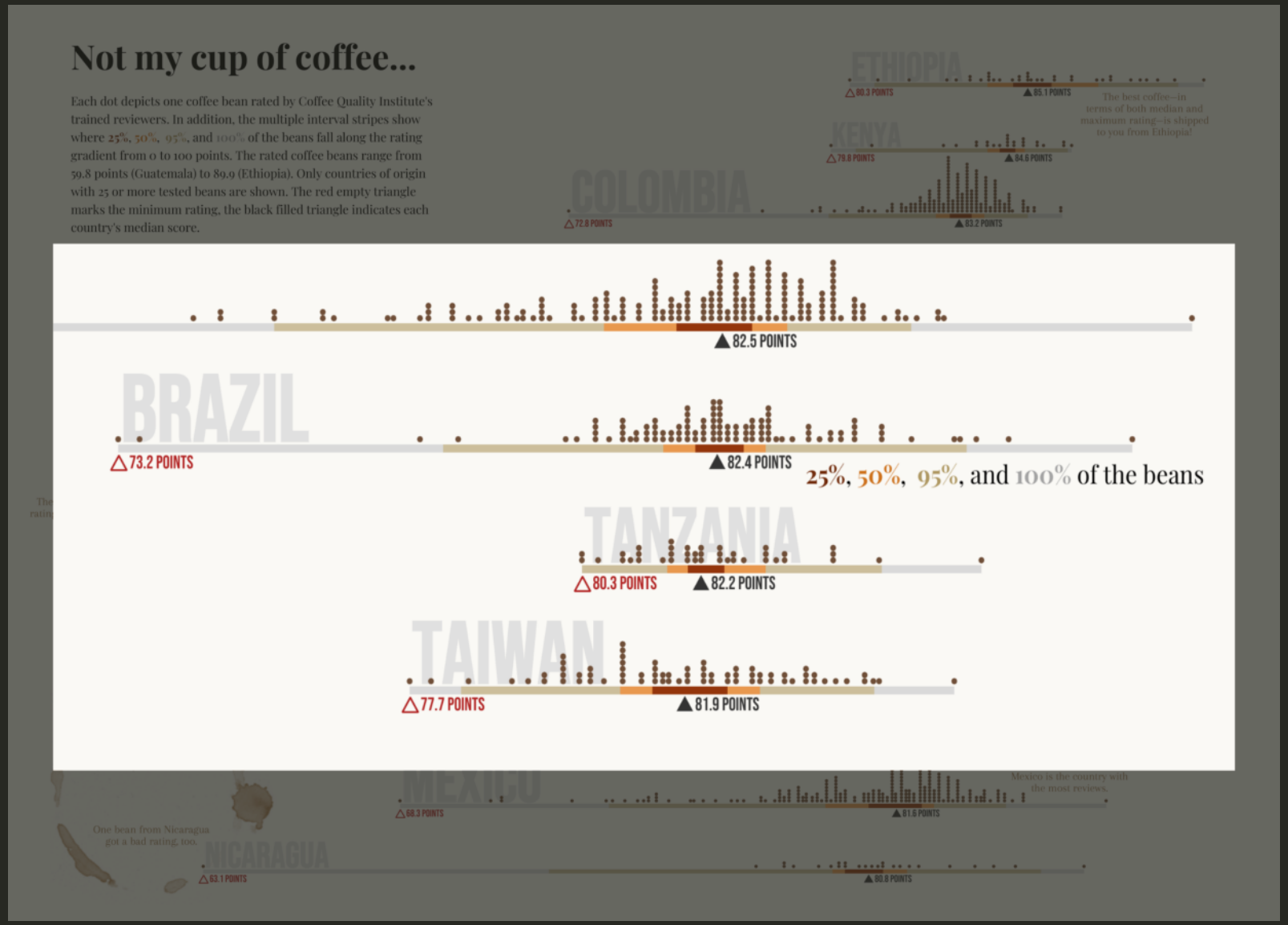

Level:
25%
50%
95%
100%
7
6
5
4
3
Group A
Group B
Group C
Group D



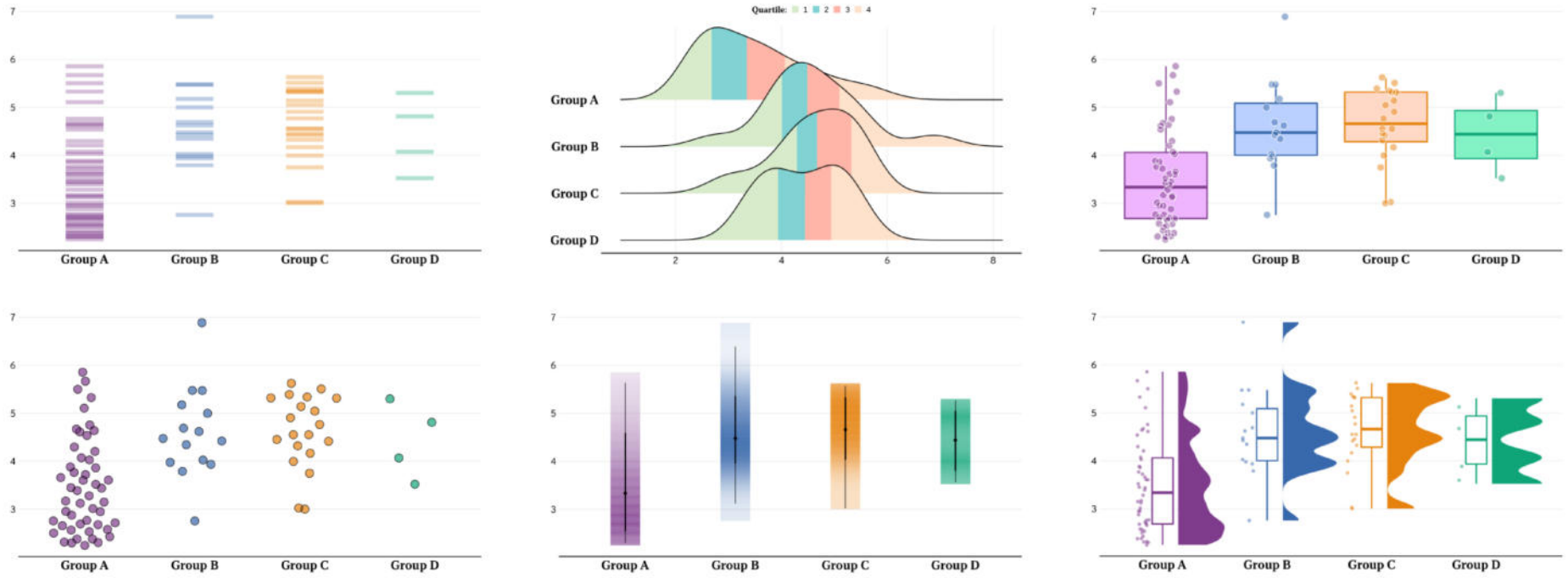
#SWDchallenge contribution September 2019



#TidyTuesday contribution Week 2020/28



#TidyTuesday contribution Week 2020/28



- Always check raw data and sample size
- Try several chart types
- Be open to combine chart types
- Choose chart type with your audience in mind

PLOS | BIOLOGY

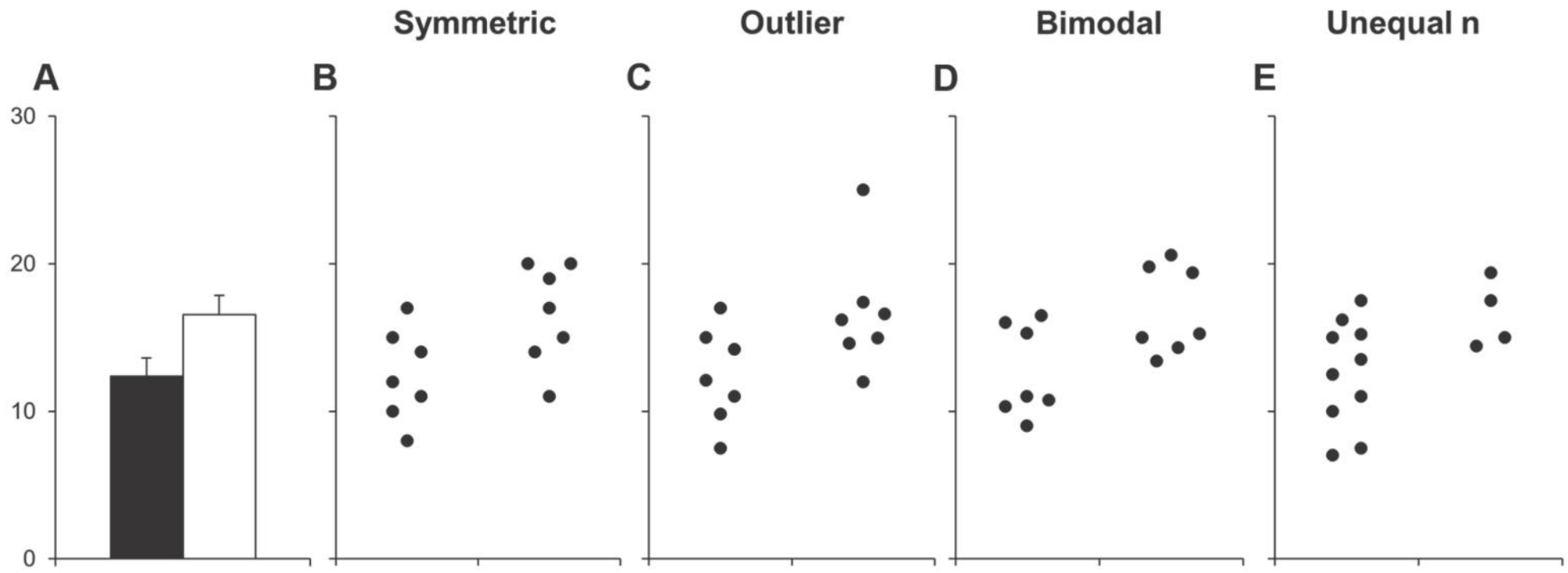
PERSPECTIVE

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

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

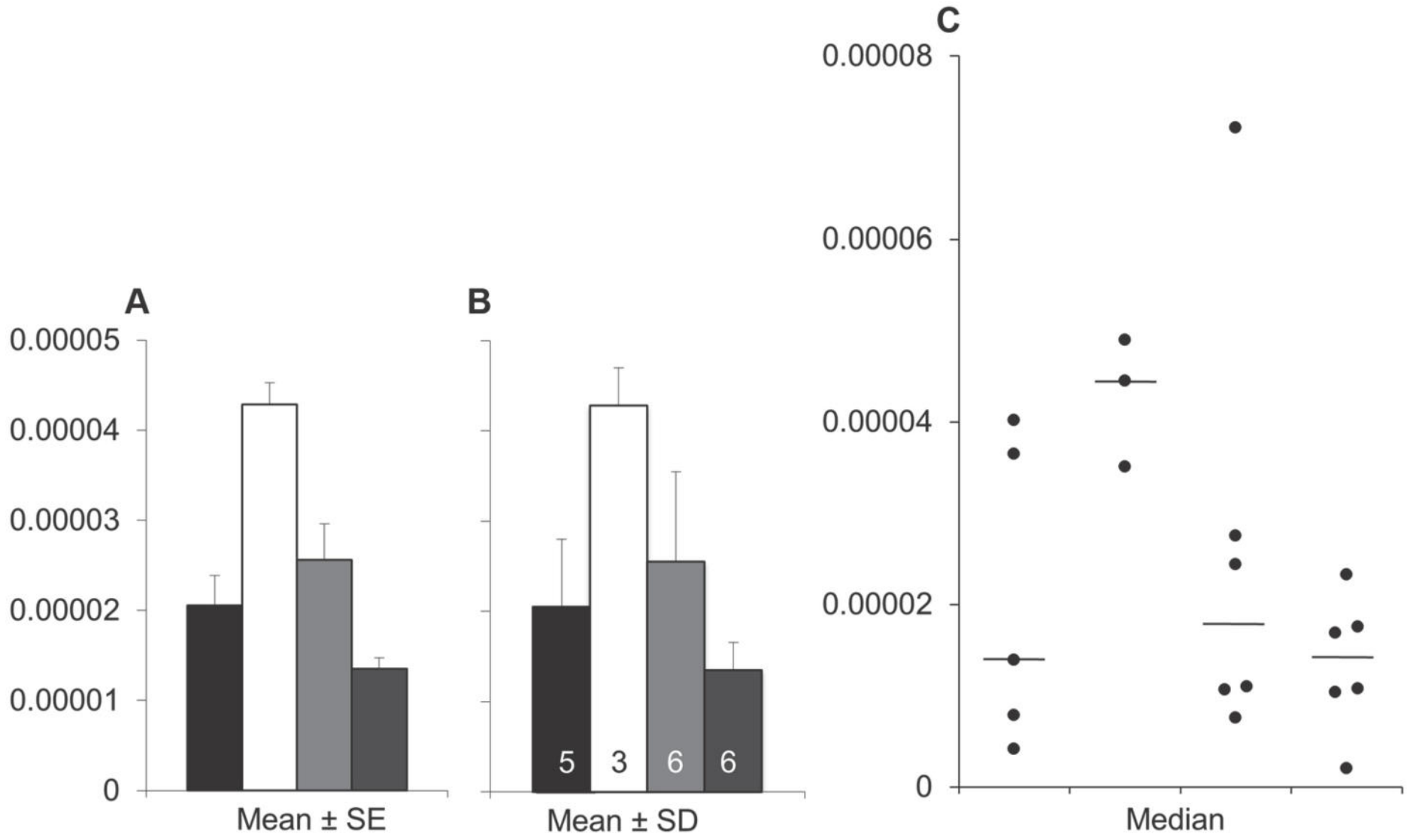
1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America, **2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States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

* weissgerber.tracey@mayo.edu

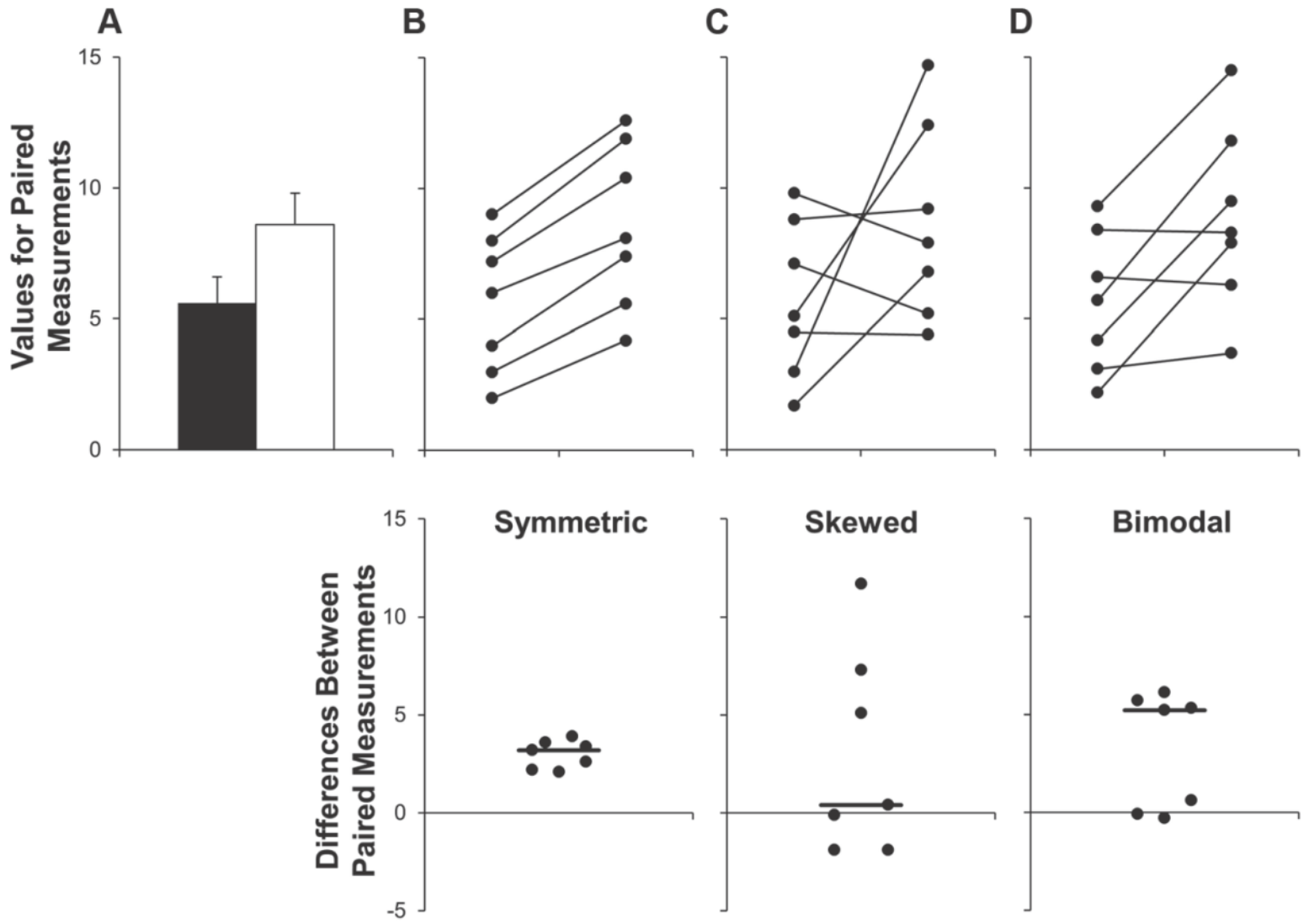


Test		p value		
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Weissgerber et al. 2015 PLoS Biology

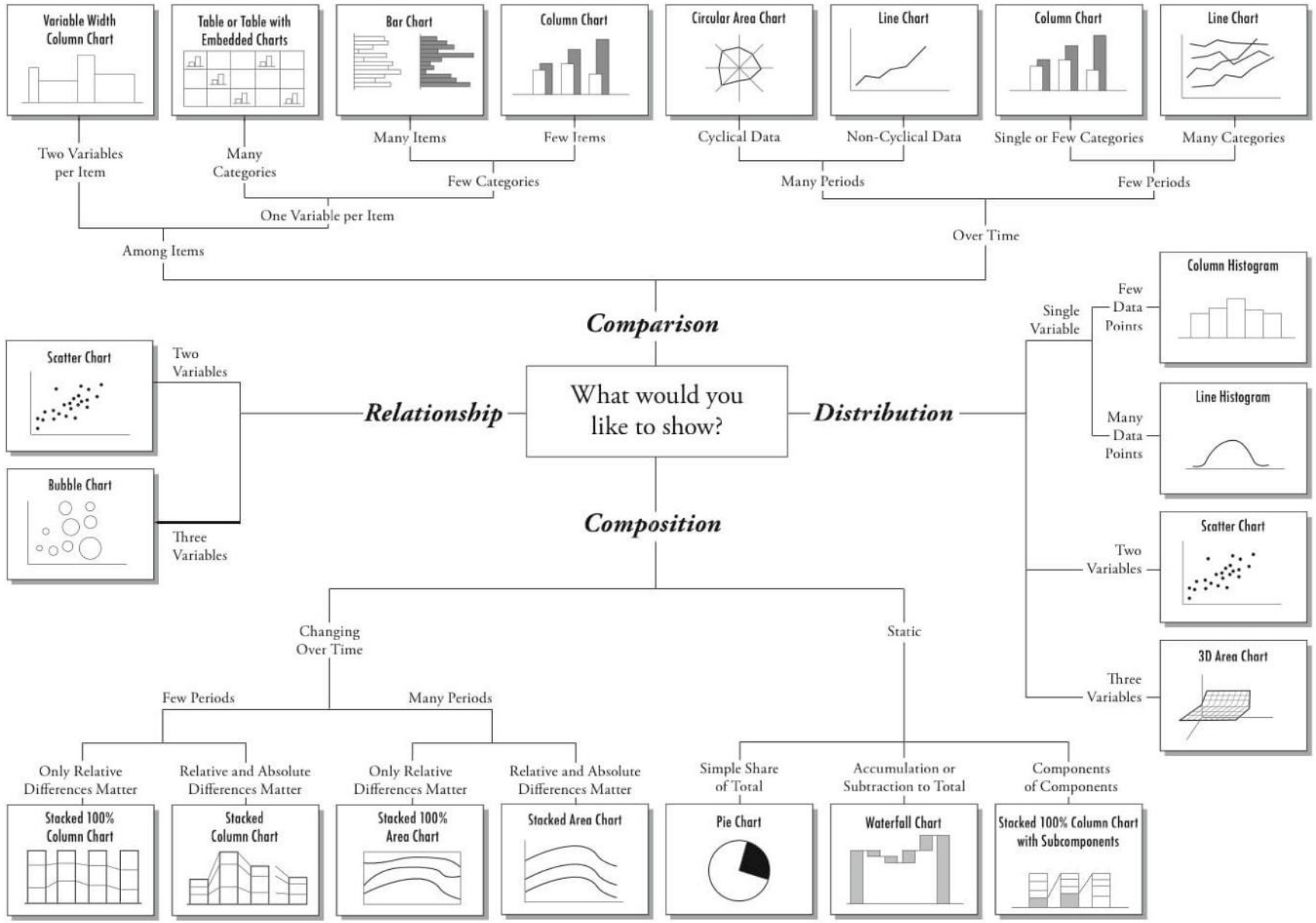


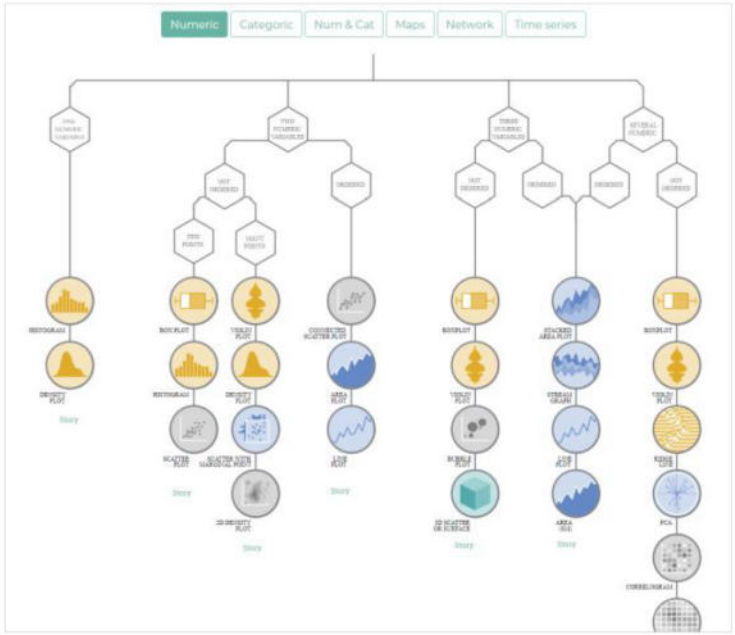
Weissgerber et al. 2015 PLoS Biology



Weissgerber et al. 2015 PLoS Biology

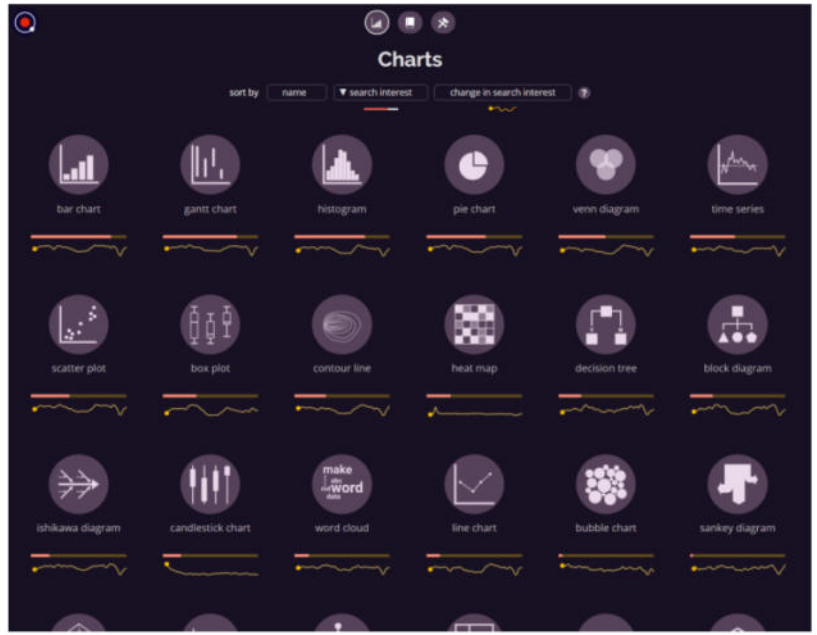
Chart Suggestions—A Thought-Starter





data-to-viz.com

datavizproject.com



visualizationuniverse.com/charts

from Data to Viz

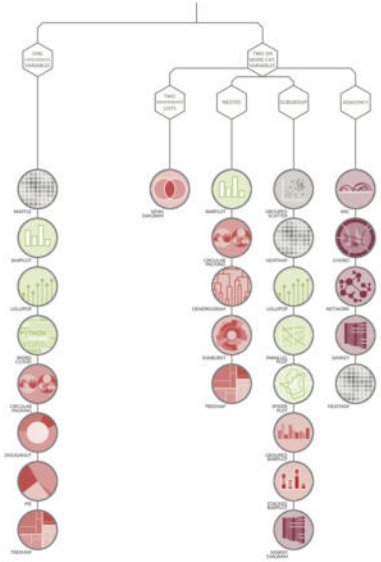
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

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

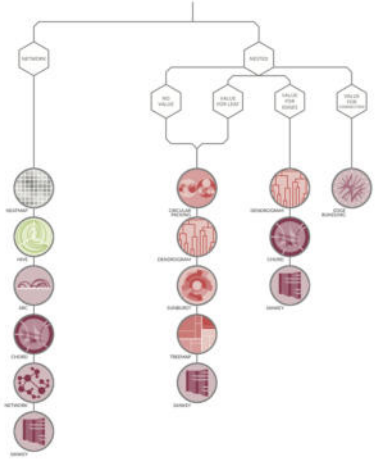
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

CATEGORIC

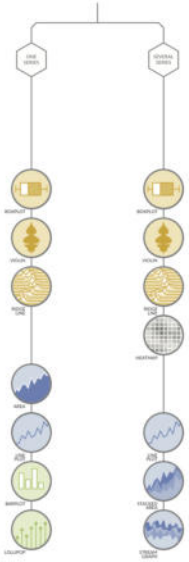


RELATIONAL



MAP

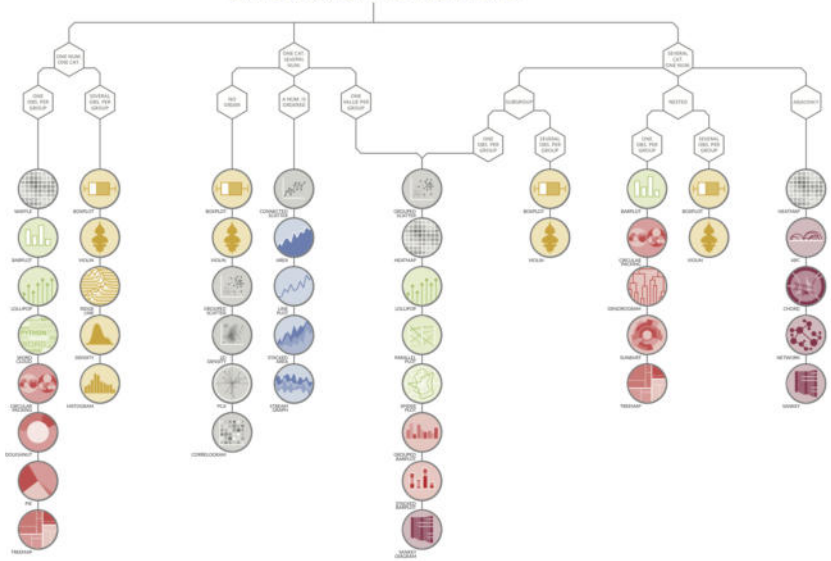
TIME SERIES



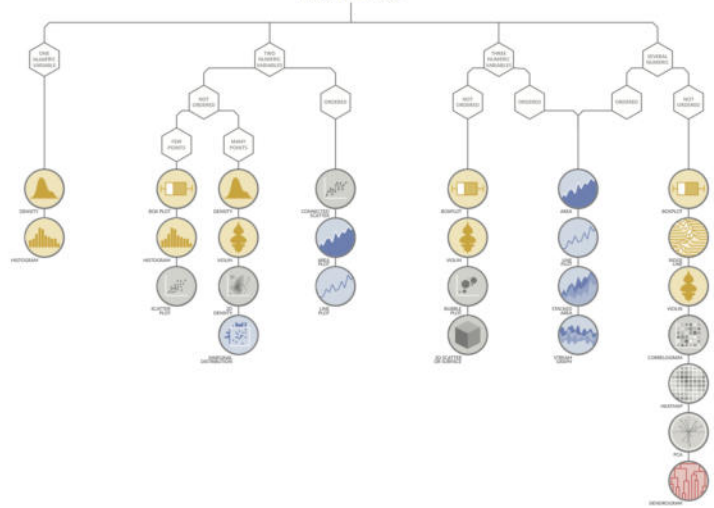
WHAT DO YOU WANT TO SHOW ?

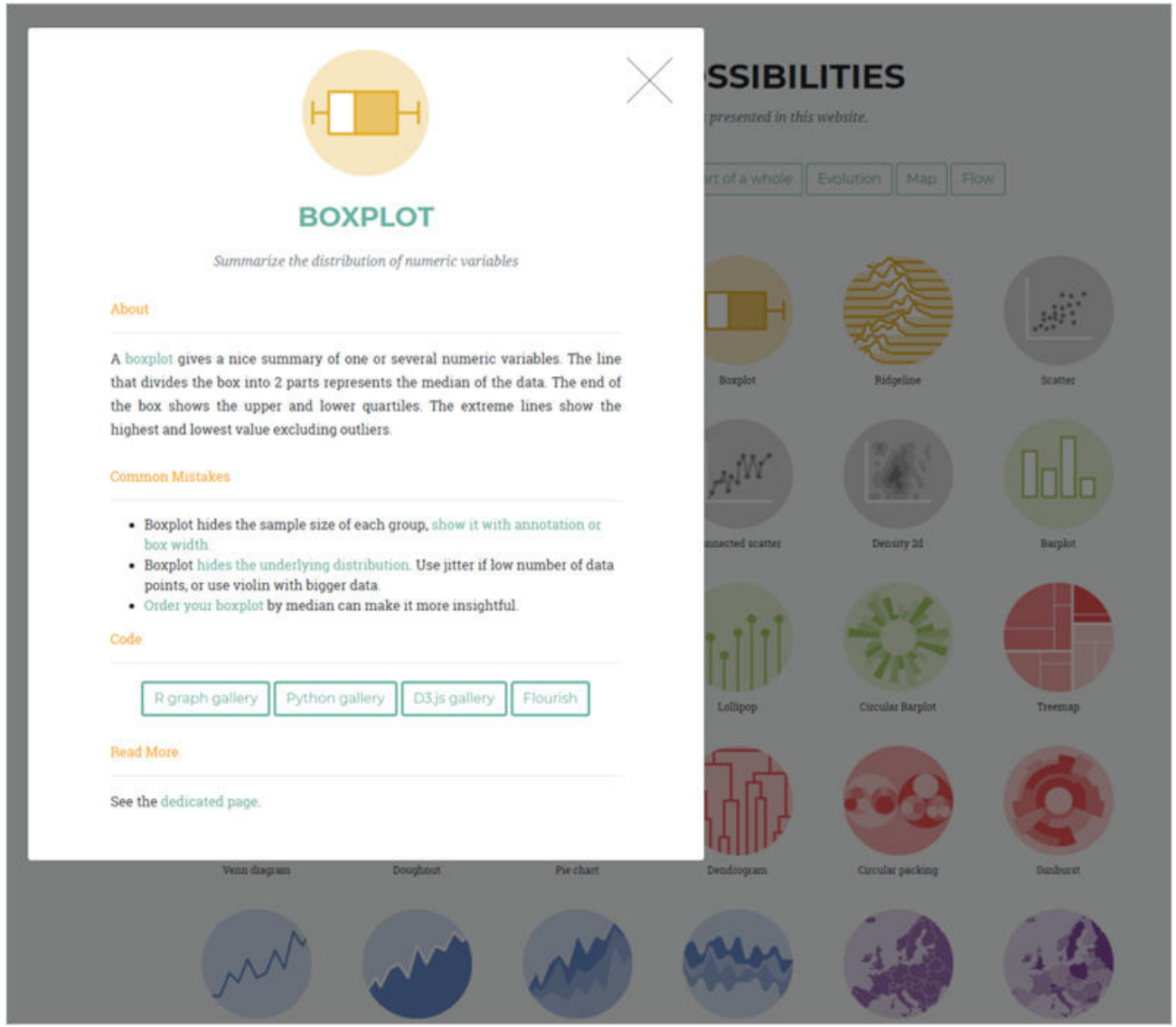
- Distribution
- Correlation
- Ranking
- Part of a whole
- Evolution
- Maps
- Flow

CATEGORIC AND NUMERIC



NUMERIC

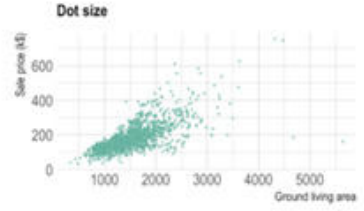


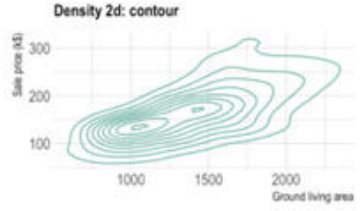

BOXPLOT
Summarize the distribution of numeric variables
About
A boxplot gives a nice summary of one or several numeric variables. The line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers.
Common Mistakes
Boxplot hides the sample size of each group, show it with annotation or box width.
Boxplot hides the underlying distribution. Use jitter if low number of data points, or use violin with bigger data.
Order your boxplot by median can make it more insightful.
Code
R graph gallery
Python gallery
D3.js gallery
Flourish
Read More
See the dedicated page.
SSIBILITIES
presented in this website.
rt of a whole
Evolution
Map
Flow
Boxplot
Ridgeline
Scatter
Density 2d
Barplot
Lollipop
Circular Barplot
Treemap
Dendrogram
Circular packing
Sunburst
Venn diagram
Doughnut
Pie chart

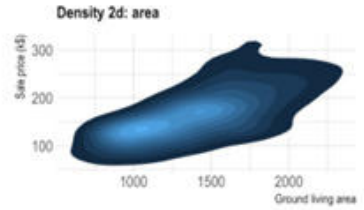
Overplotting

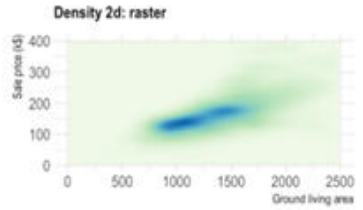
The most common pitfall with scatterplot is overplotting: when the sample size gets big, dots are plotted on top of each other what makes the chart unreadable. There are several work around to avoid this issue as describe in this specific post. Here is a summary of the different offered techniques:

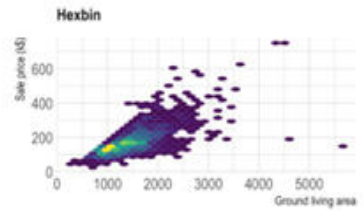
CODE

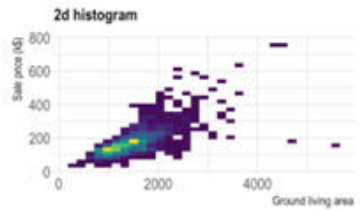








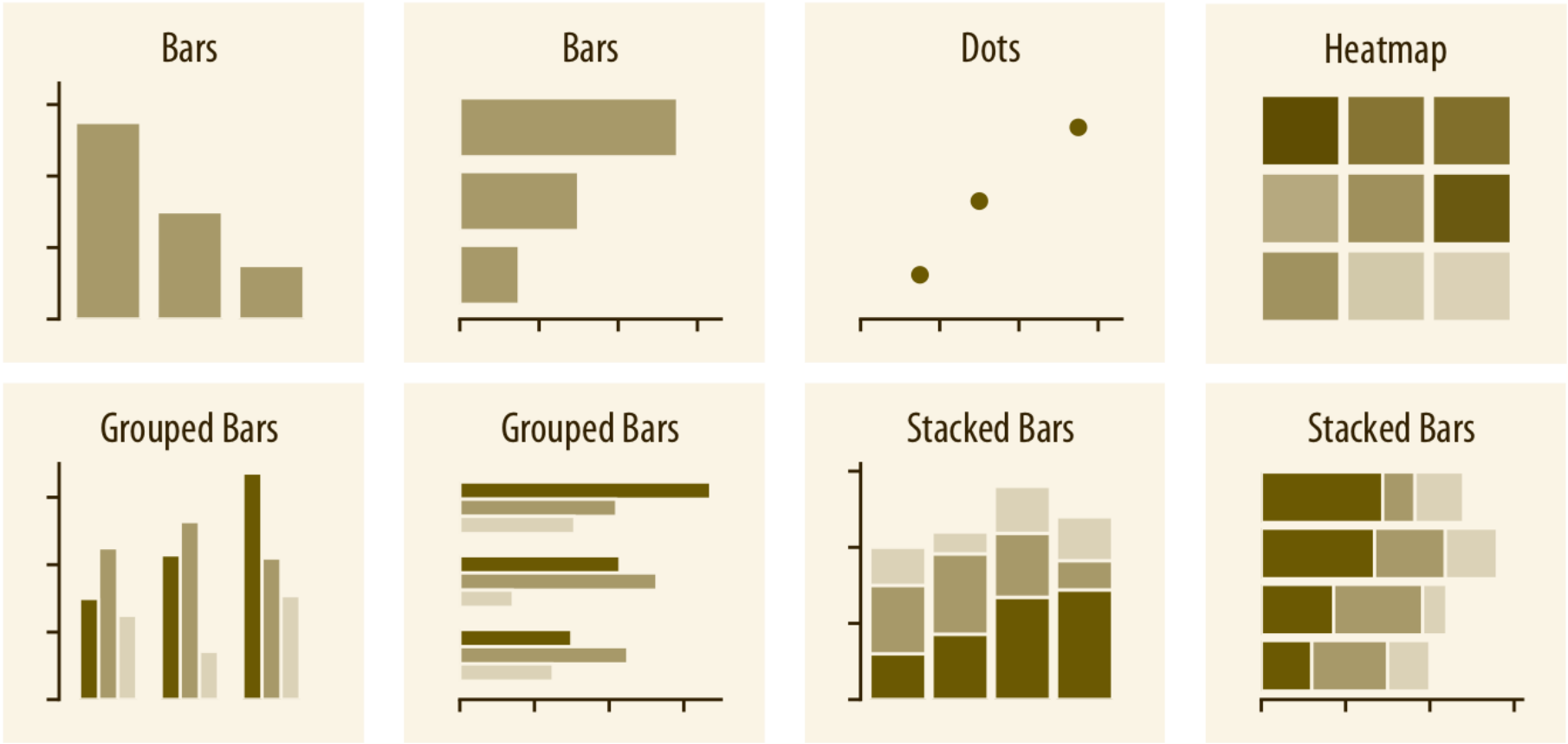




Going further

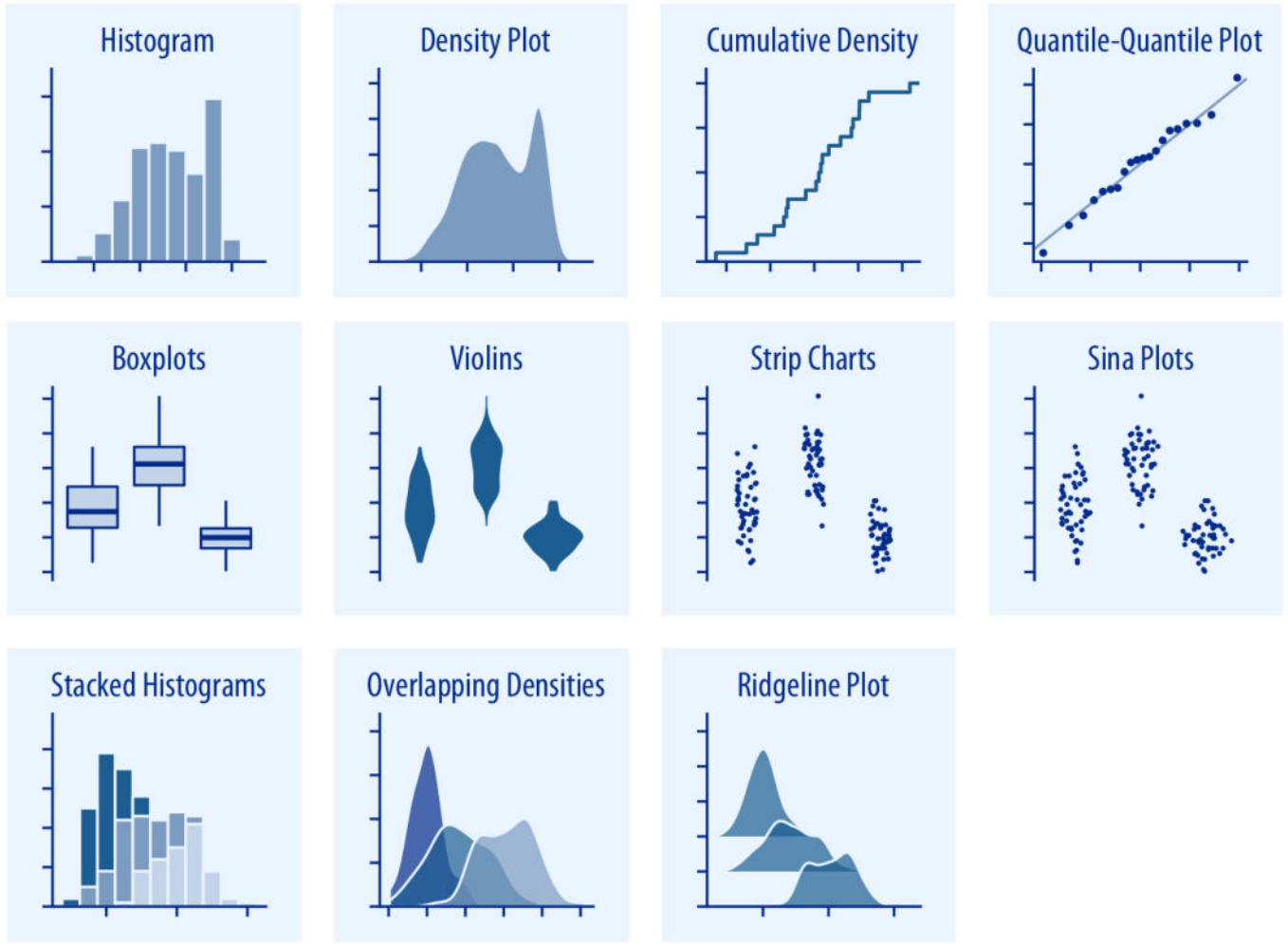
You can learn more about each type of graphic presented in this story in the dedicated

Charts to Visualize Amounts



"Fundamentals of Data Visualization" by Claus Wilke

Charts to Visualize Distributions

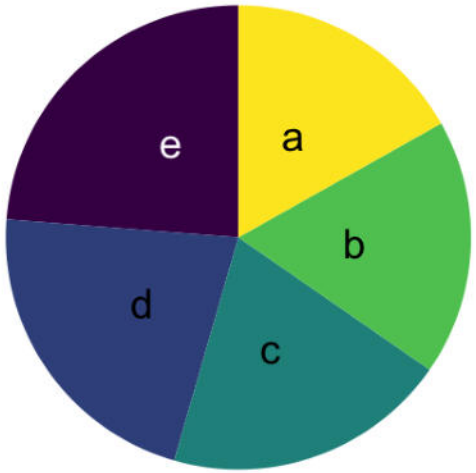


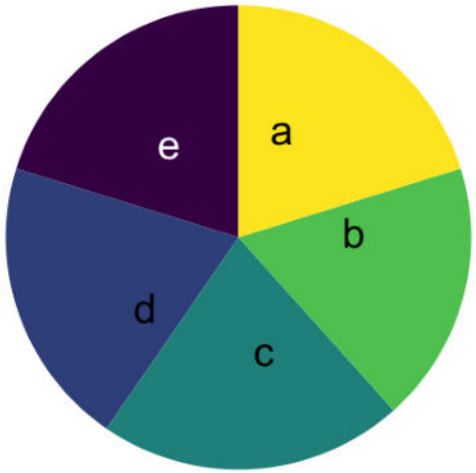
"Fundamentals of Data Visualization" by Claus Wilke

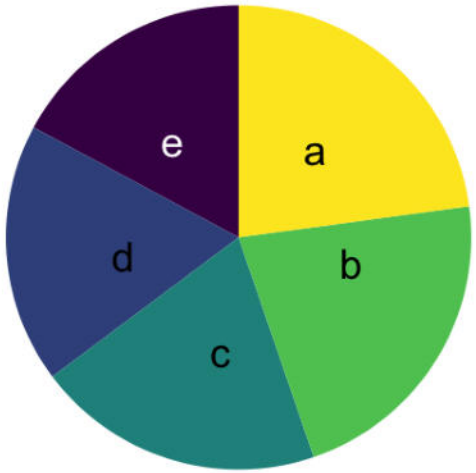
Charts to Visualize Proportions

"Fundamentals of Data Visualization" by Claus Wilke

Beware of Pie Charts

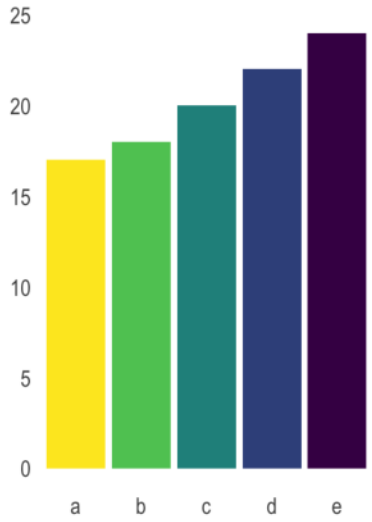


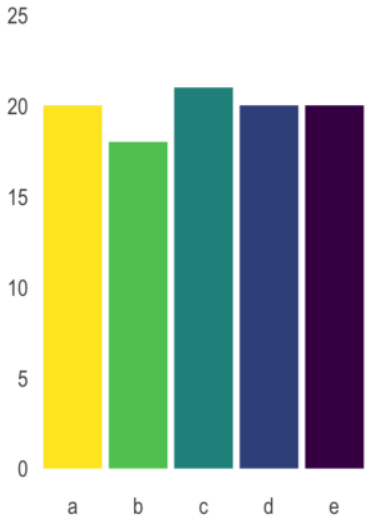


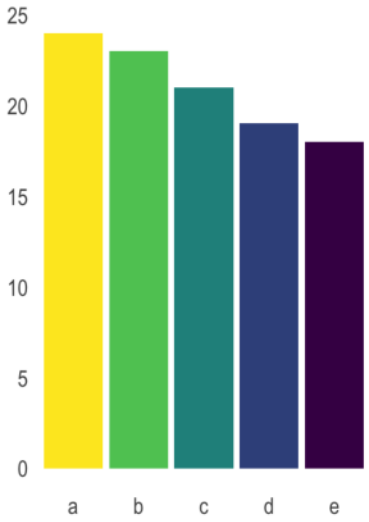


From Data to Viz

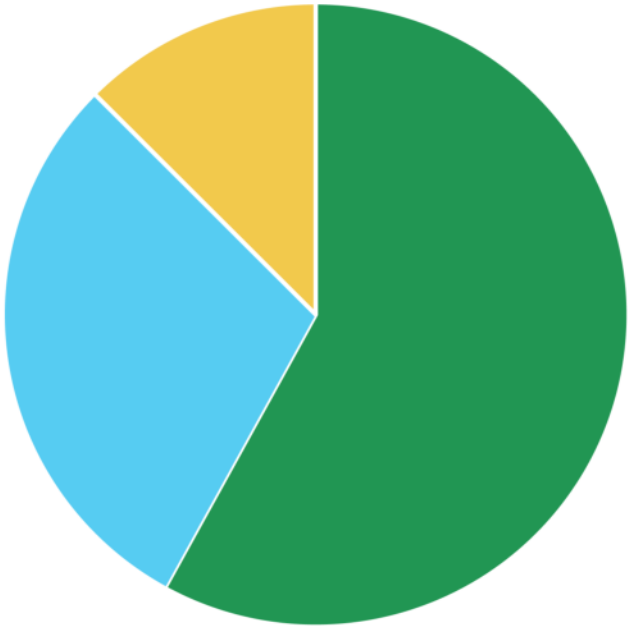
Beware of Pie Charts

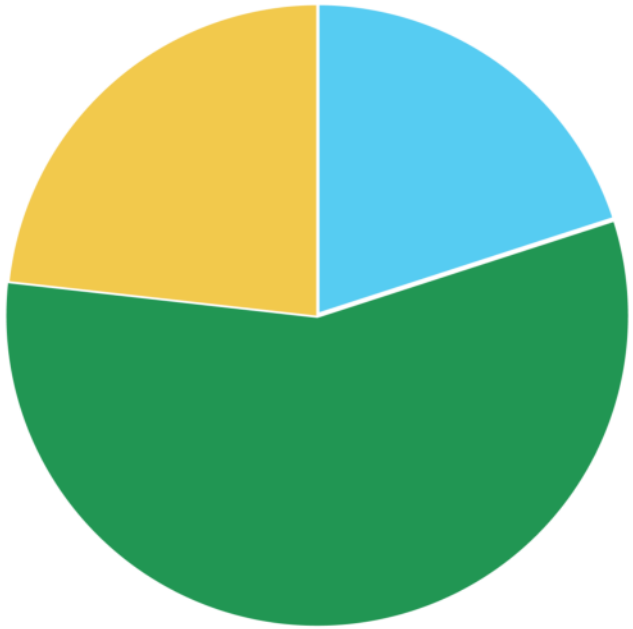






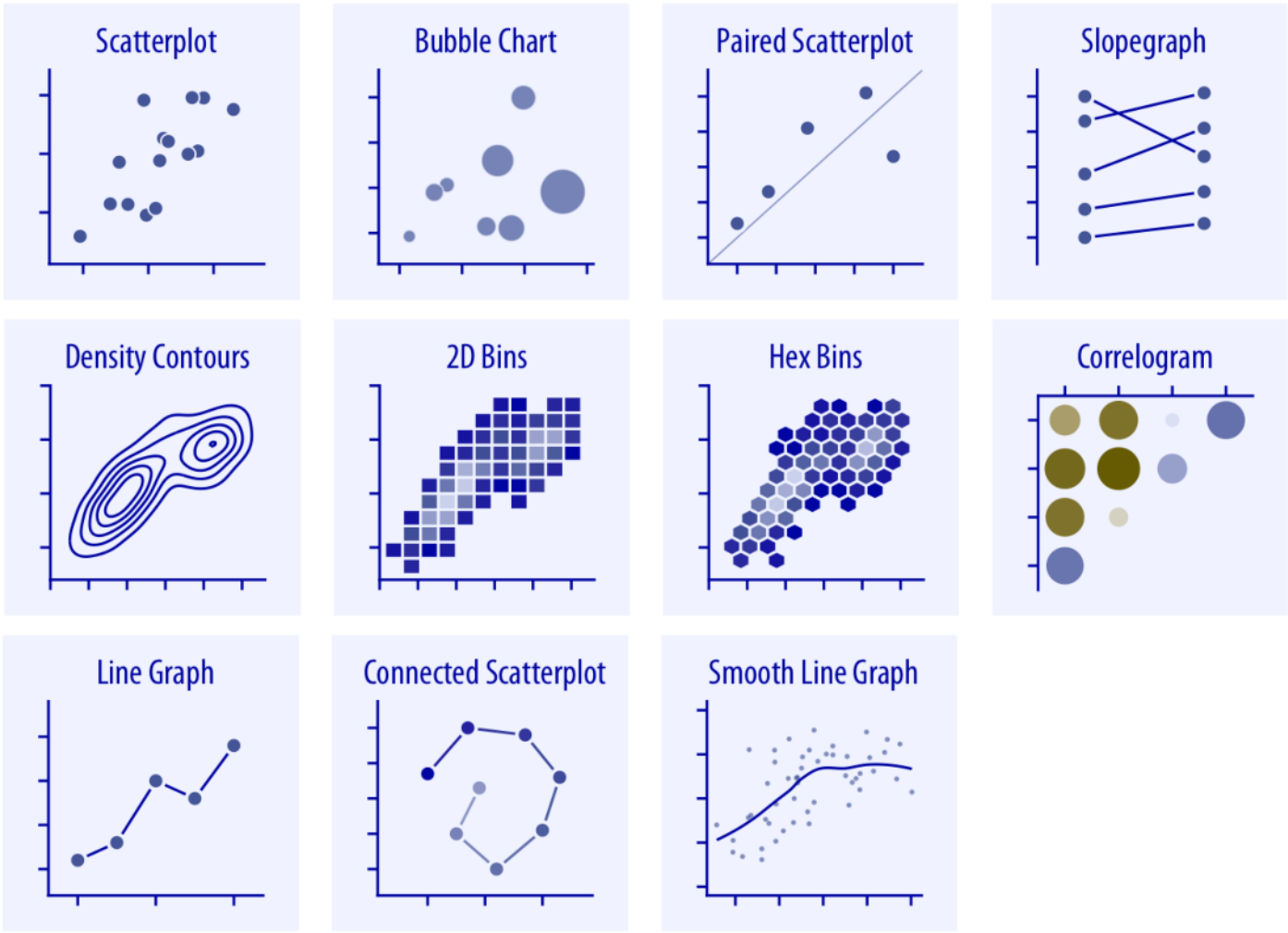
Beware of Pie Charts



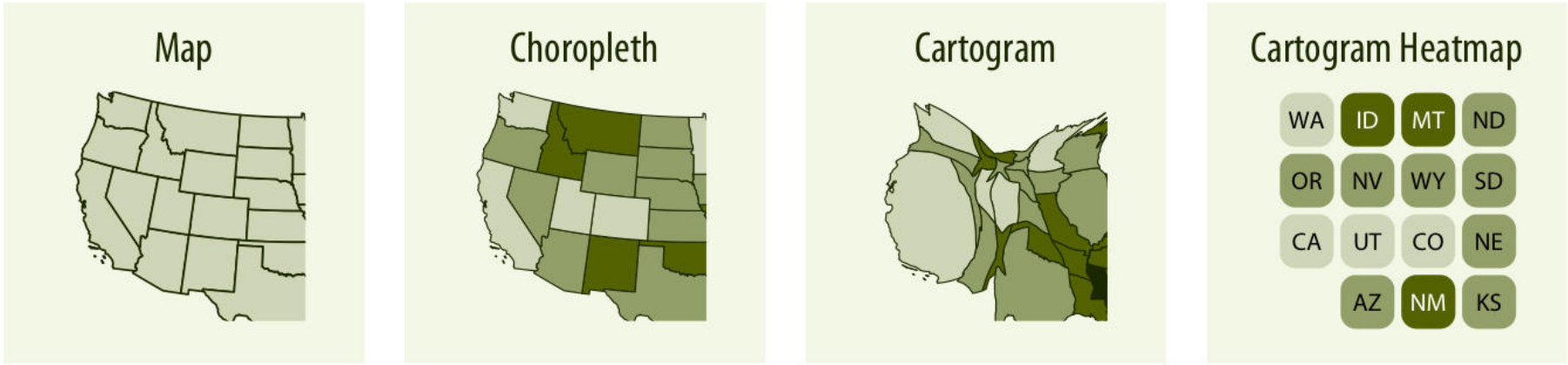


"Hands-On Data Visualization" by Jack Dougherty & Ilya Ilyankou

Charts to Visualize x–y Relationships

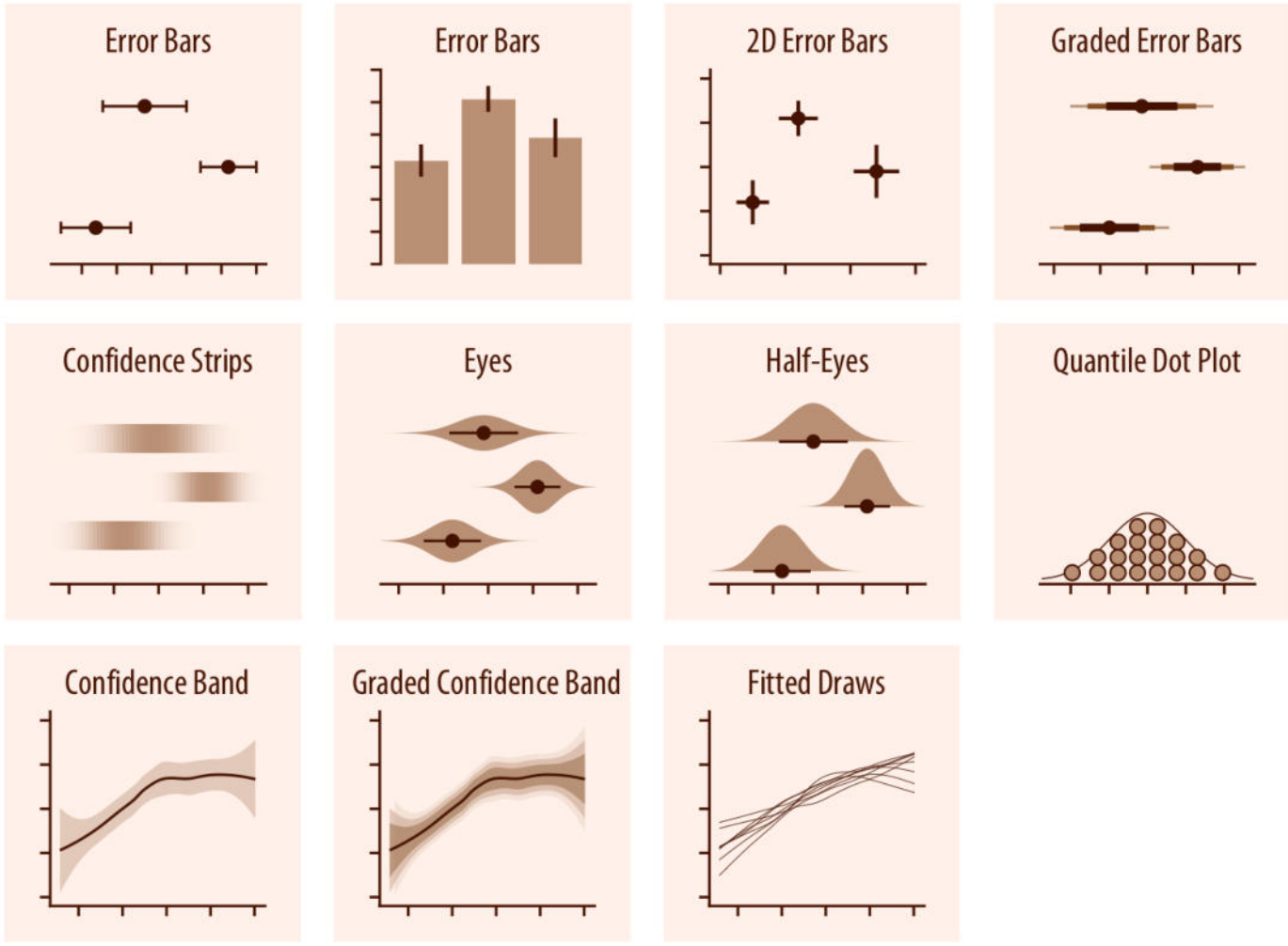


Charts to Visualize Geospatial Data



"Fundamentals of Data Visualization" by Claus Wilke

Charts to Visualize Uncertainty



"Fundamentals of Data Visualization" by Claus Wilke

Visual Form

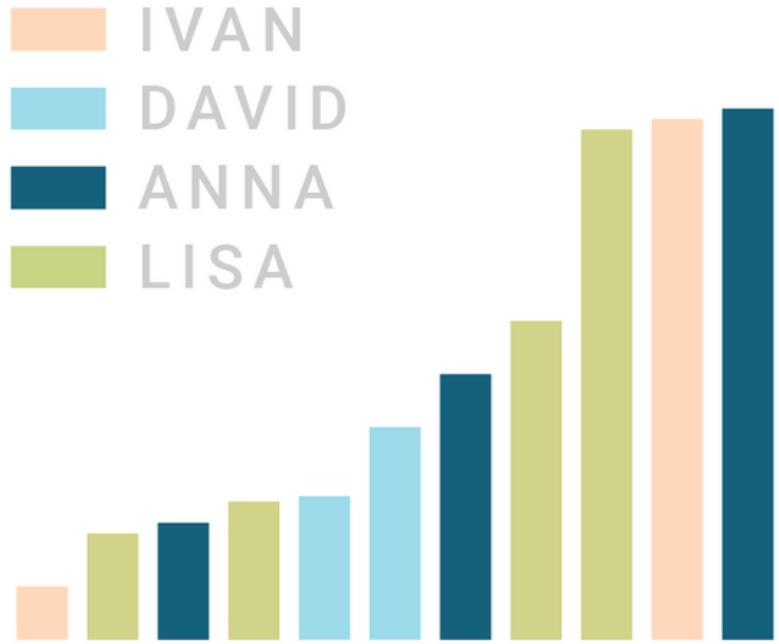
Follow design rules and data visualization principles


Colors
and Common Pitfalls

Color Choice & Accessibility

NOT IDEAL

BETTER



"What to consider when choosing colors for data visualization" by Lisa Charlotte Muth/DataWrapper

Color Palette Types

Categorical

default_uchicago

Sequential: Single-Hue

cyan_material

Sequential: Multi-Hue

Sunset

Diverging

Geyser

Cyclical

ocean.phase

Color Choice & Accessibility

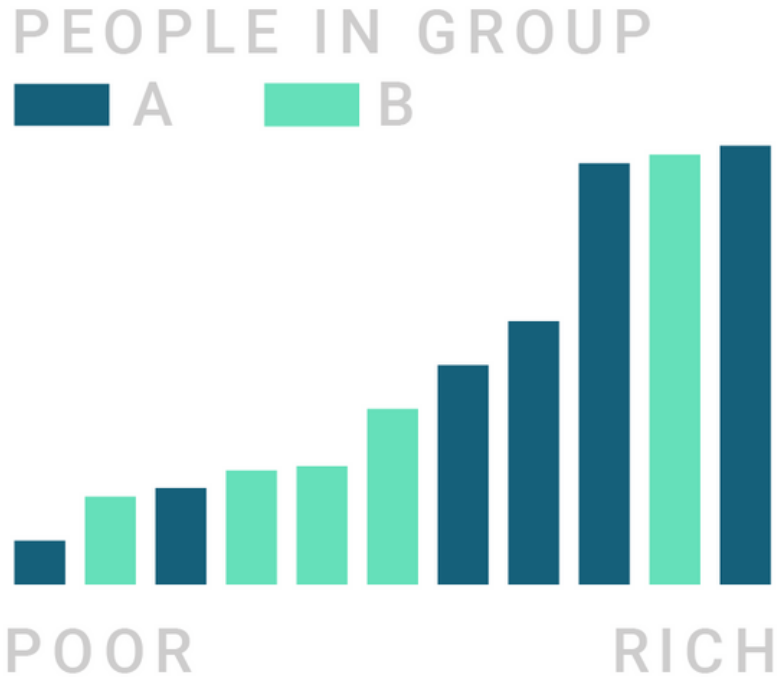
NOT IDEAL

POOR RICH

PEOPLE IN GROUP A

PEOPLE IN GROUP B

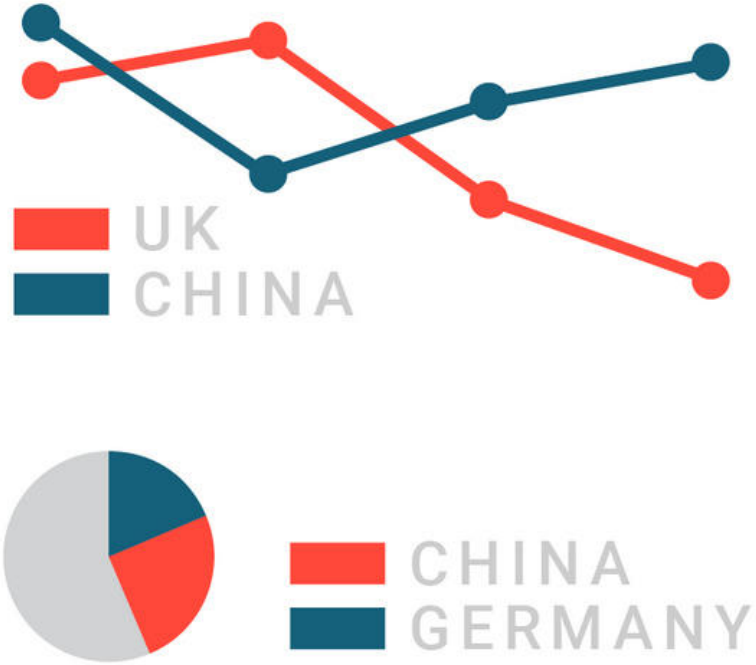
BETTER



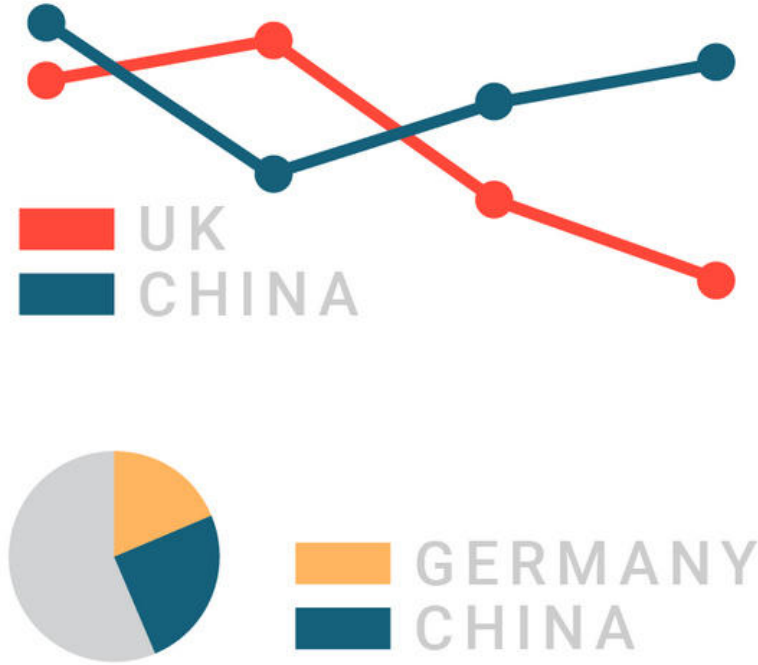
"What to consider when choosing colors for data visualization" by Lisa Charlotte Muth/DataWrapper

Color Choice & Accessibility

NOT IDEAL



BETTER



"What to consider when choosing colors for data visualization" by Lisa Charlotte Muth/DataWrapper

Journals & Magazines > IEEE Computer Graphics and Ap... > Volume: 27 Issue: 2

Rainbow Color Map (Still) Considered Harmful

Publisher: IEEE

2 Author(s) David Borland ; Russell M. Taylor Ii **View All Authors**

172	3	9091
Paper Citations	Patent Citations	Full Text Views

Medical Physics

Current Issue | Authors | Submissions | Advertise | Search

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

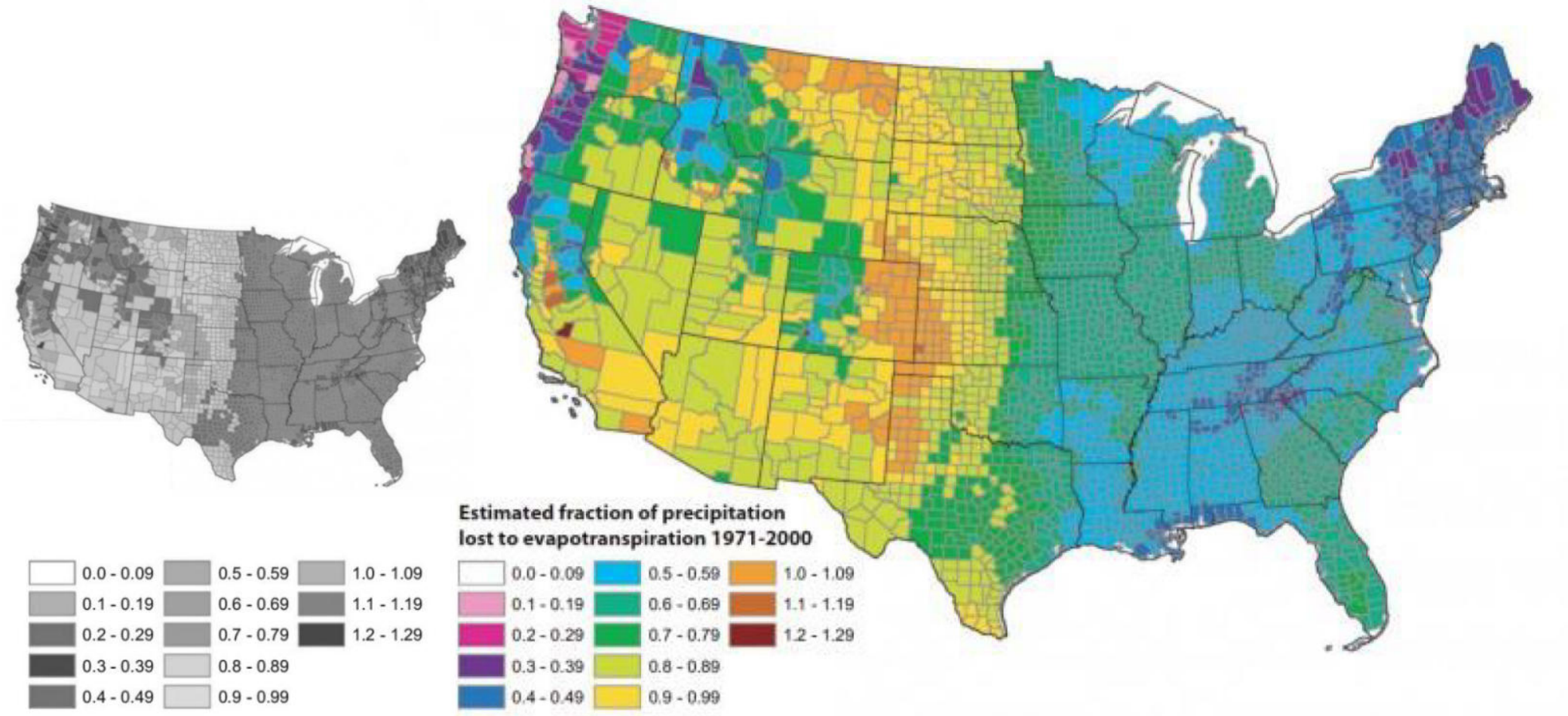
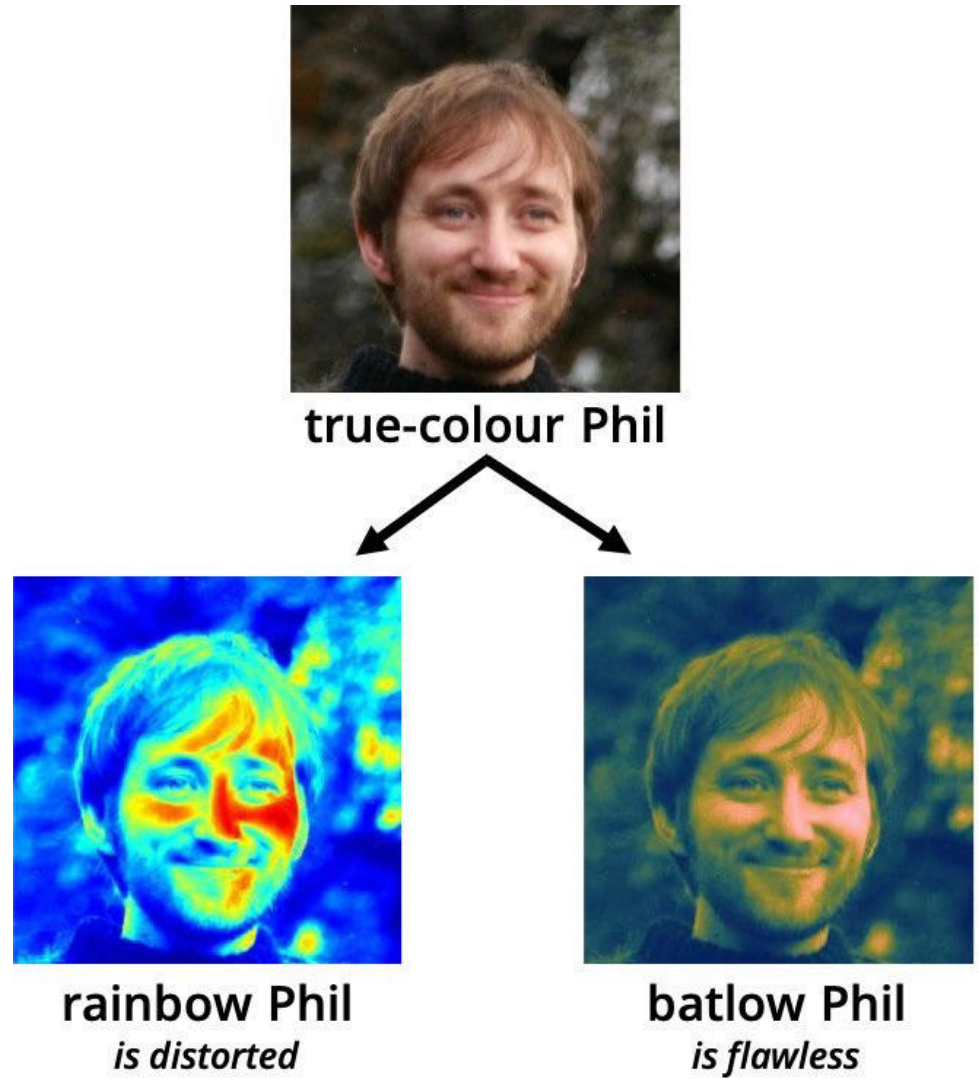


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

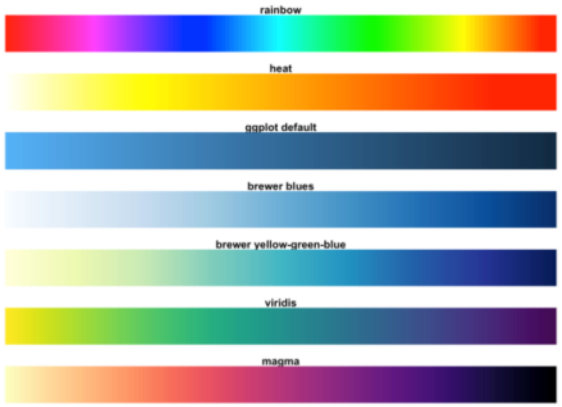
Modified from eagereyes.org/basicss/rainbow-color-map


true-colour Phil
rainbow Phil
is distorted
batlow Phil
is flawless

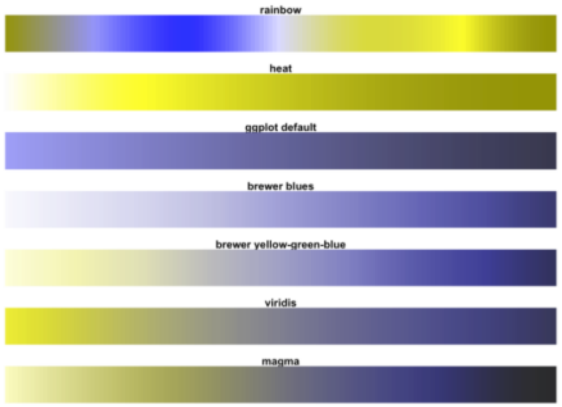
Color Choice & Accessibility

Vignette {viridis} package

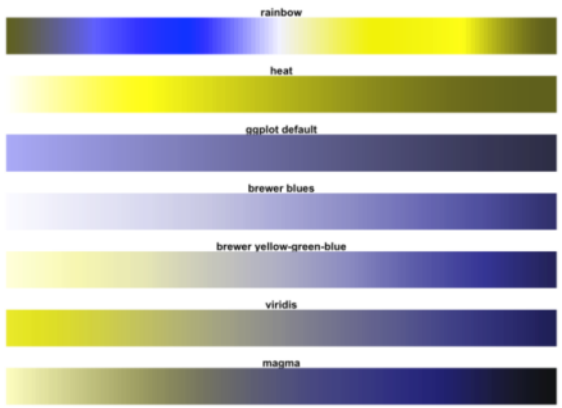
Color Choice & Accessibility



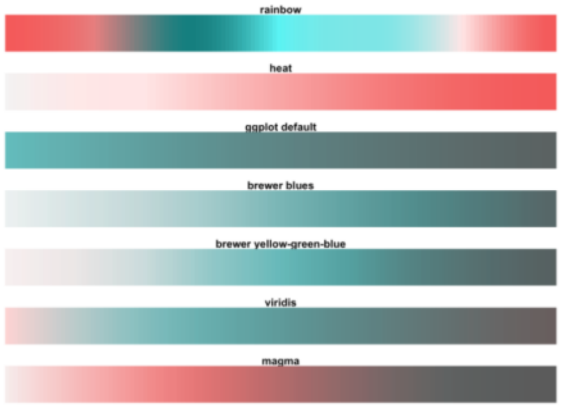
Deutranopia: present in 6% of males



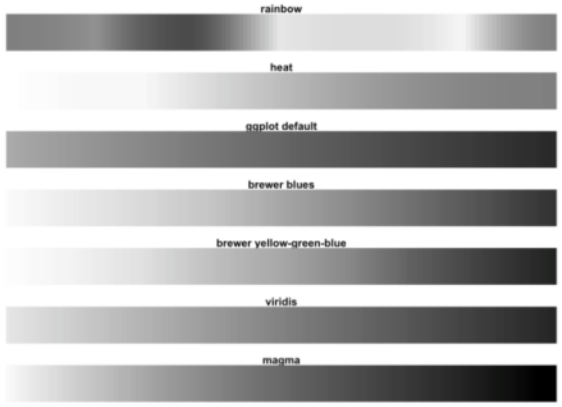
Protanopia: present in 1% of males



Tritanopia: present in 0.008% of humans

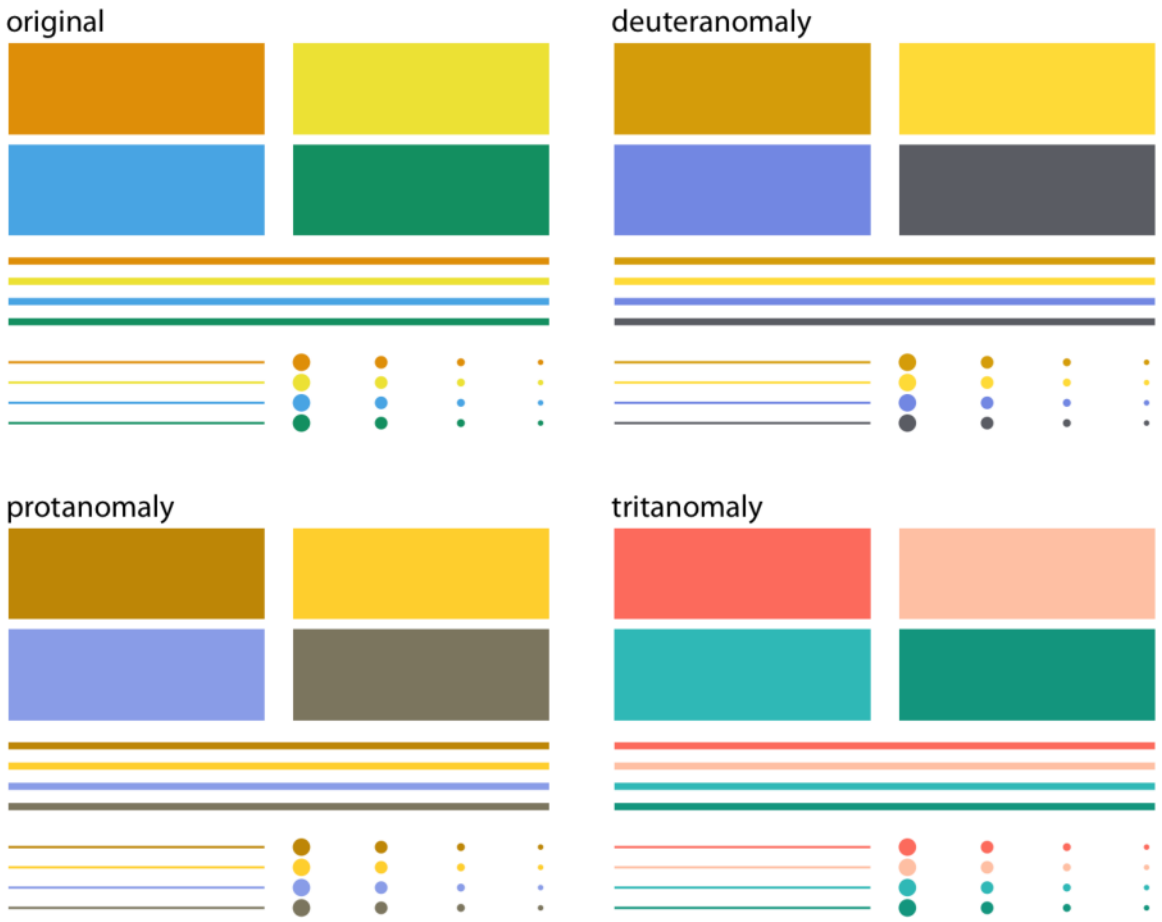


Monochromacy: present in 0.001% of humans



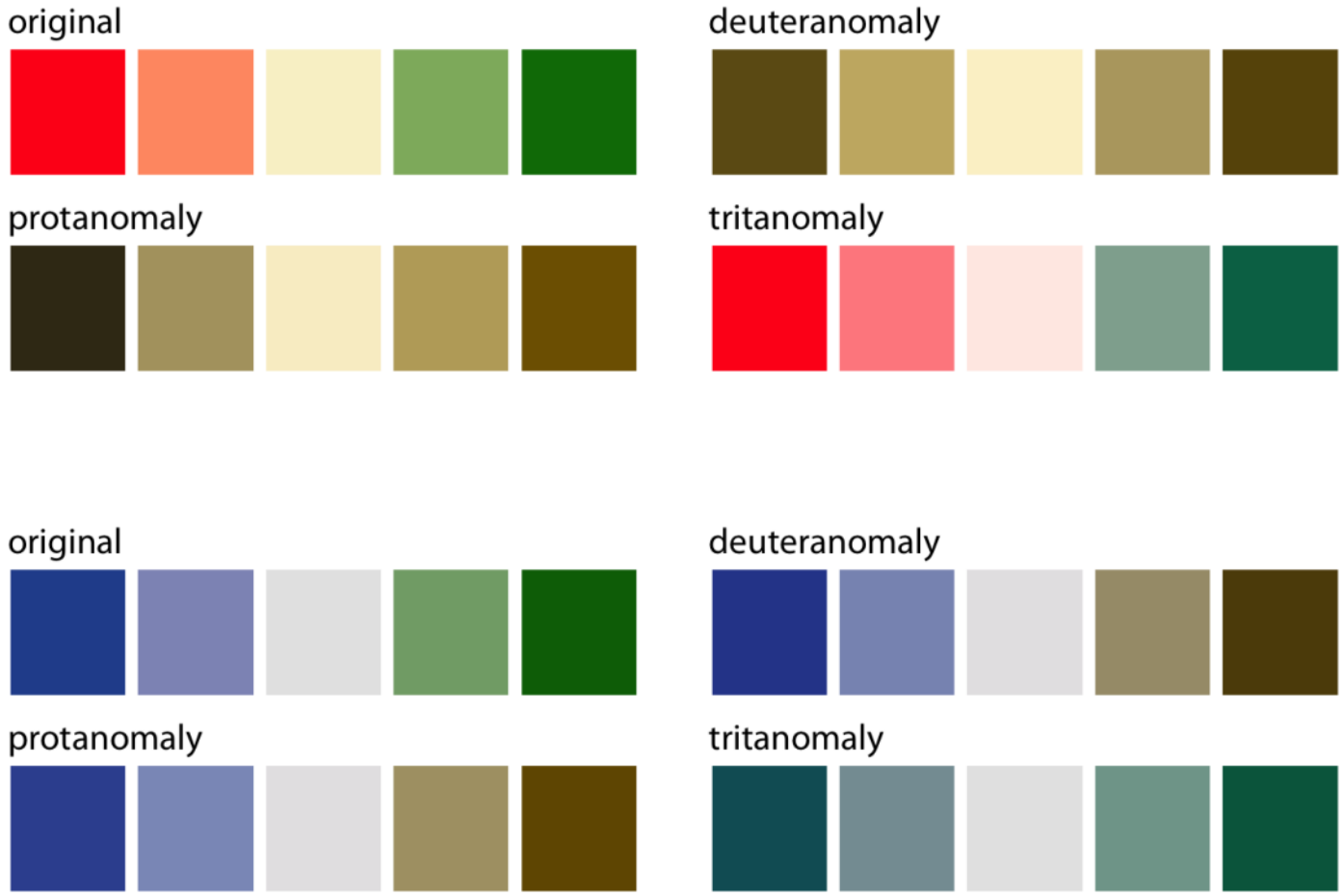
... and present in ~75% of university printers! ☺

Choice of the Color Palette & Accessibility



"Fundamentals of Data Visualization" by Claus Wilke

Choice of the Color Palette & Accessibility



"Fundamentals of Data Visualization" by Claus Wilke

To make sure your visualizations work for people with CVD don't just rely on provided color palettes.

Instead, test your figures in a color-blindness simulator!

VIZ PALETTE

By: Elijah Meeks & Susie Lu

PICK

Use Chroma.js

Use Colorgorical

Use ColorBrewer

Add | Replace

EDIT

7 Colors

Add

#hex rgb hsl

1 #ffd700
2 #ffb14e
3 #fa8775
4 #ea5f94
5 #cd34b5
6 #9d02d7
7 #0000ff

GET

String quotes
Object with metadata

#hex rgb hsl

```
["#ffd700",
"#ffb14e",
"#fa8775",
"#ea5f94",
"#cd34b5",
"#9d02d7",
"#0000ff"]
```

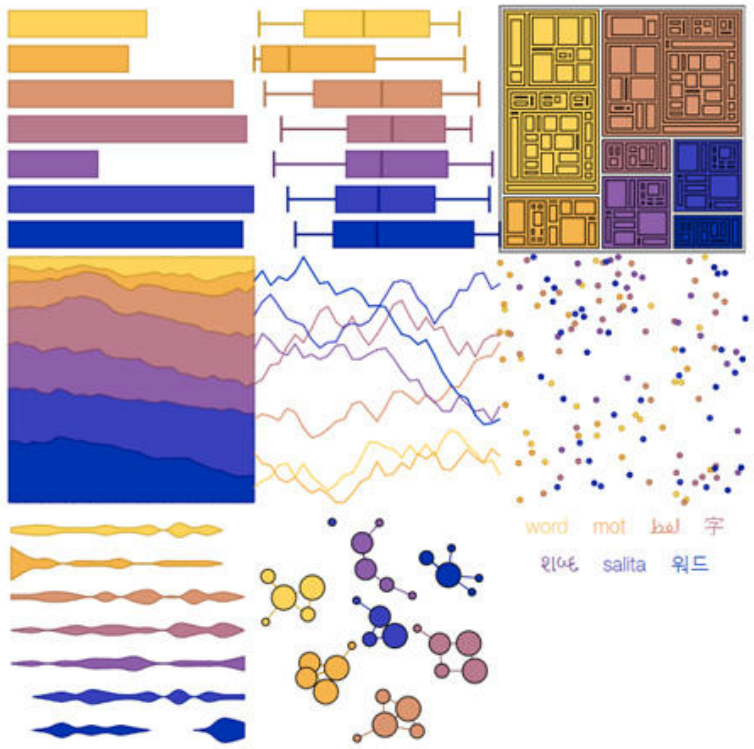
COLORS IN ACTION

Background color: #ffffff
Font color: #000000

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



ourworldindata.org/covid-vaccination-global-projections

Which countries are on track to have vaccinated 40% of their population with at least 1 dose by the end of 2021?

Based on our international COVID-19 vaccination data, we have made projections of which countries are on track to achieve this.

This map shows which countries have already surpassed this 40% target; those that are on track to meet it by the end of 2021 based on recent vaccination rates; and those that are not on track.

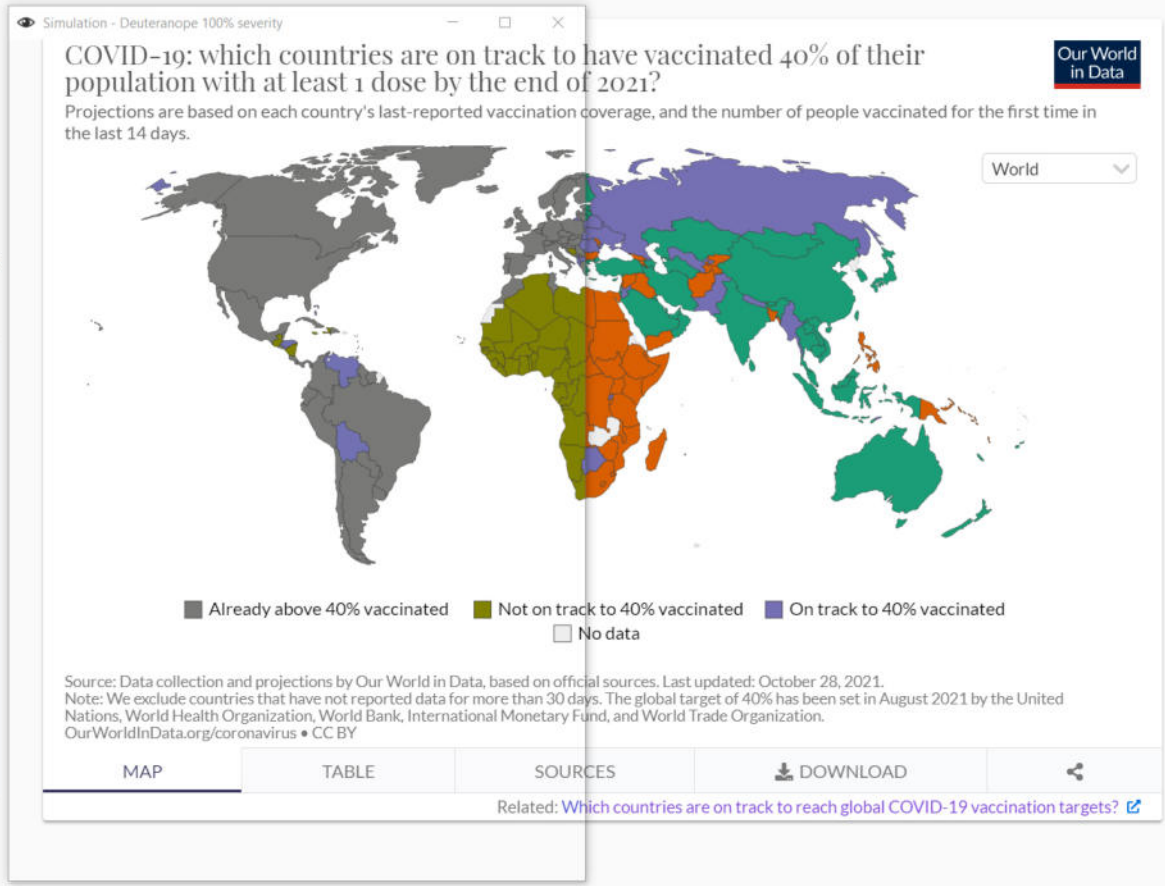
We will update these projections weekly to continue to monitor global progress towards this goal.

How does Our World in Data calculate these projections?

For each country, we calculate the current vaccination rate as the average number of people who received their first dose of a vaccine per day, *over the last 14 days*.

We then assume that this recent vaccination rate remains constant for the remainder of the year. By adding this expected share to the share of the population that have already been vaccinated, we project what share of people will have received at least one dose by January 1, 2022. We exclude from our projections countries that have not reported figures for more than 30 days.

This method means that the projections do not take into account future increases or decreases in the speed of vaccination, which could result from changes in eligibility criteria, vaccine deliveries or shortages, or new government policies.



Subscribe Feedback

Color Choice & Accessibility

Choose color-blind friendly palettes:

projects.susielu.com/viz-palette

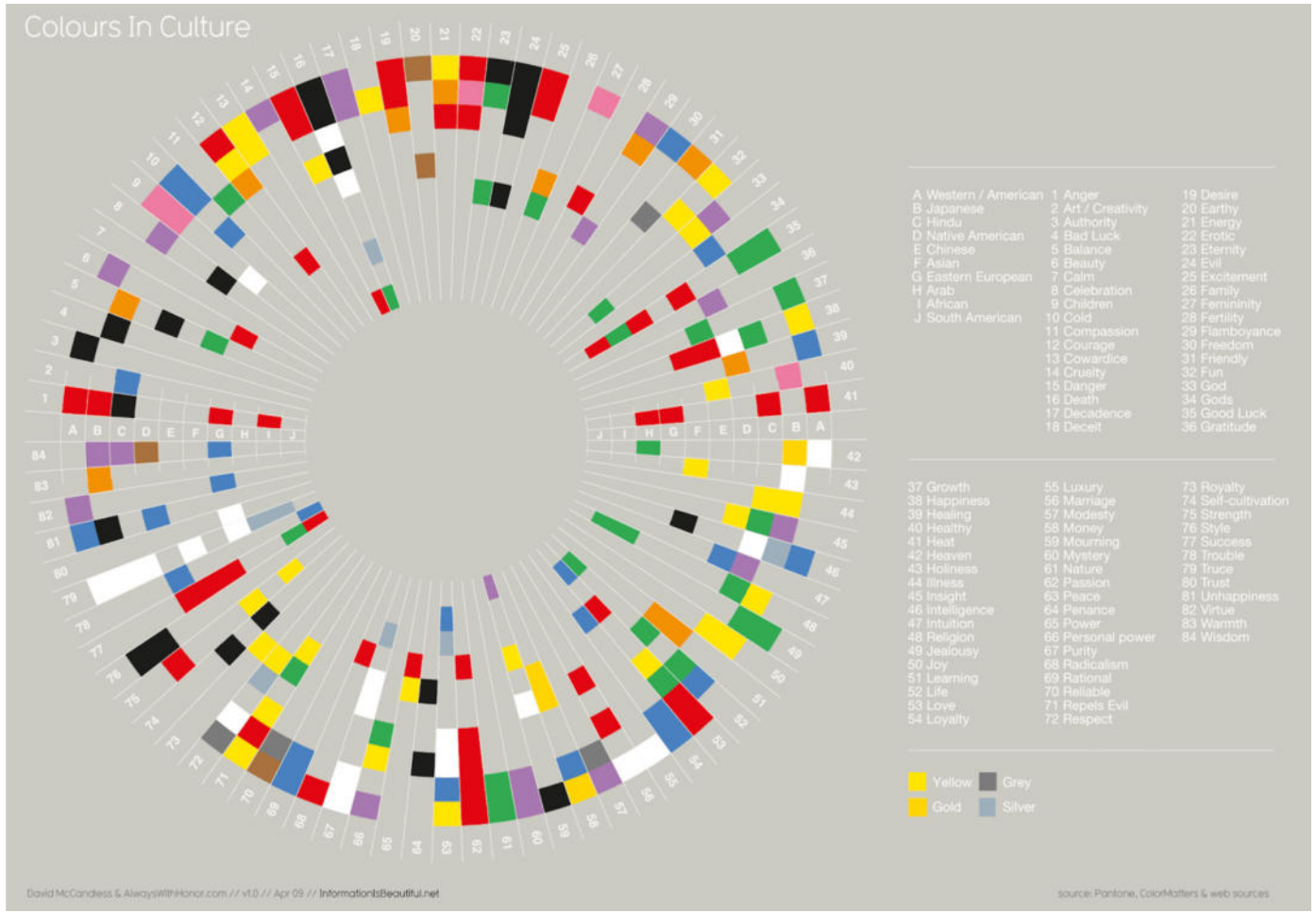
Simulate color deficient users:

coloursimulations.com

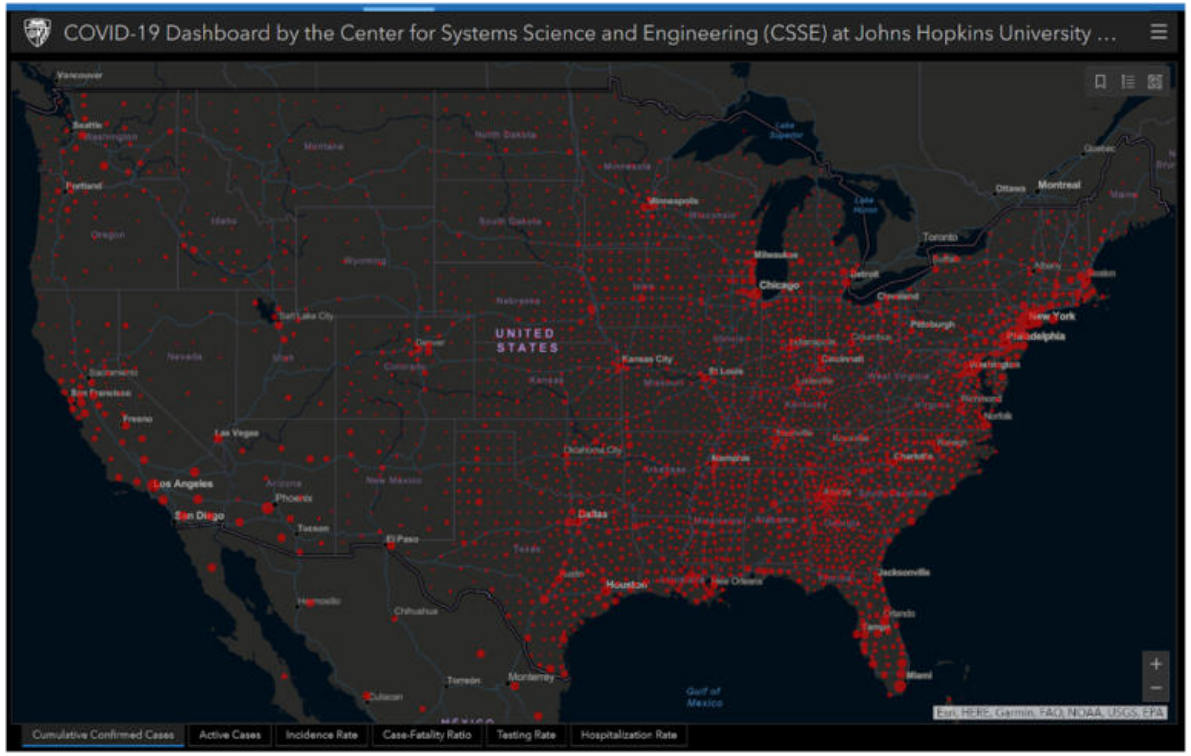
Test your final visualization:

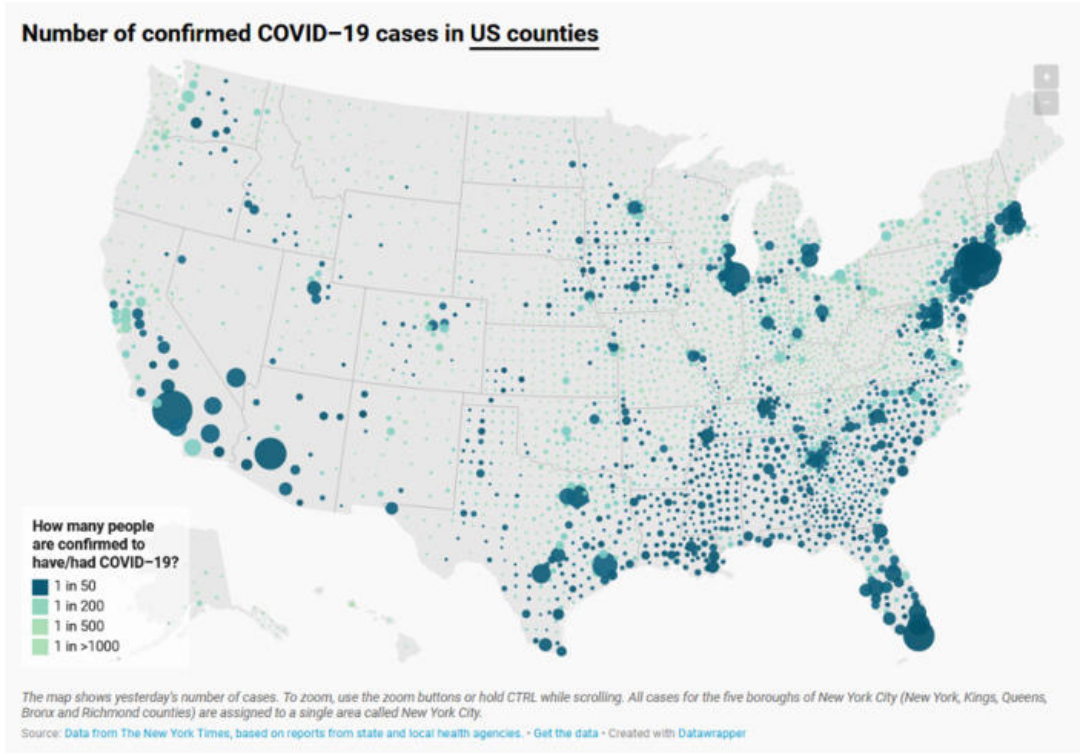
color-blindness.com/coblis-color-blindness-simulator

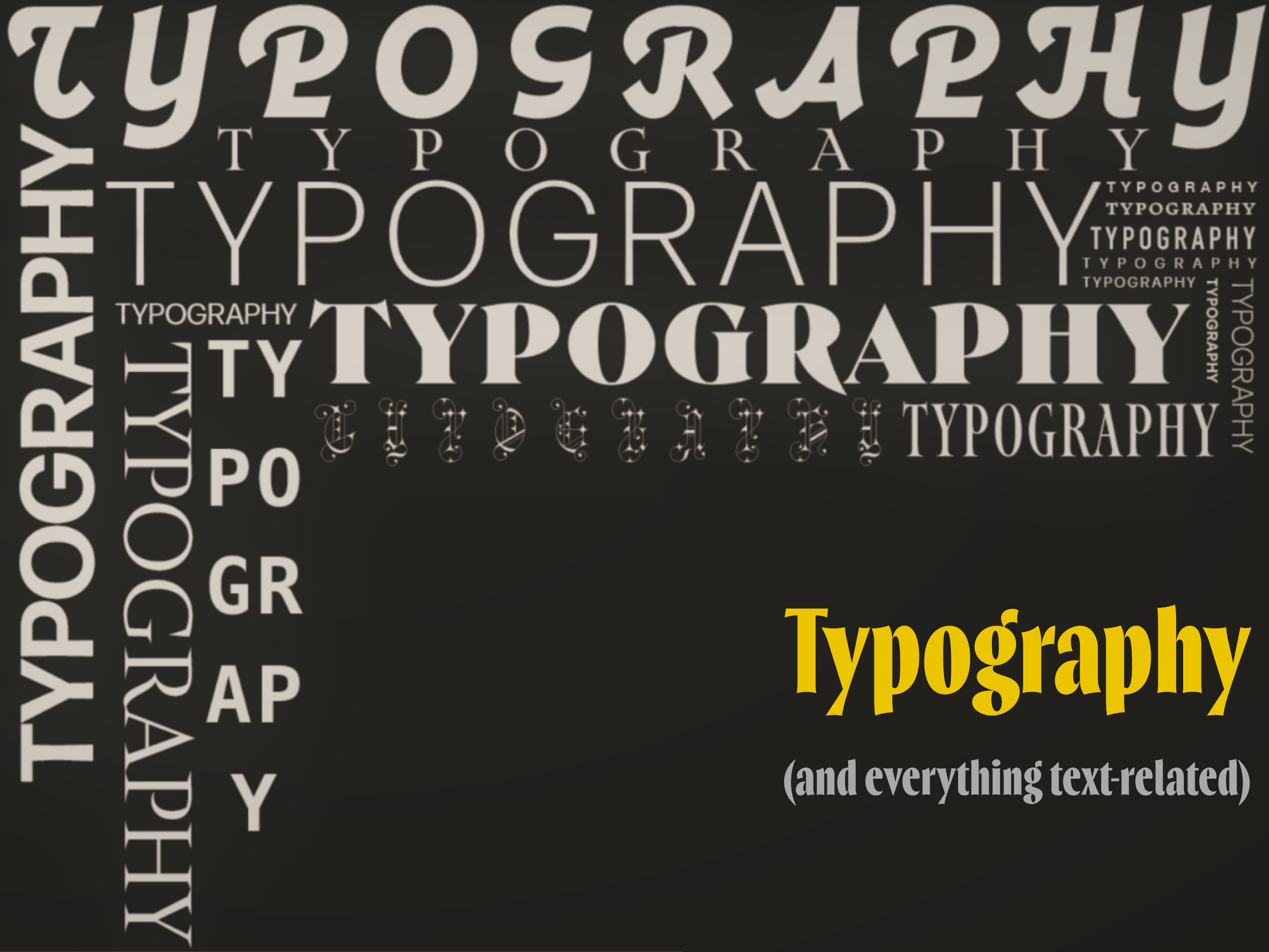
Color Choice & Emotions



Color Choice & Emotions







Typography

(and everything text-related)

I'LL BE WAITING FOR YOU!

I'll be waiting for you!

The Choice of the Font(s)

- The font(s) should fit the topic and audience - **context matters**.
- **Avoid fancy** fonts and squiggeled letters.
- Use ways to **visualize hierarchy**.
- Avoid using **ALL CAPS**.
- Use a **monospaced** font with lining for numbers.

The Choice of the Font(s)

- The font(s) should fit the topic and audience - **context matters**.
- **Avoid fancy** fonts and squiggeled letters.
- Use ways to **visualize hierarchy**.
- Avoid using **ALL CAPS**.
- Use a **monospaced** font with lining for numbers.
- **Consistency is key!**

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too.

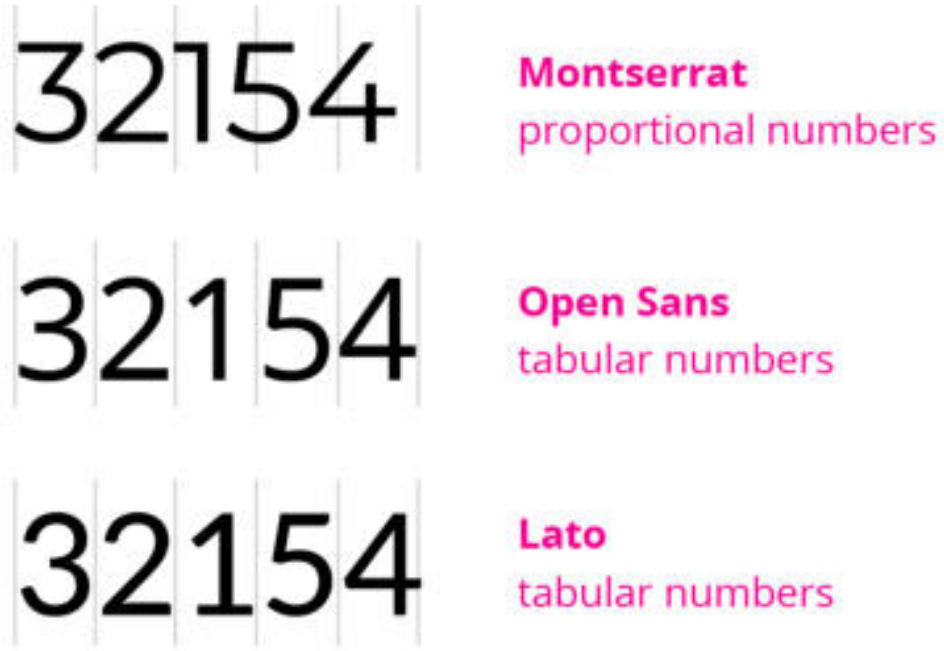
Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

The 1Il Test

1Il	Calibri
1Il	Open Sans
1Il	Roboto
1Il	Avenir Next Condensed
1Il	Lato
1Il	Oswald
1Il	Bitter
1Il	Monda
1Il	Chivo
1Il	Fira Sans
1Il	Noto Sans
1Il	Bahnschrift

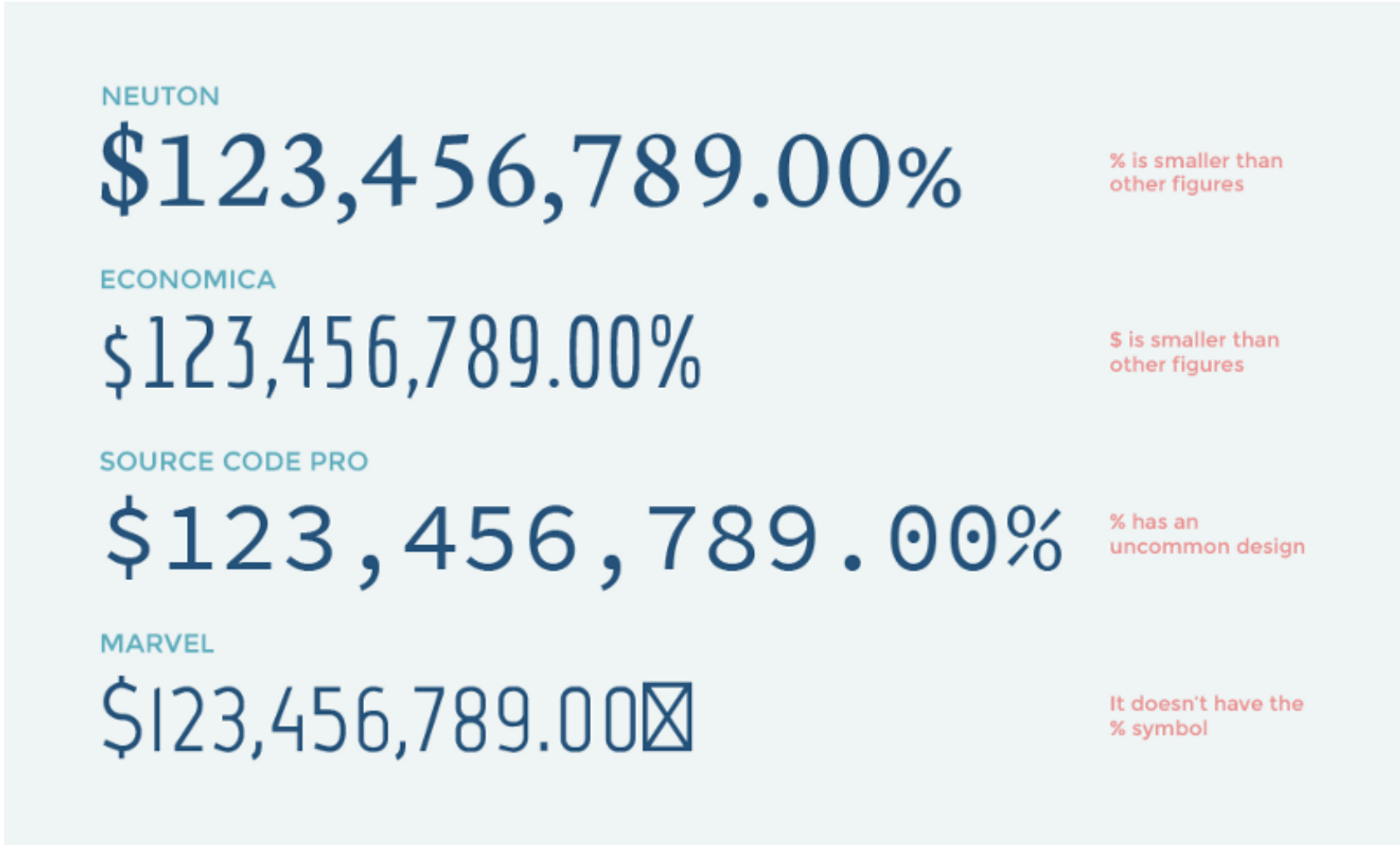
Keep it Simple

Tabular (Monospaced) Numbers



Choosing Fonts for Your Data Visualization by Tiffany France

Quality of Number Symbols



Source: invisionapp.com/inside-design/best-free-fonts-for-numbers

Droid Serif

REGULAR $123,456,789.00%

BOLD **$123,456,789.00%**

Crimson Text

REGULAR $123,456,789.00%

SEMIBOLD **$123,456,789.00%**

BOLD **$123,456,789.00%**

Copse

REGULAR $123,456,789.00%

Kameron

REGULAR $123,456,789.00%

BOLD **$123,456,789.00%**

Open Sans

LIGHT $123,456,789.00%

REGULAR $123,456,789.00%

BOLD **$123,456,789.00%**

Roboto Condensed

LIGHT $123,456,789.00%

REGULAR $123,456,789.00%

BOLD **$123,456,789.00%**

Lato

LIGHT $123,456,789.00%

REGULAR $123,456,789.00%

BOLD **$123,456,789.00%**

Varela Round

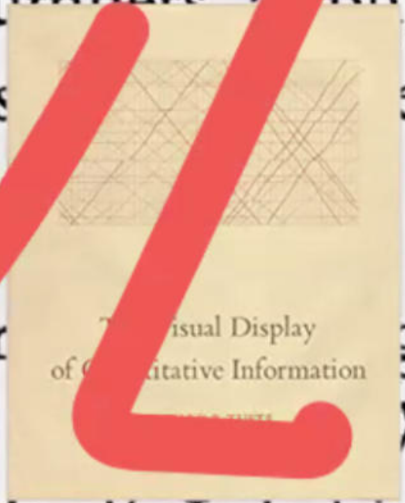
REGULAR $123,456,789.00%

Chart Design Principles

Graphical integrity is more likely to result if these six principles are followed:

RULES

The [illegible] presenta[illegible]on [illegible]f numbers [illegible] physically measured on [illegible] su[illegible] the grap[illegible]ic its[illegible] e direc[illegible] proporti[illegible]al to t[illegible] merical qua[illegible]tities[illegible]

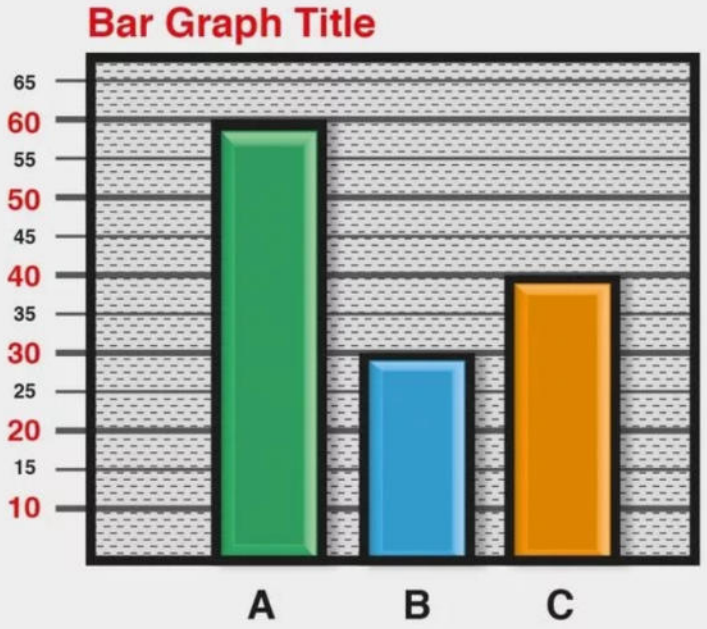


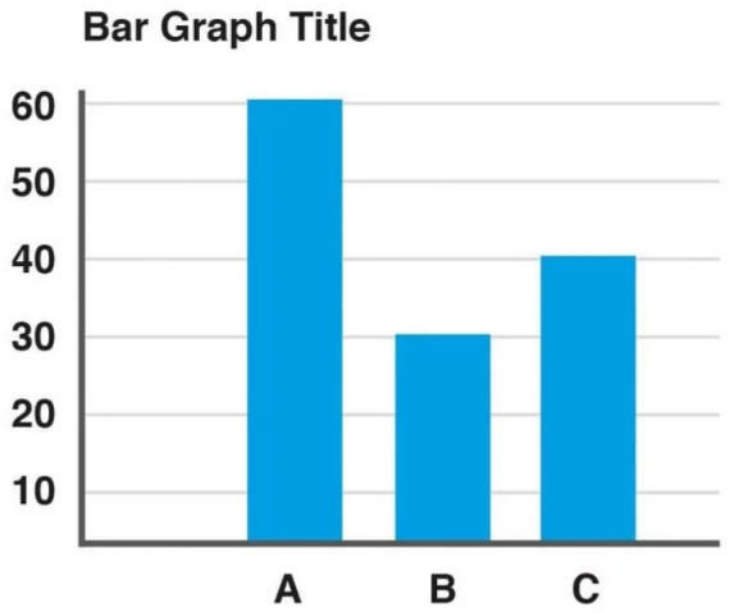
[illegible]lear, de[illegible]iled, a[illegible]d [illegible]or[illegible] g s[illegible]ould be used to [illegible]efeat graphical d[illegible]tortion and [illegible] W[illegible] [illegible]xplanation[illegible] of the data on the graphic itself. Label important eve[illegible] in th[illegible] data.

Show data variation, not design variation.

In time-series displays of money, deflated and standardized units

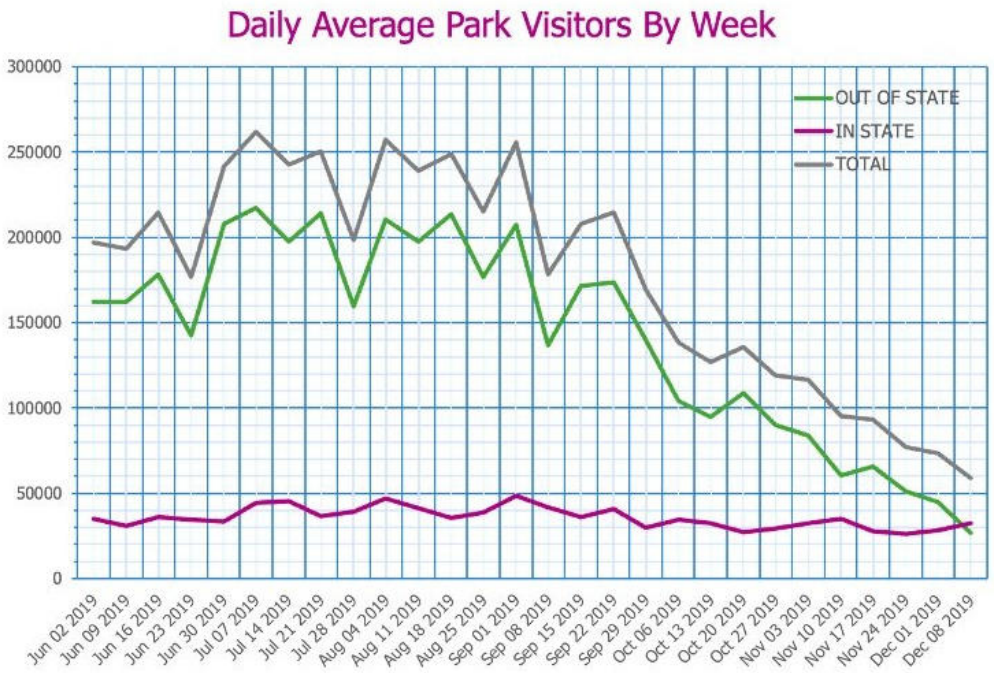
Don't Clutter Your Plot



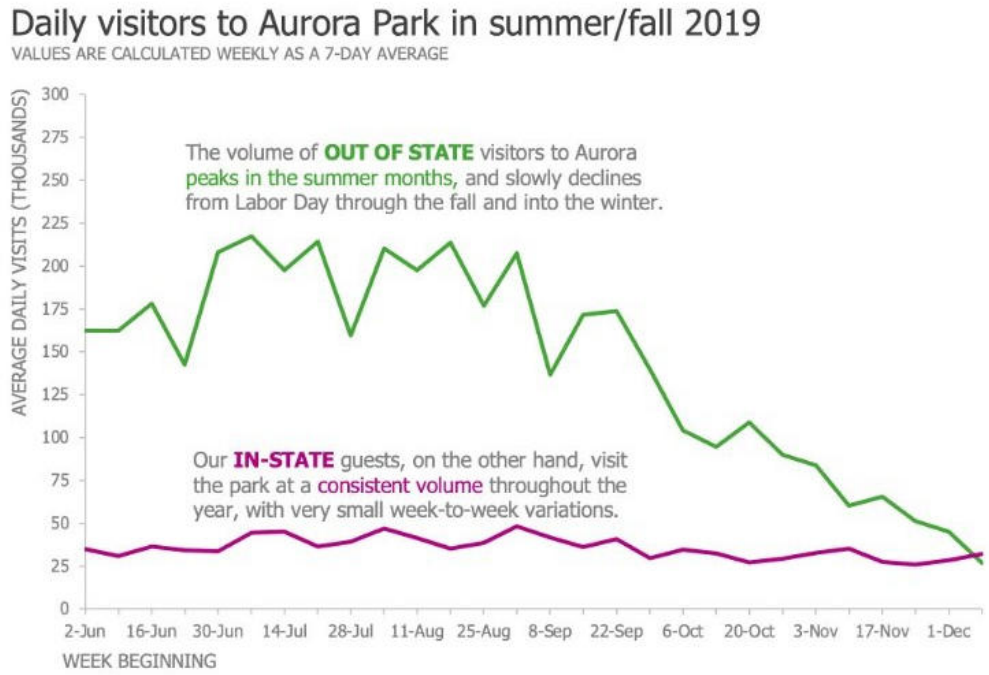


Don't Clutter Your Plot

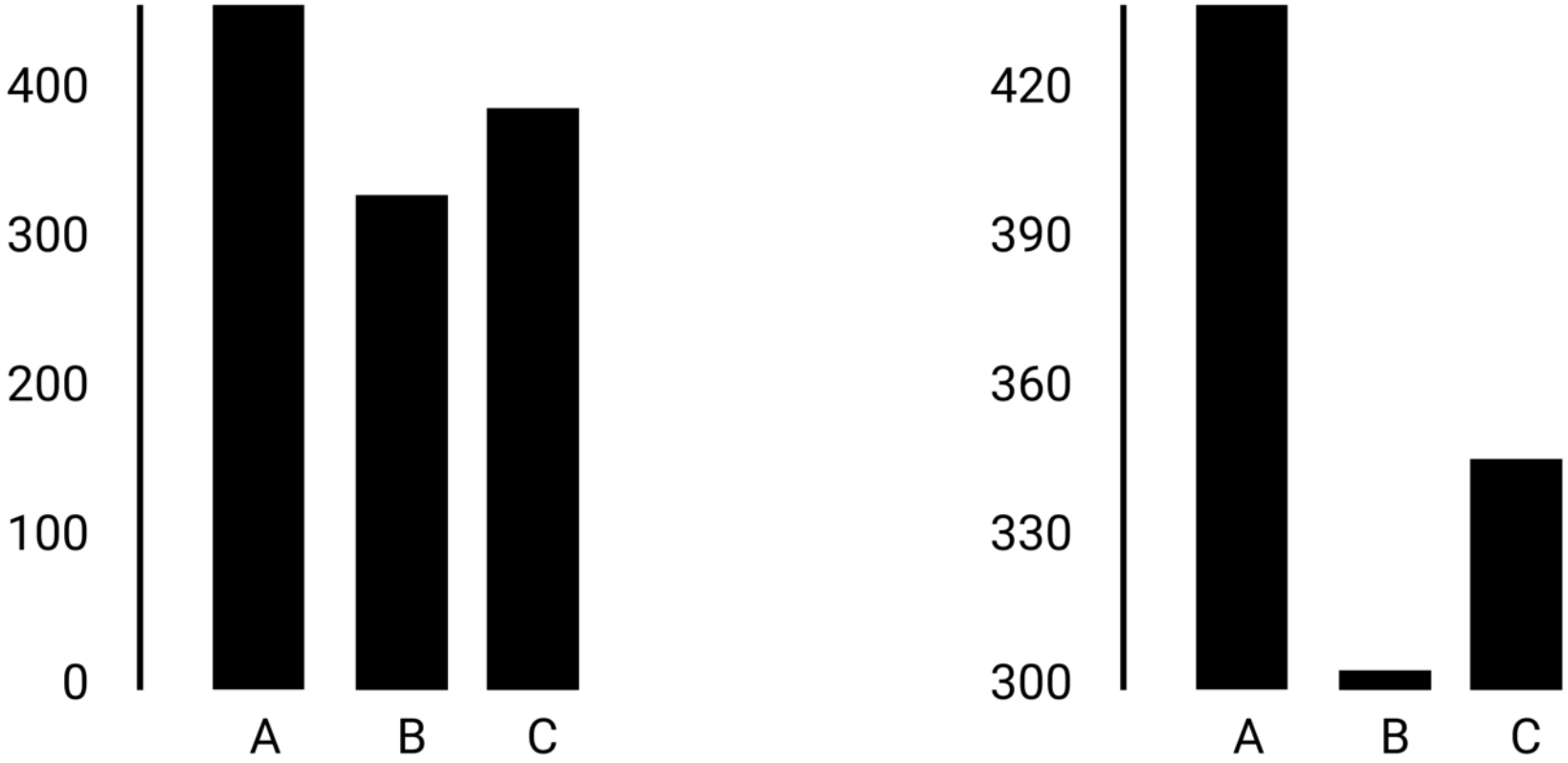
BEFORE: SPOOKY SKELETON



AFTER: GOOD BONES

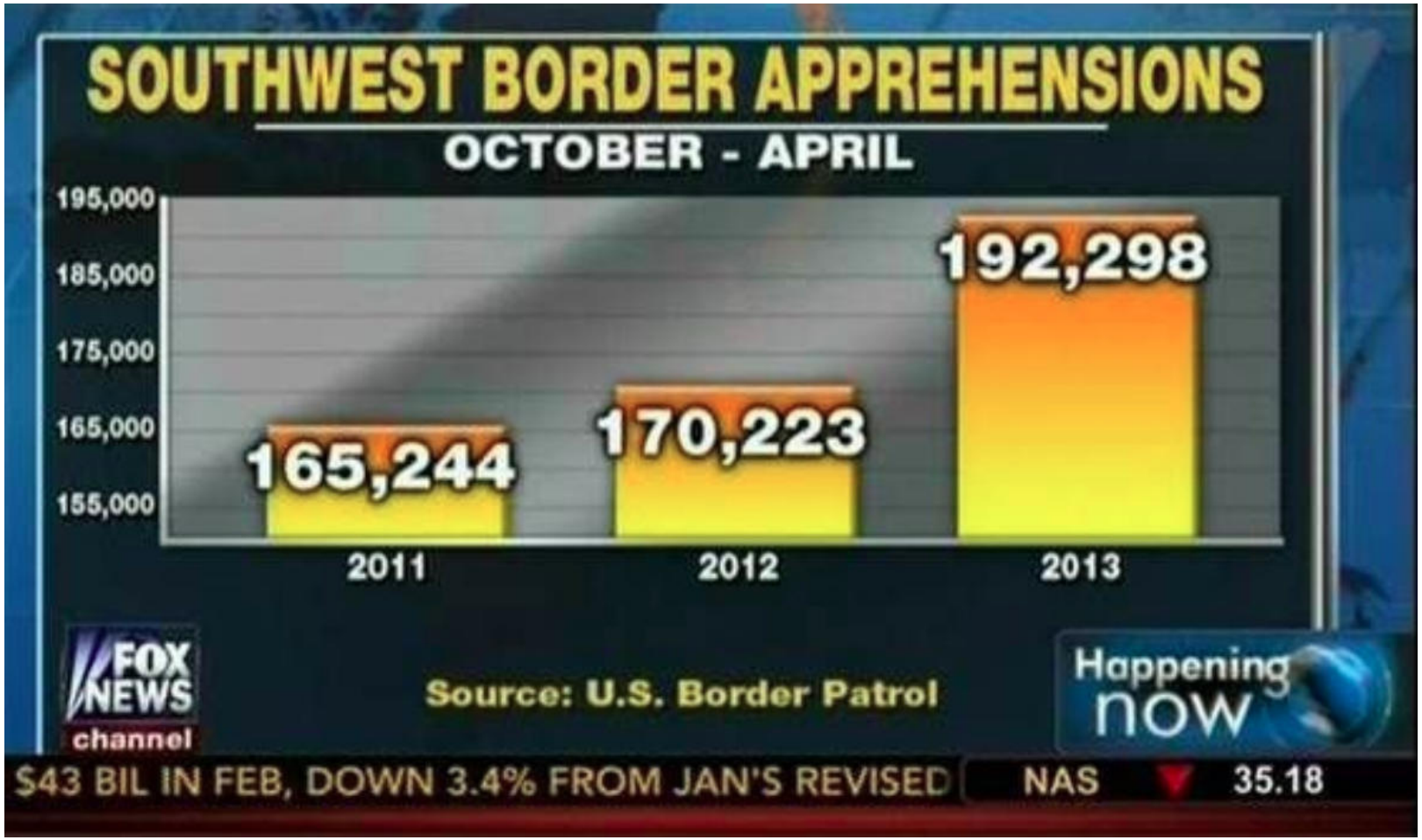


Start at Zero



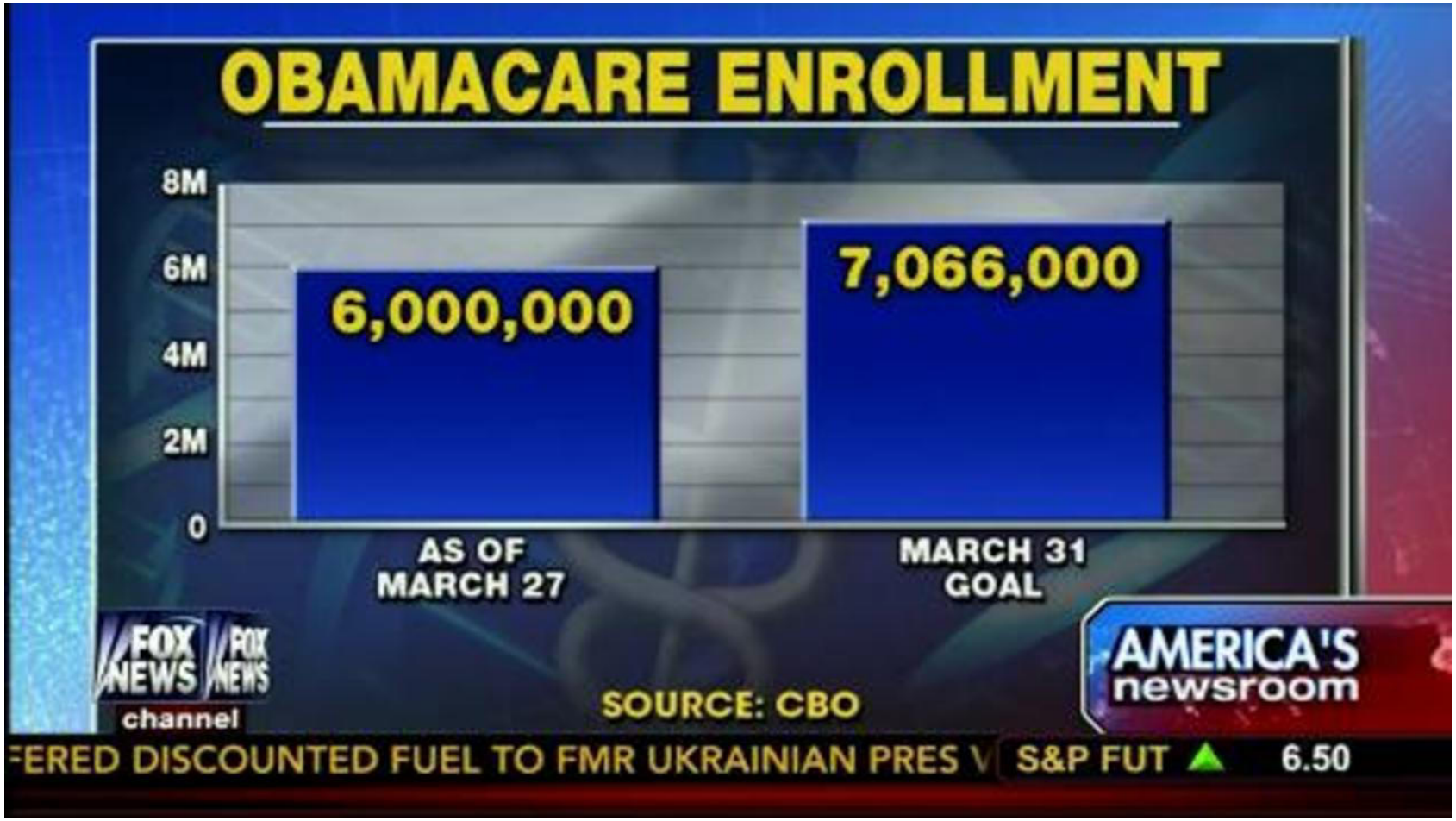
"Hands-On Data Visualization" by Jack Dougherty & Ilya Ilyankou

Always Start at Zero



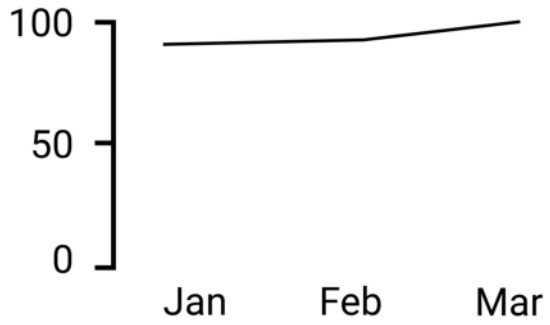
Fox News

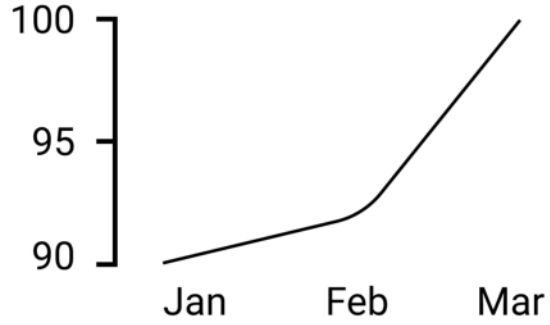
Always Start at Zero?

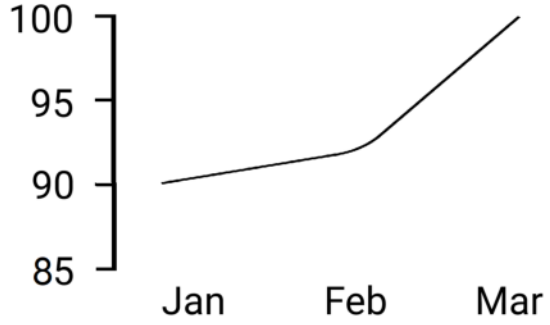


Fox News

Always Start at Zero?

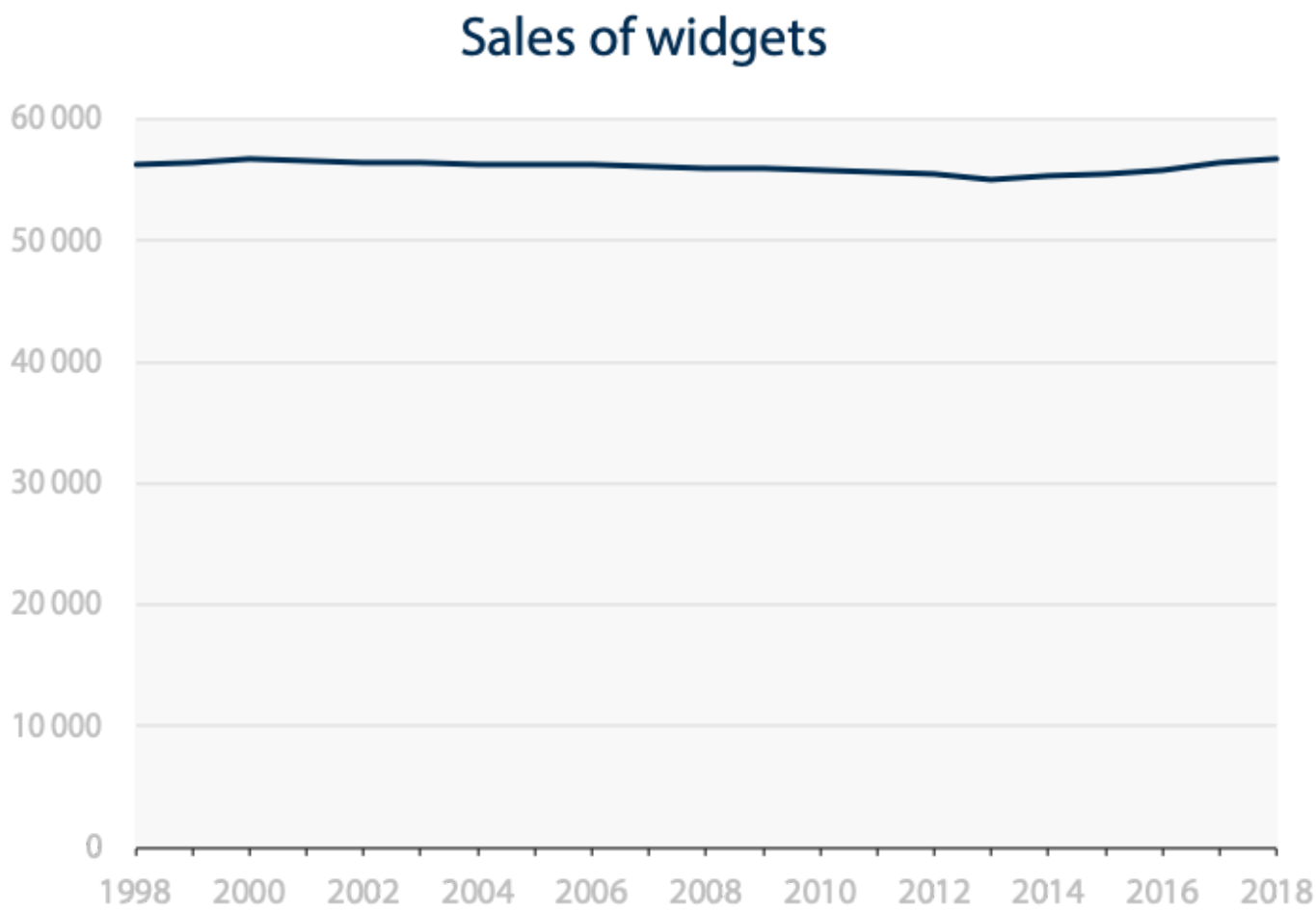






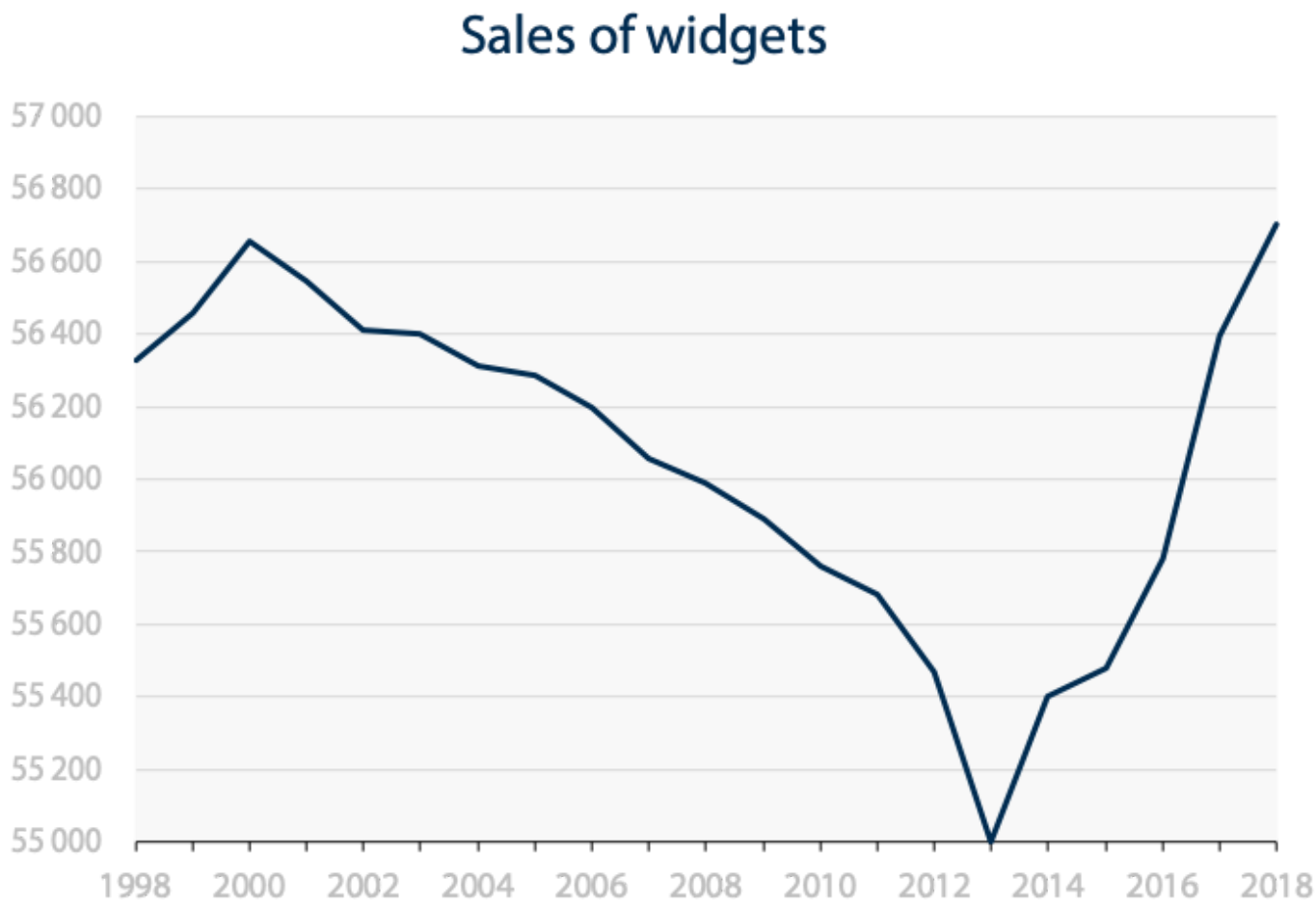
Modified from "Hands-On Data Visualization" by Jack Dougherty & Ilya Ilyankou

Always Start at Zero?



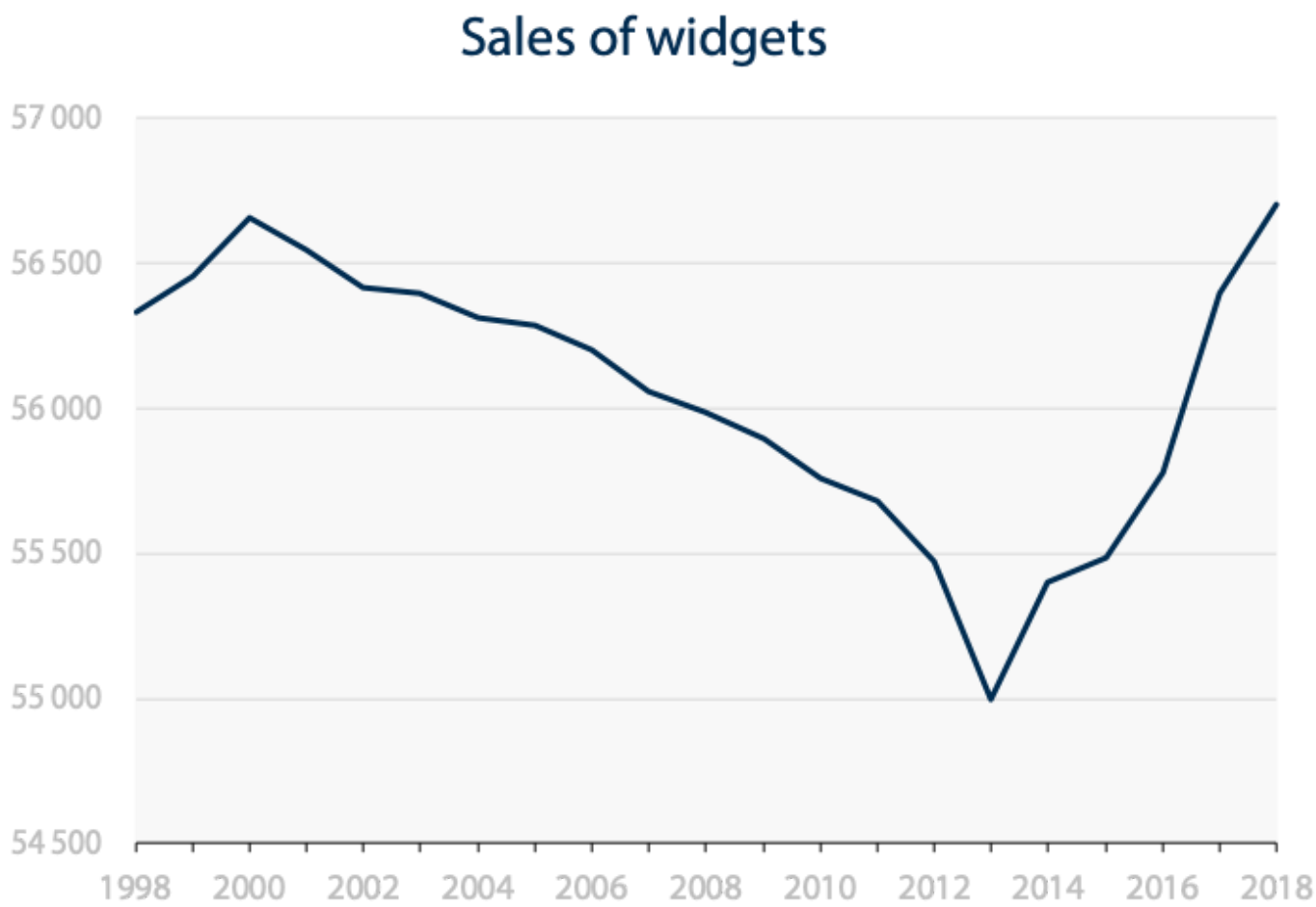
Francis Gagnon, Voilà

Always Start at Zero?



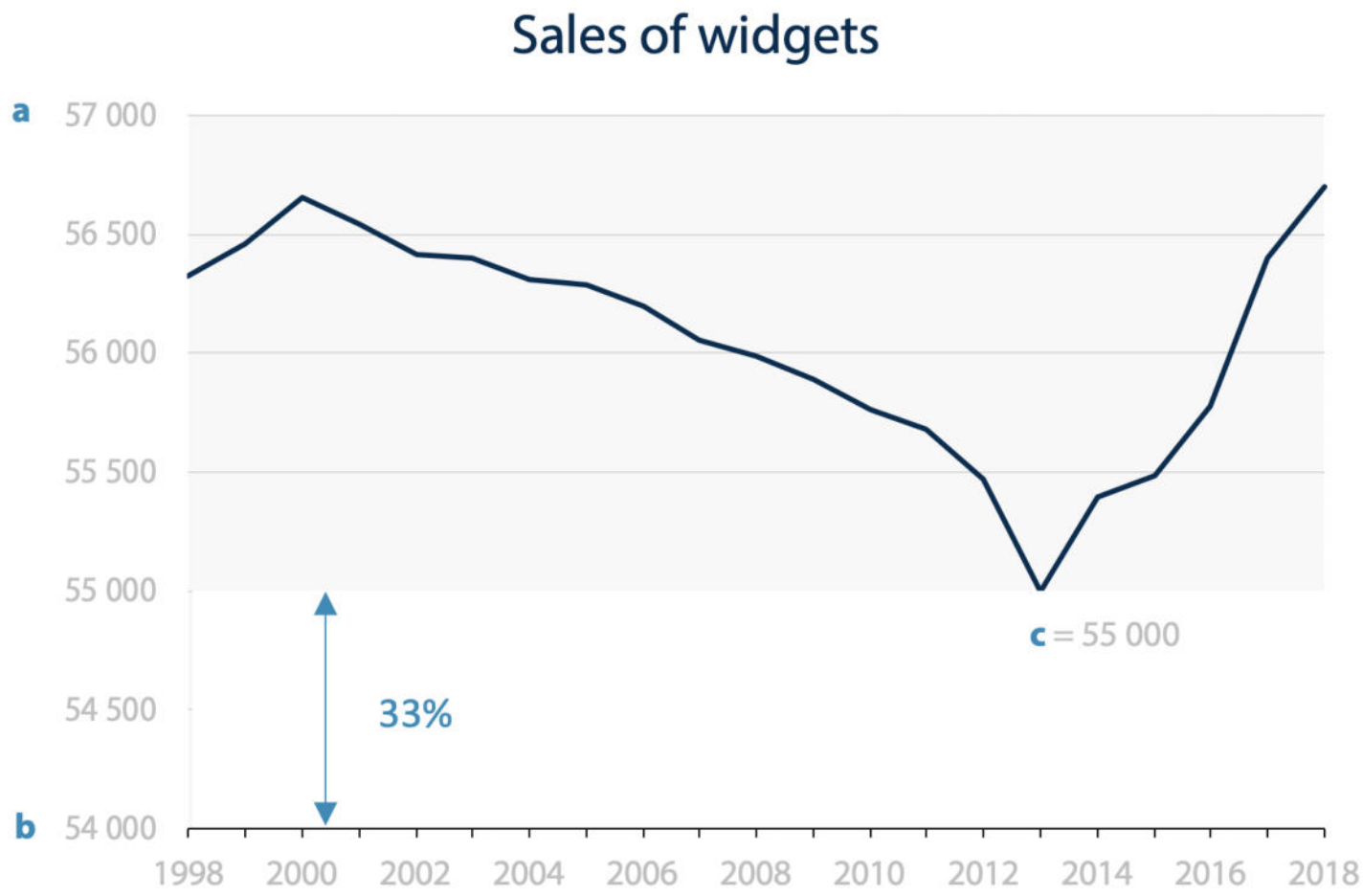
Francis Gagnon, Voilà

The Golden Ratio



Francis Gagnon, Voilà

The Golden Ratio

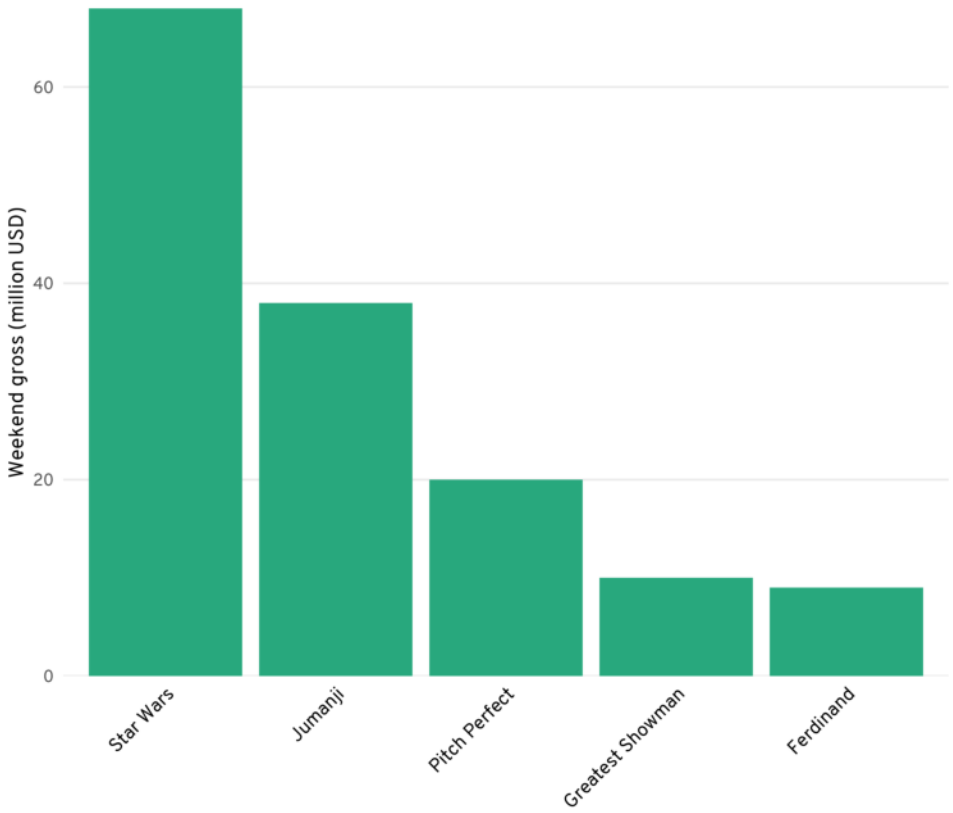


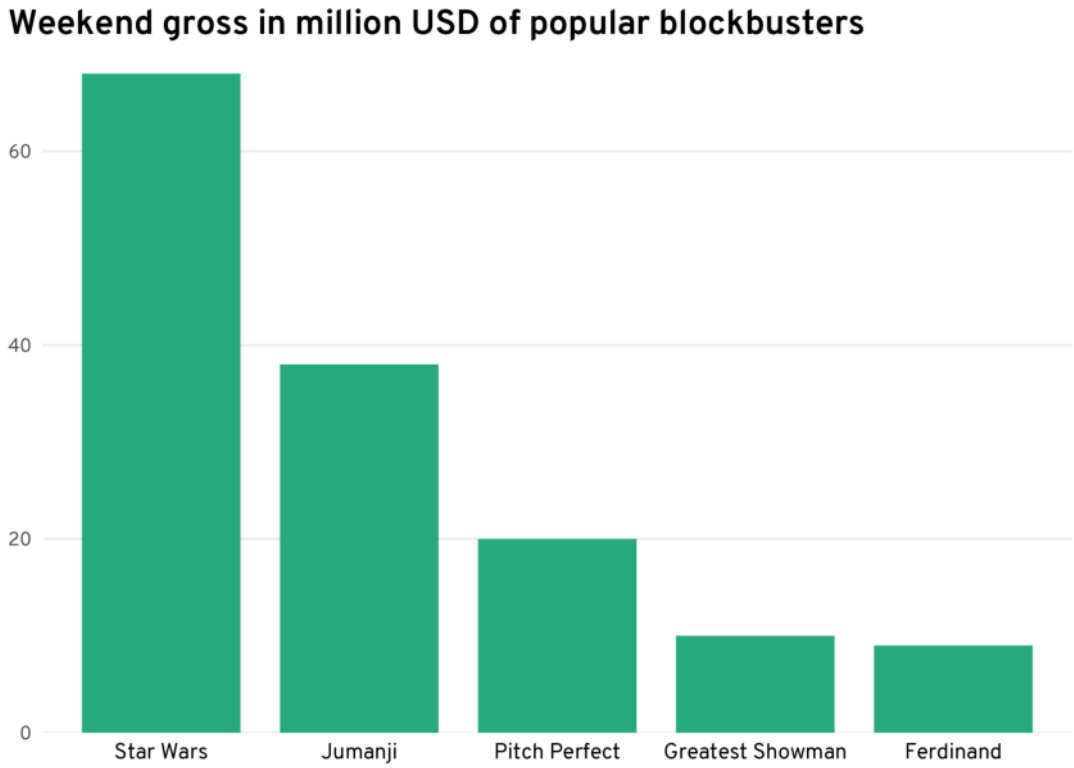
Francis Gagnon, Voilà

Order Your Data

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

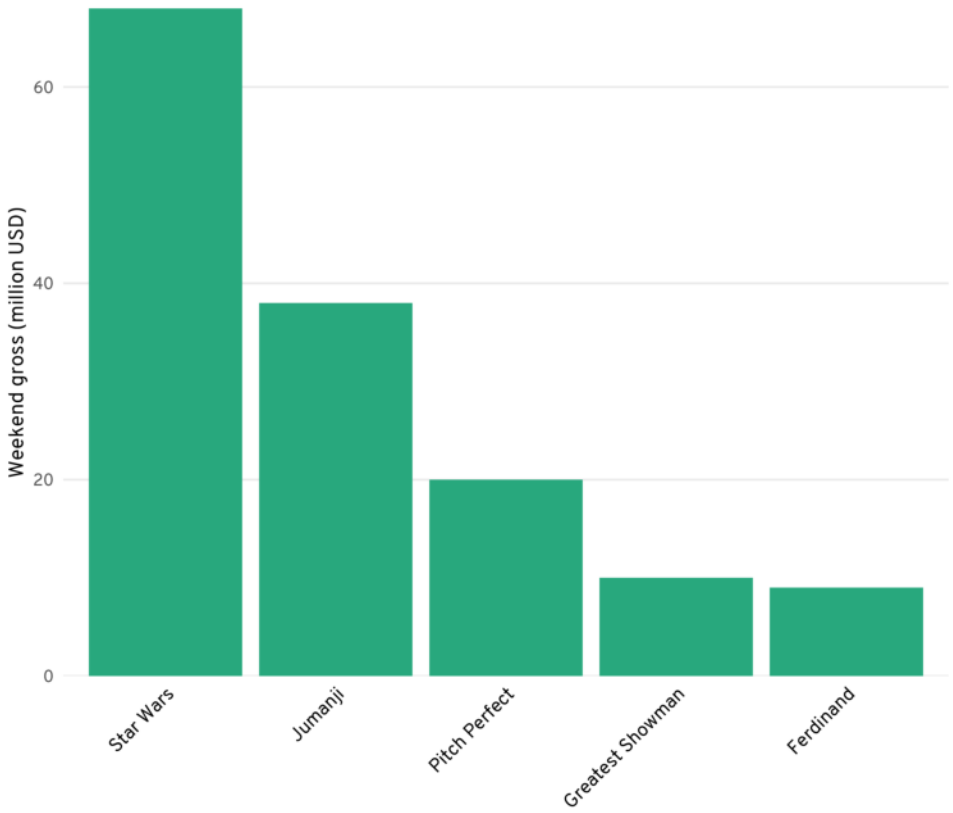
(Don't) Rotate Your Text

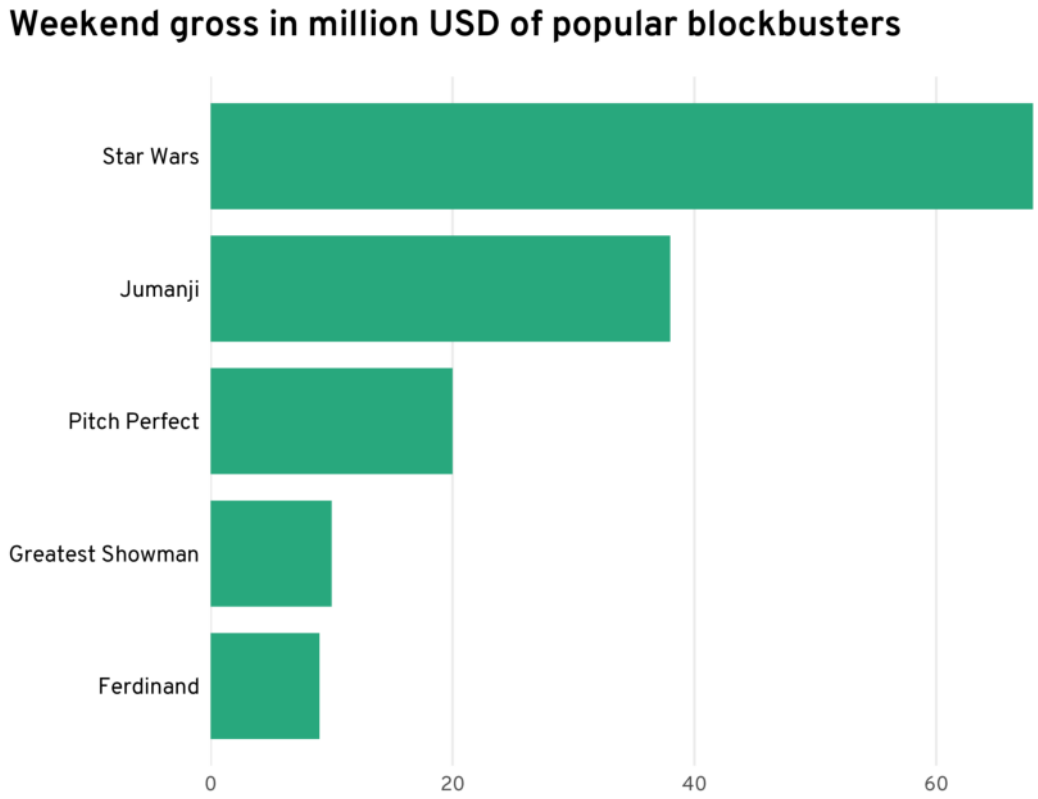




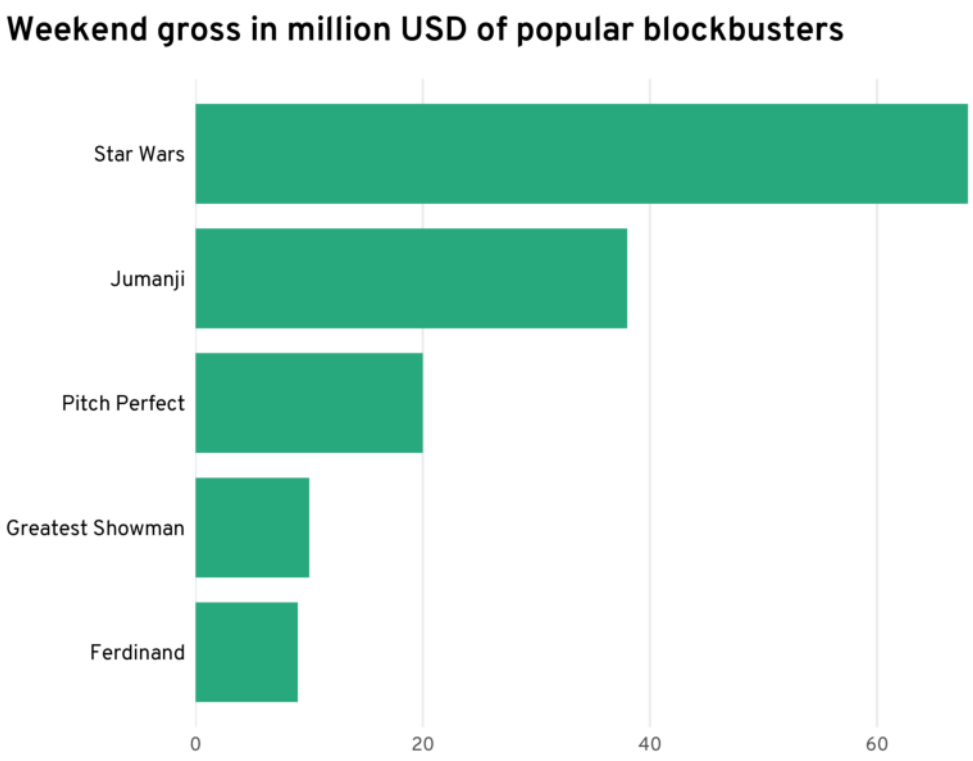
Modified from "Fundamentals of Data Visualization" by Claus Wilke

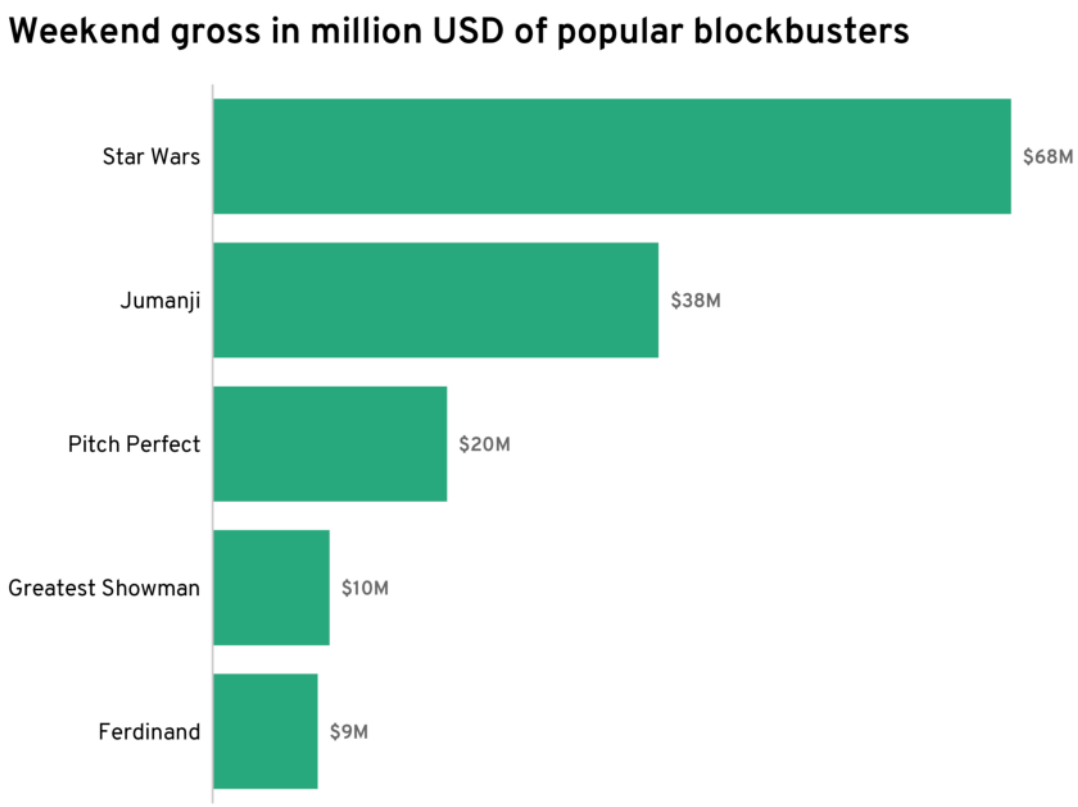
(Don't) Rotate Your Text



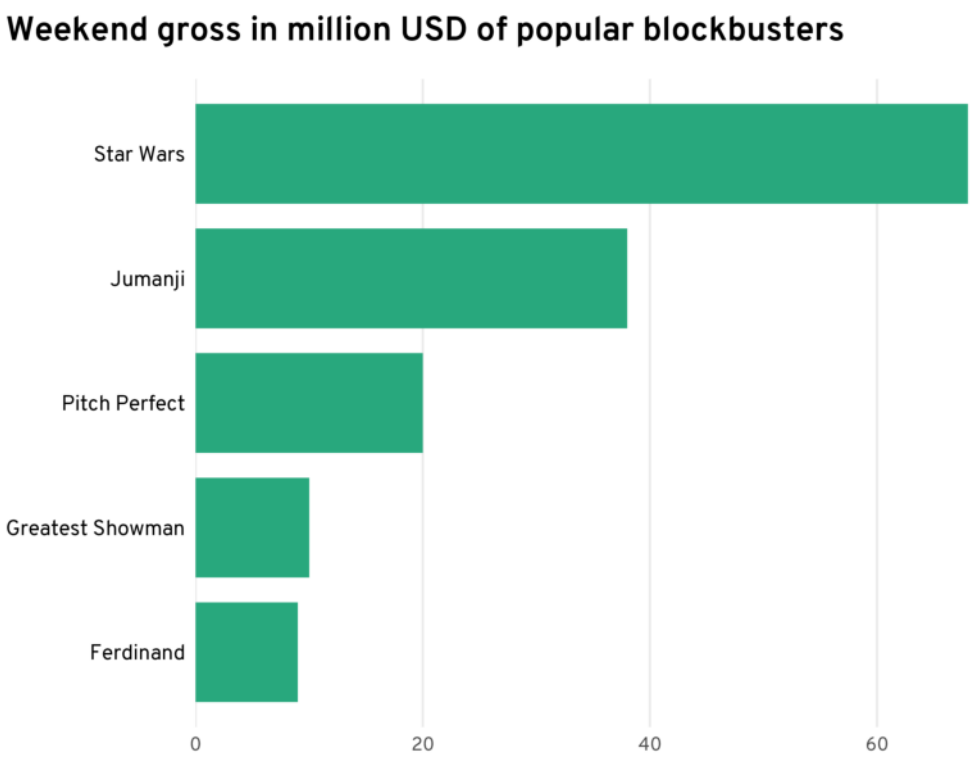


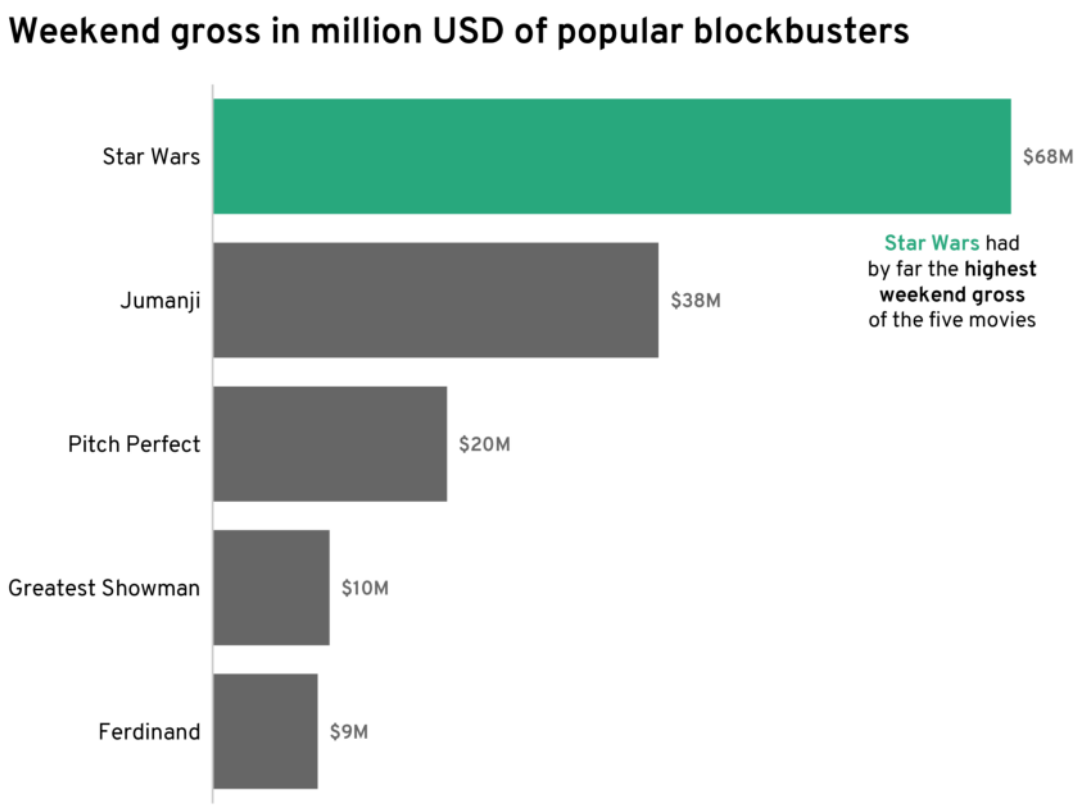
Use Annotations





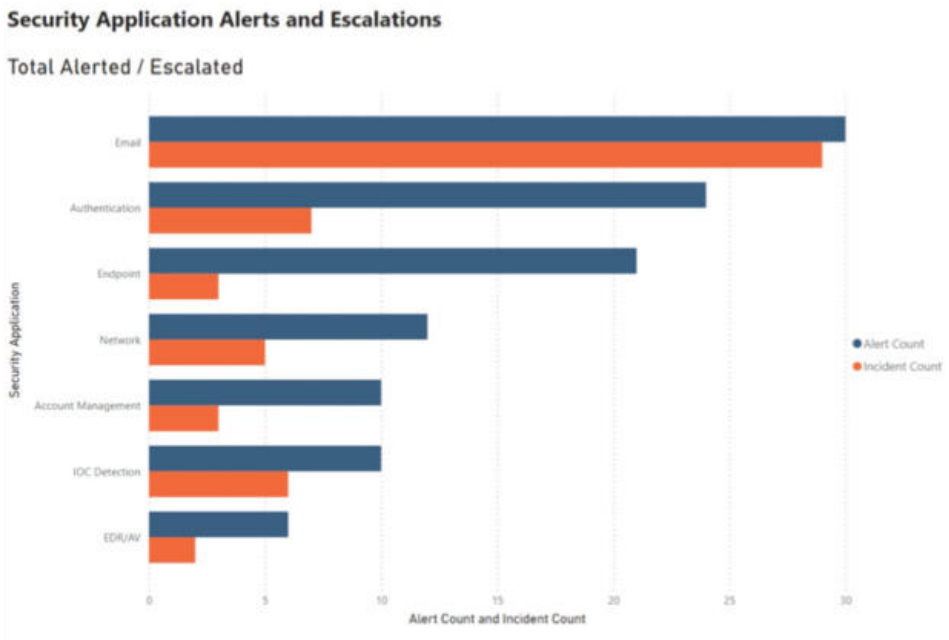
Use Annotations



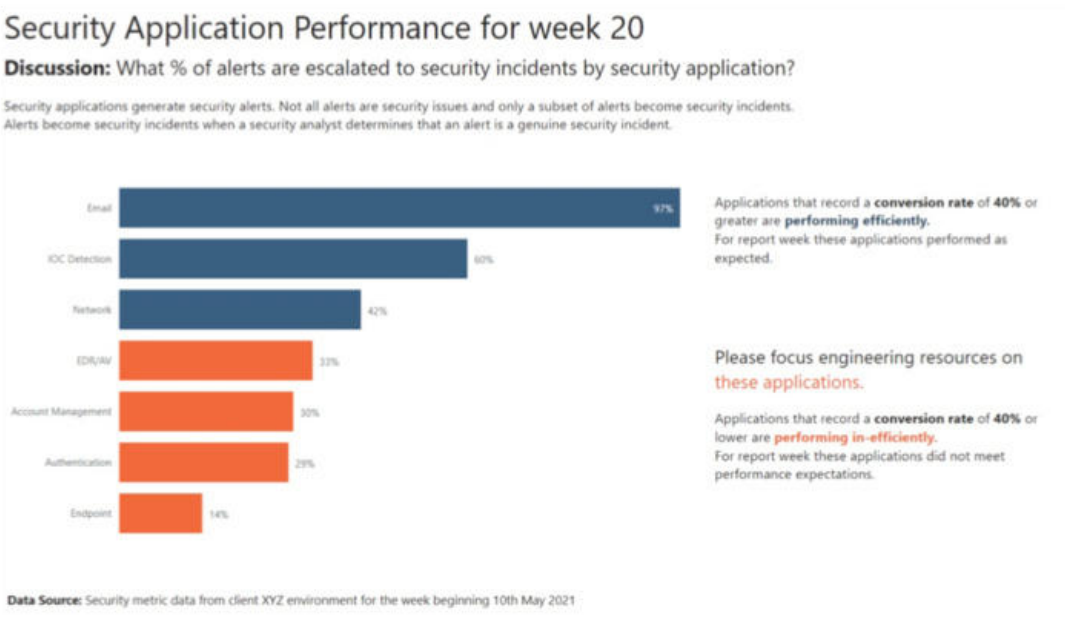


Use Annotations

BEFORE:

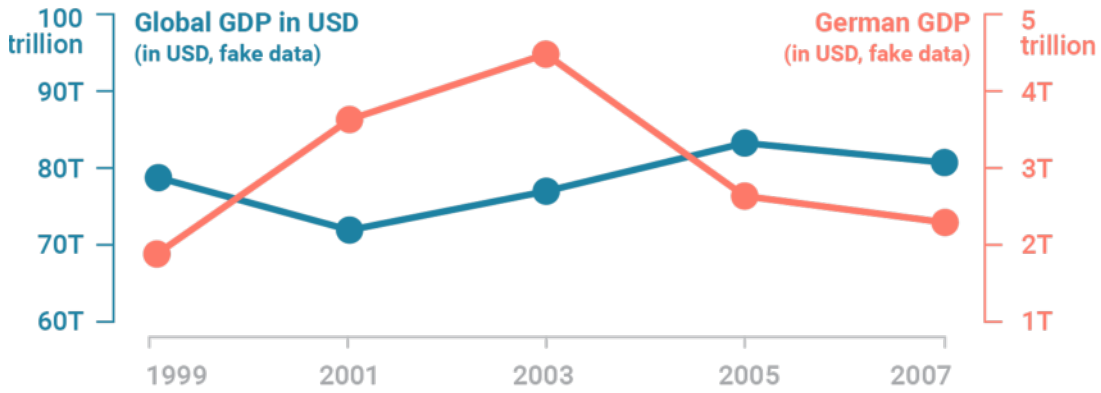


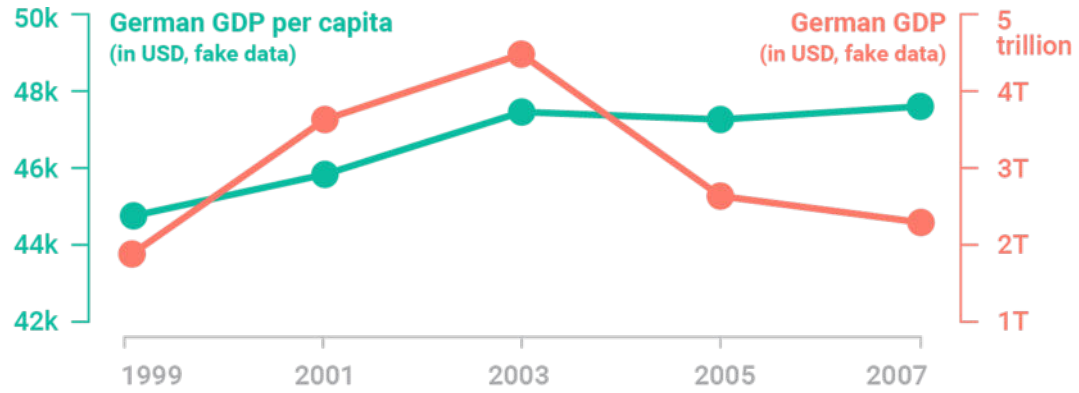
AFTER:

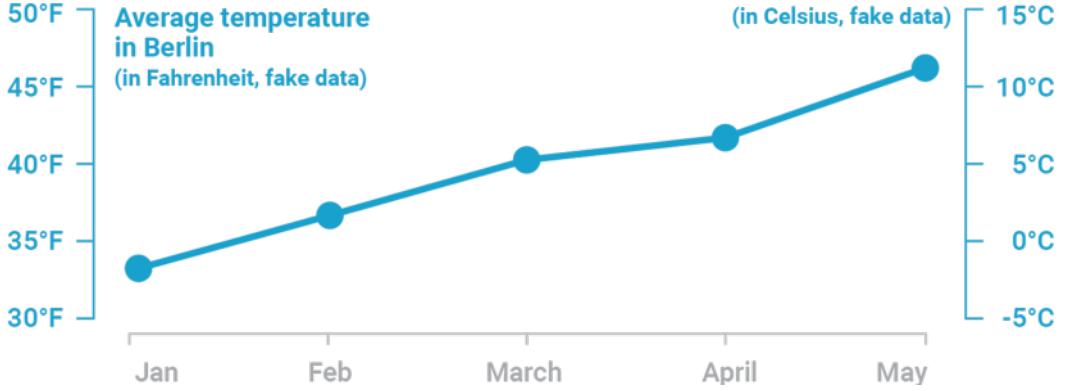


storytellingwithdata.com

Avoid Dual Axes

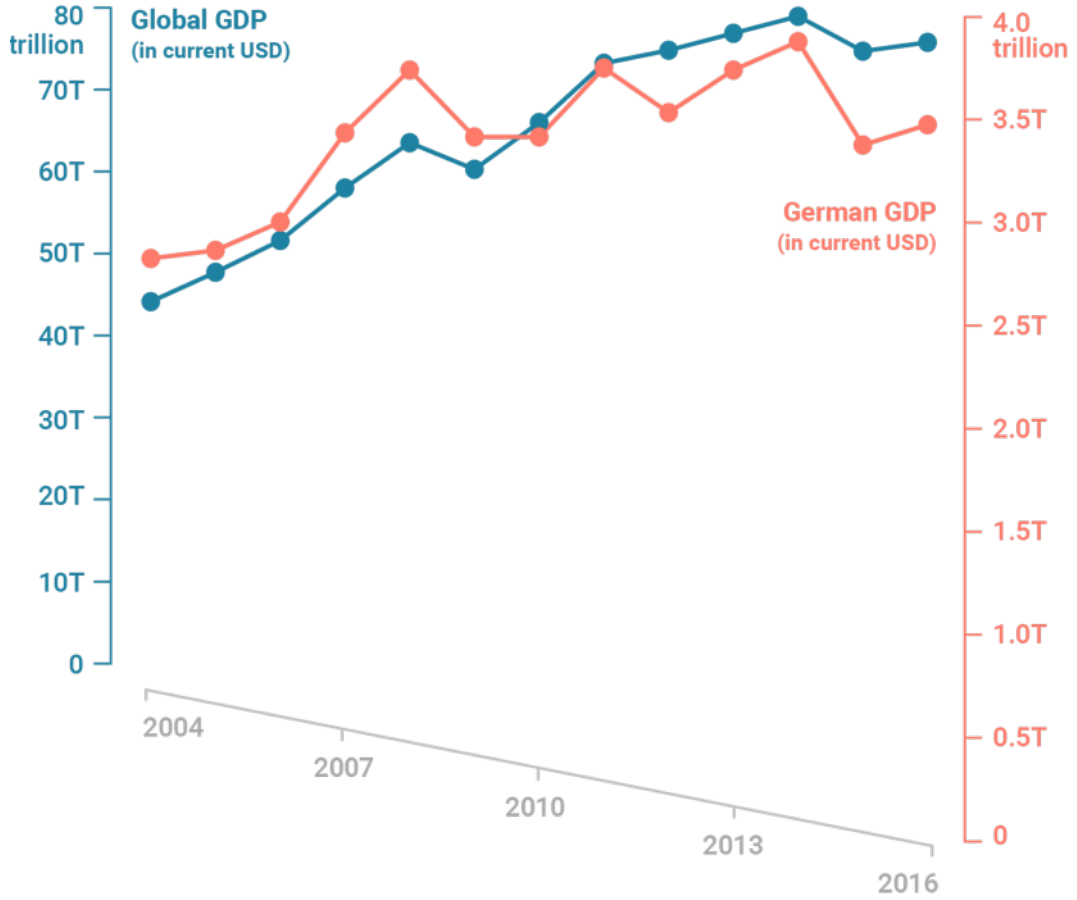






"Why not to use two axes, and what to use instead" by Lisa Charlotte Muth/DataWrapper

Avoid Dual Axes

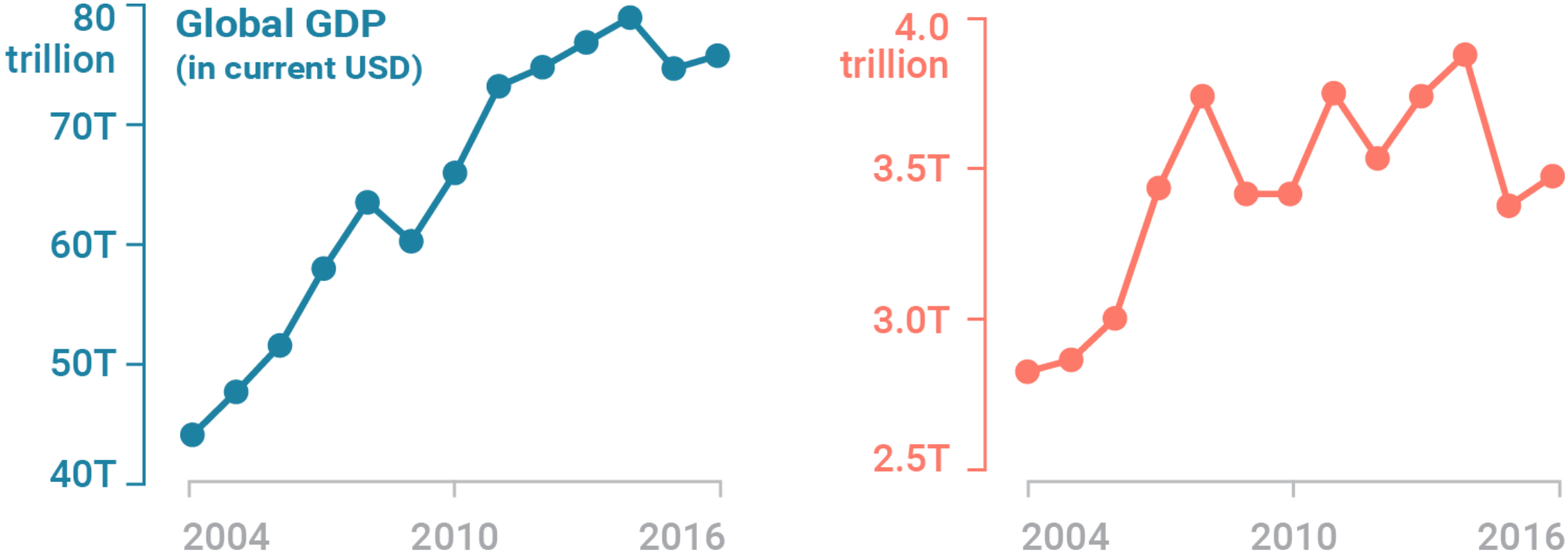


"Why not to use two axes, and what to use instead" by Lisa Charlotte Muth/DataWrapper

Avoid Dual Axes

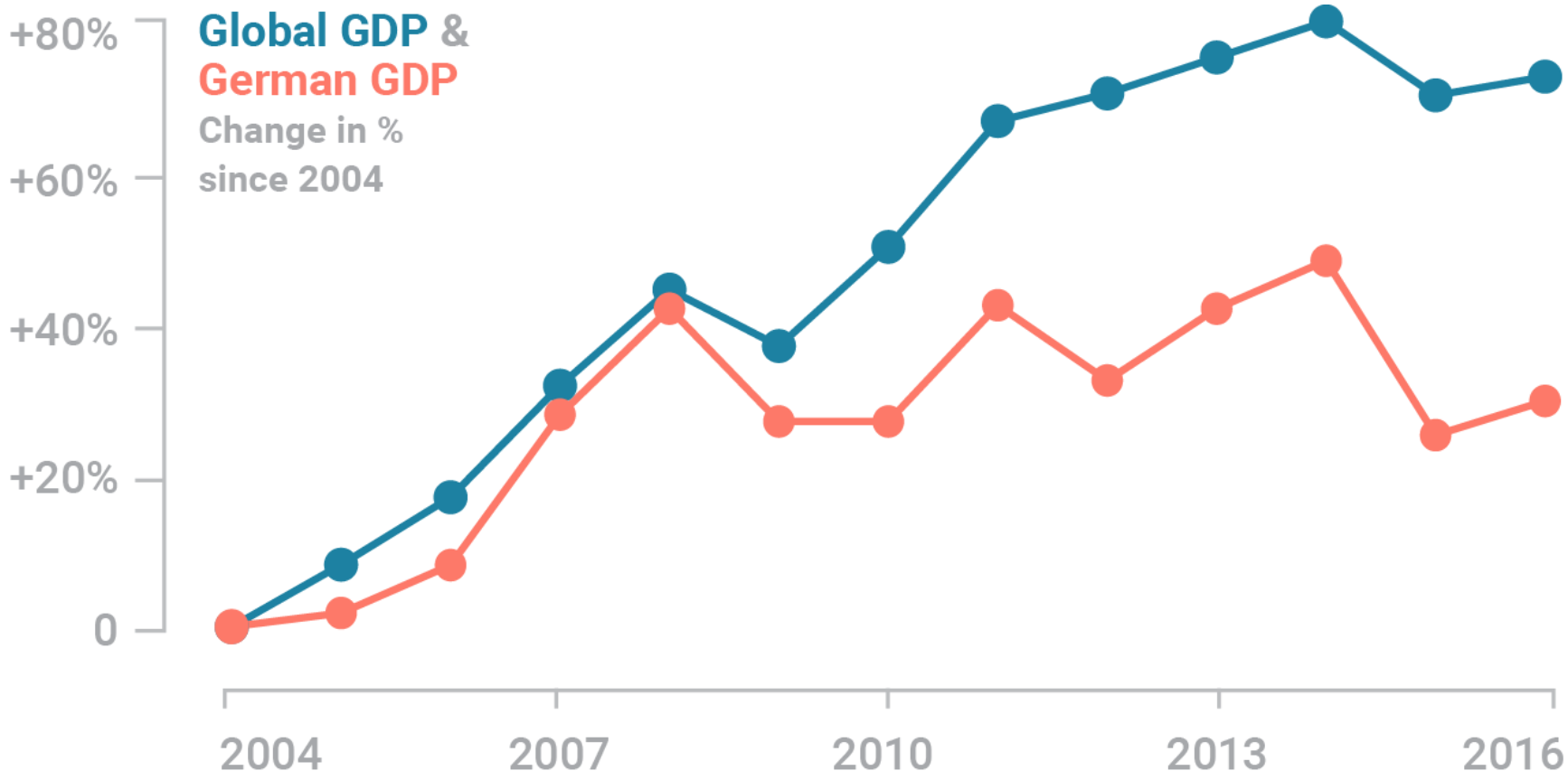
"Why not to use two axes, and what to use instead" by Lisa Charlotte Muth/DataWrapper

Avoid Dual Axes



"Why not to use two axes, and what to use instead" by Lisa Charlotte Muth/DataWrapper

Avoid Dual Axes



"Why not to use two axes, and what to use instead" by Lisa Charlotte Muth/DataWrapper

A word about
Tables

Make Your Data Tables More Visual

J. Benefit Cost Anal. 2020; 11(2):151–178
doi:10.1017/bca.2020.11

Jonathan A. Schwabish

Ten Guidelines for Better Tables

Abstract: Tables are a unique form of visualizing data because, unlike many charts, they are not usually intended to give a quick, visual representation of data. Instead, tables are useful when you want to show the exact values of your data or estimates. They are not the best solution if you want to show a lot of data or if you want to show the data in a compact space, but a well-designed table can help your reader find specific numbers and discover patterns and outliers. In this article, I present 10 guidelines for creating better, more effective tables; I then model these lessons by redesigning six tables from articles previously published in the *Journal of Benefit-Cost Analysis*.

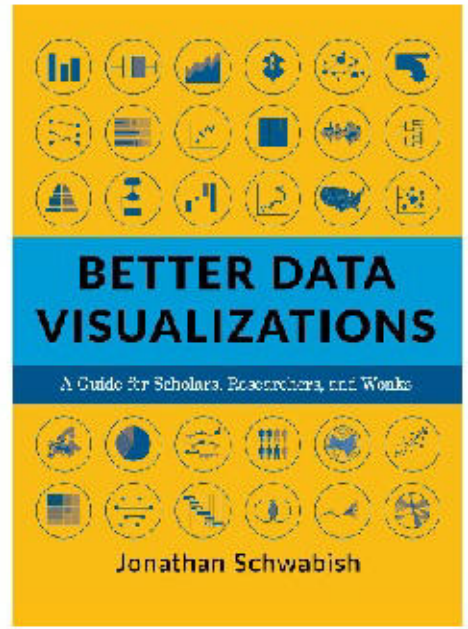
John Schwabisch

J. Benefit Cost Anal. 2020; 11(2):151–178
doi:10.1017/bca.2020.11

Jonathan A. Schwabish

Ten Guidelines for Better Tables

Abstract: Tables are a unique form of visualizing data because, unlike many charts, they are not usually intended to give a quick, visual representation of data. Instead, tables are useful when you want to show the exact values of your data or estimates. They are not the best solution if you want to show a lot of data or if you want to show the data in a compact space, but a well-designed table can help your reader find specific numbers and discover patterns and outliers. In this article, I present 10 guidelines for creating better, more effective tables; I then model these lessons by redesigning six tables from articles previously published in the *Journal of Benefit-Cost Analysis*.



	Potato Yield in Tonnes/Hectare		
Country	2013	2017	2013-2017
Brazil	27.75	30.94	
China	17.09	18.21	
Denmark	41.57	43.68	
El Salvador	42.60	29.22	
France	43.16	44.05	
Germany	39.83	46.79	
India	22.76	22.31	
Indonesia	16.02	15.40	
Ireland	38.33	44.83	
Italy	25.25	27.73	
Lebanon	26.08	25.14	
Mexico	26.78	28.95	
Netherlands	42.21	45.97	
Pakistan	21.81	21.45	
United States	46.36	48.39	

Table: @thomas_mock | **Data**: OurWorldInData.org

Inspiration: @jschwabish

	Potato Yield in Tonnes/Hectare				
Country	2013	2014	2015	2016	2017
United States	46.36	47.15	46.90	48.64	48.39
France	43.16	47.98	42.51	38.83	44.05
El Salvador	42.60	26.25	26.01	25.54	29.22
Netherlands	42.21	45.66	42.73	42.00	45.97
Denmark	41.57	43.12	41.41	42.48	43.68
Germany	39.83	47.42	43.81	44.40	46.79
Ireland	38.33	40.32	42.36	39.11	44.83
Brazil	27.75	27.94	29.32	29.66	30.94
Mexico	26.78	27.34	27.14	27.93	28.95
Lebanon	26.08	25.23	24.82	25.07	25.14
Italy	25.25	26.08	27.16	28.44	27.73
India	22.76	22.92	23.13	20.51	22.31
Pakistan	21.81	18.15	23.44	22.43	21.45
China	17.09	17.14	17.27	17.69	18.21
Indonesia	16.02	17.67	18.20	18.25	15.40

Table: @thomas_mock | **Data**: OurWorldInData.org

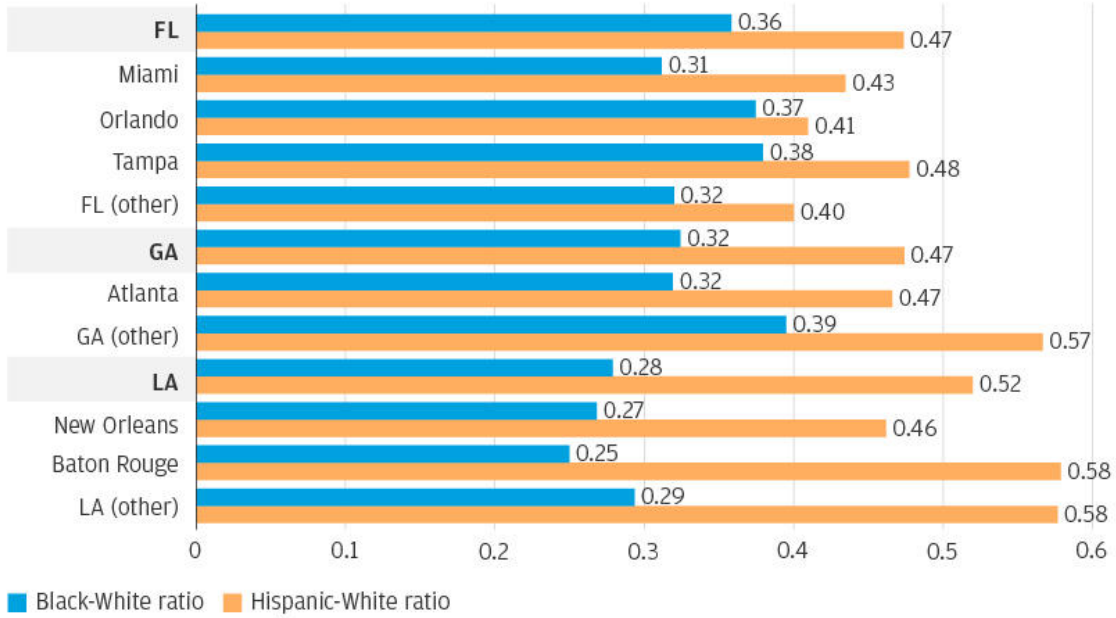
Inspiration: @jschwabish

Thomas Mock

How to improve this visualization?

Finding Four: Across geographies, the financial outcomes of Hispanic families vary the most, while the financial outcomes of Black families vary the least. Black-White gaps in financial outcomes are largest in Louisiana, while Hispanic-White gaps are largest in Florida.

Black-White and Hispanic-White ratios of annual median liquid assets (2018), by geography

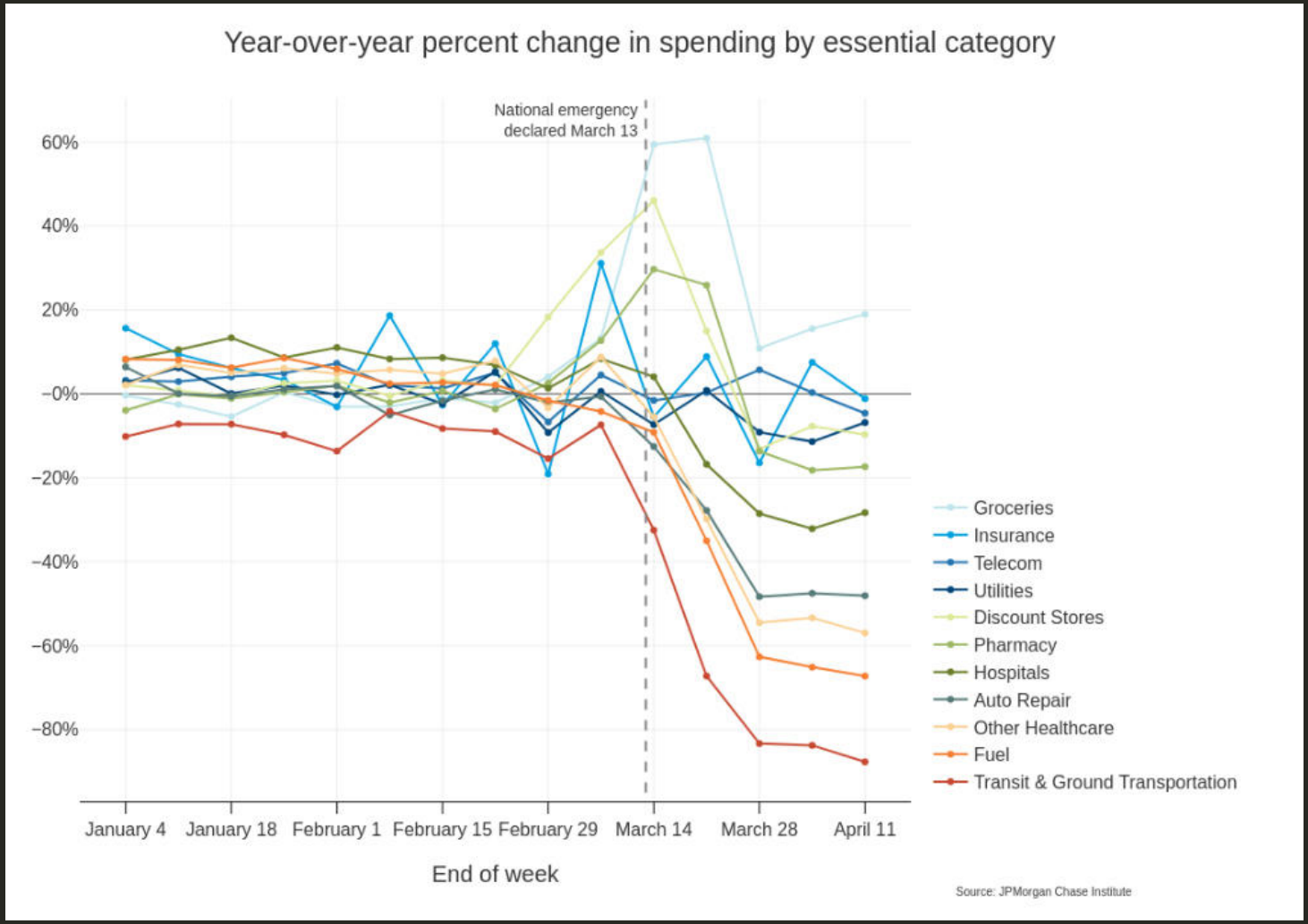


Note: Liquid assets is the sum of balances in one's checking, prepaid debit cards, savings, money market, and certificates of deposit accounts. Cities refer to CBSAs (e.g., Miami refers to the Miami-Fort Lauderdale-West Palm Beach CBSA).

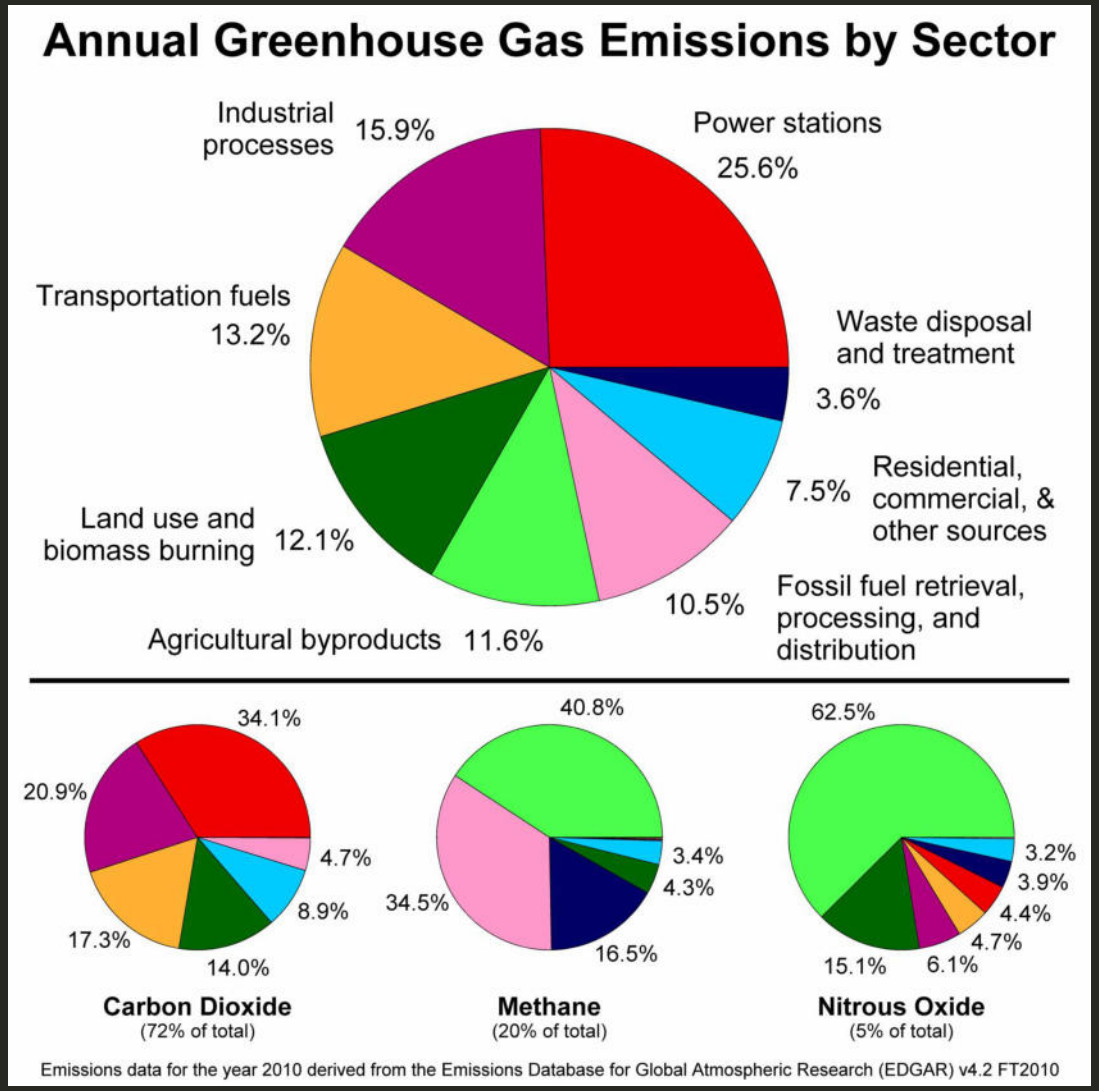
Source: JPMorgan Chase Insitute

View the Text Version >

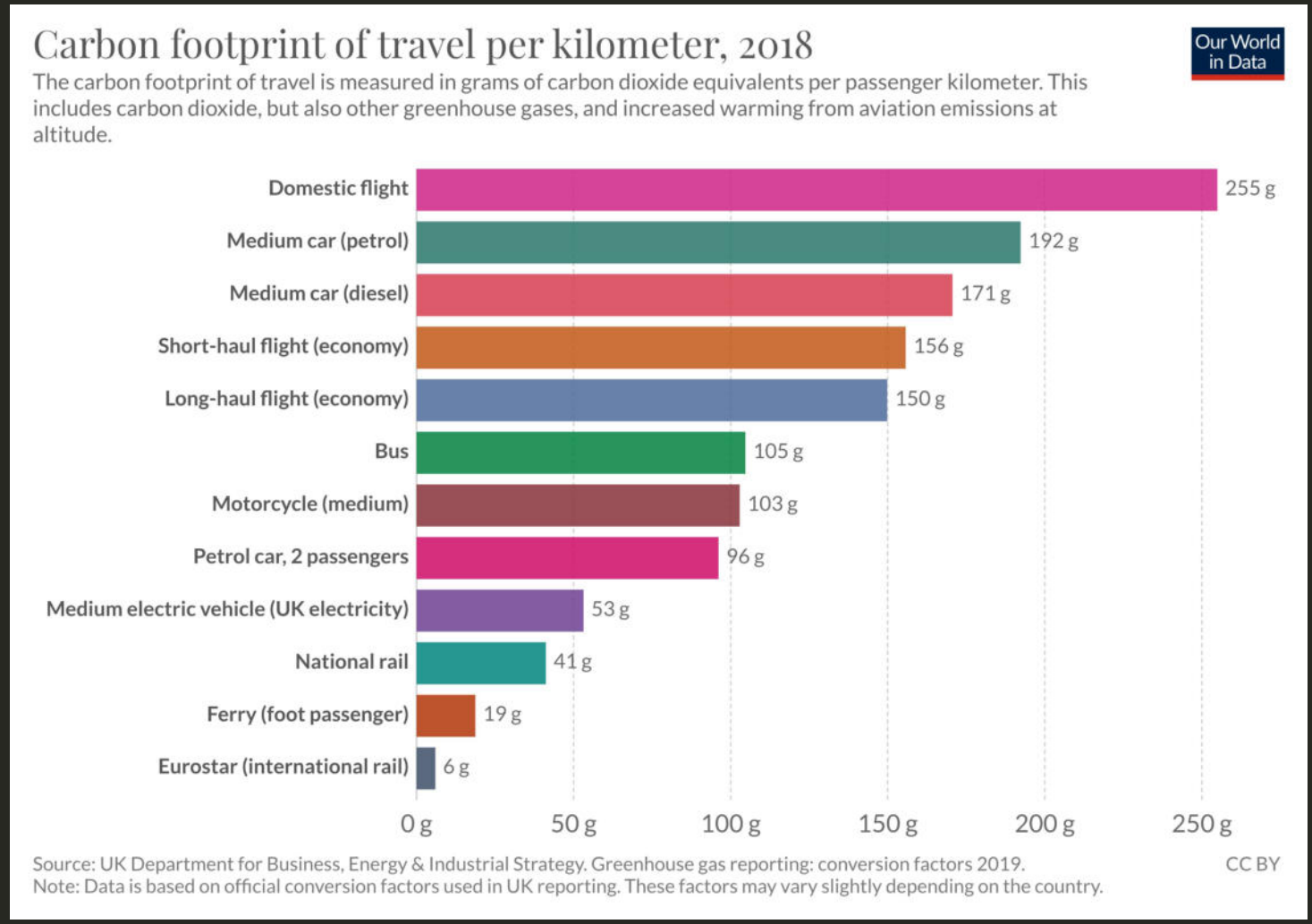
How to improve this visualization?



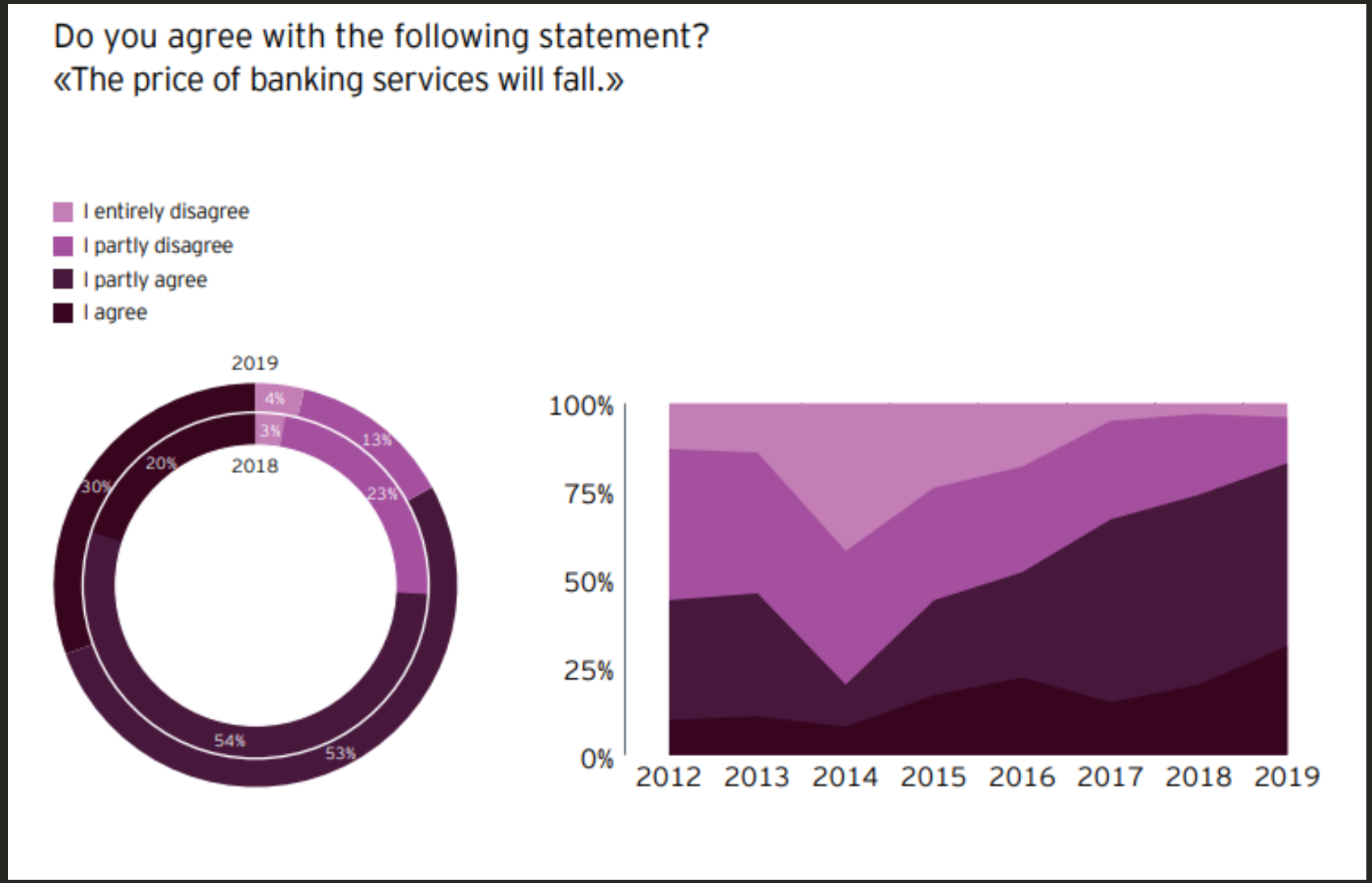
How to improve this visualization?



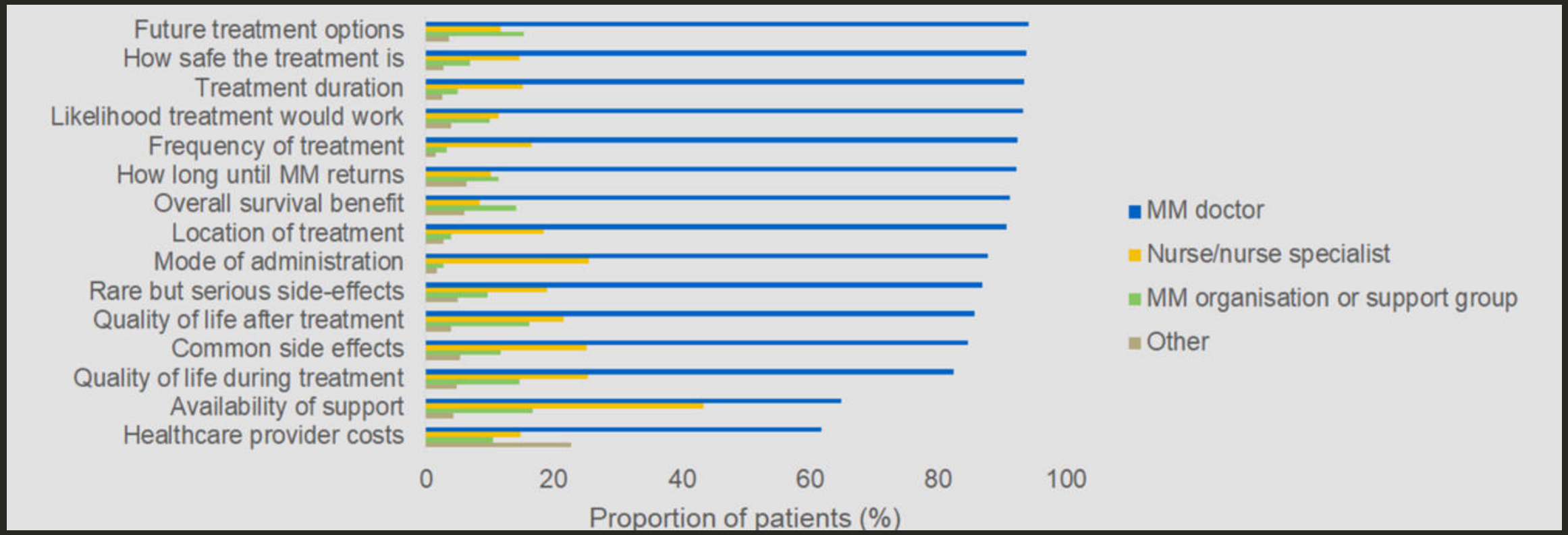
How to improve this visualization?



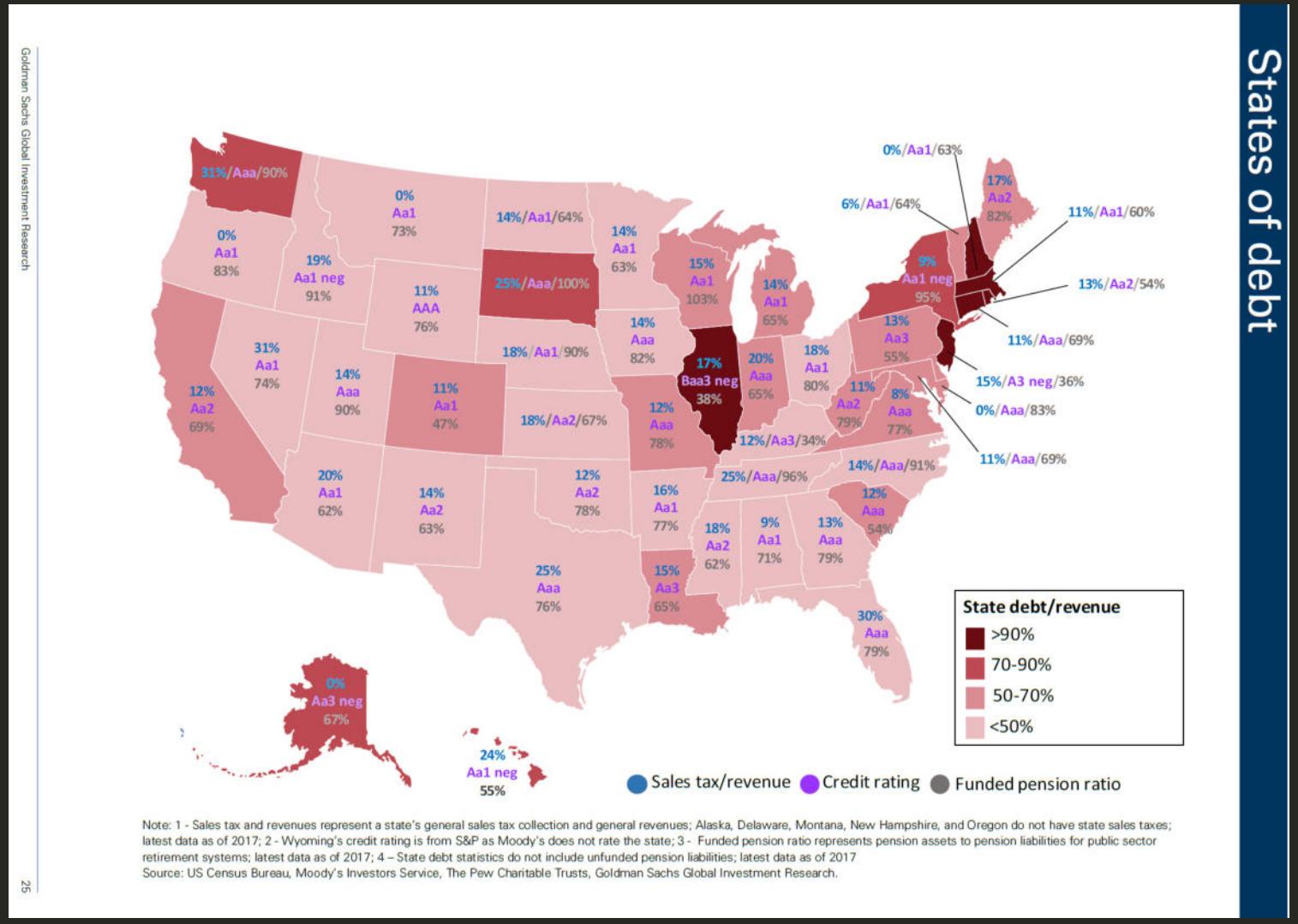
How to improve this visualization?



How to improve this visualization?



How to improve this visualization?



Wrap-Up

Information

→ Understand your data and be accurate.

Story

→ Be clear about the message of your visualization.

Goal

→ Select charts that successfully transport your story.

Visual Form

→ Follow design rules and data visualization principles.

Design for Your Audience

- Choose charts based on your goal not tradition or novelty (only).
- Make sure your visualization is accessible for everyone (colors, readability).
- Use visual contrast to highlight important information.
- Provide meaningful labels and titles.

Be Honest

- Show raw data if possible.
- Don't truncate bar charts. Add spacing to truncated axes.
- Be consistent with axis scaling.

Lend A Helping Hand

- Use annotations and direct labels instead of/in addition to text and legends.
- Order your data, either by value or intrinsic ranking.
- Reveal information step by step (if applicable).

Books

- "Avoiding Data Pitfalls" by Ben Jones
- "Data Visualization", an open-access book by Kieran Healy
- "The Functional Art", "The Truthful Art", and "How Charts Lie" by Alberto Cairo
- "Better Data Visualizations" by John Schwabisch
- "Storytelling with Data" by Cole Nussbaumer Knaflic
- "Data Visualization Handbook" by Juuso Koponen & Jonatan Hildén
- "Fundamentals of Data Visualization", an open-access book by Claus Wilke

Blogs

- "Nightingale", the journal of the Data Visualization Society
- "Chartable", the blog by DataWrapper

Chart Choice Helpers

- From Data to Viz
- DataViz Project
- Visualizaiton Universe
- Material Design

Color Choice Helpers

- Viz Palette by Elijah Meeks & Susie Lu
- "How to pick more beautiful colors for your data visualizations" by Lisa Charlotte Muth/Data Wrapper