Clear figures, stronger stories

Scientific workflows: Tools and Tips 🛠️

Dr. Selina Baldauf

2025-04-17

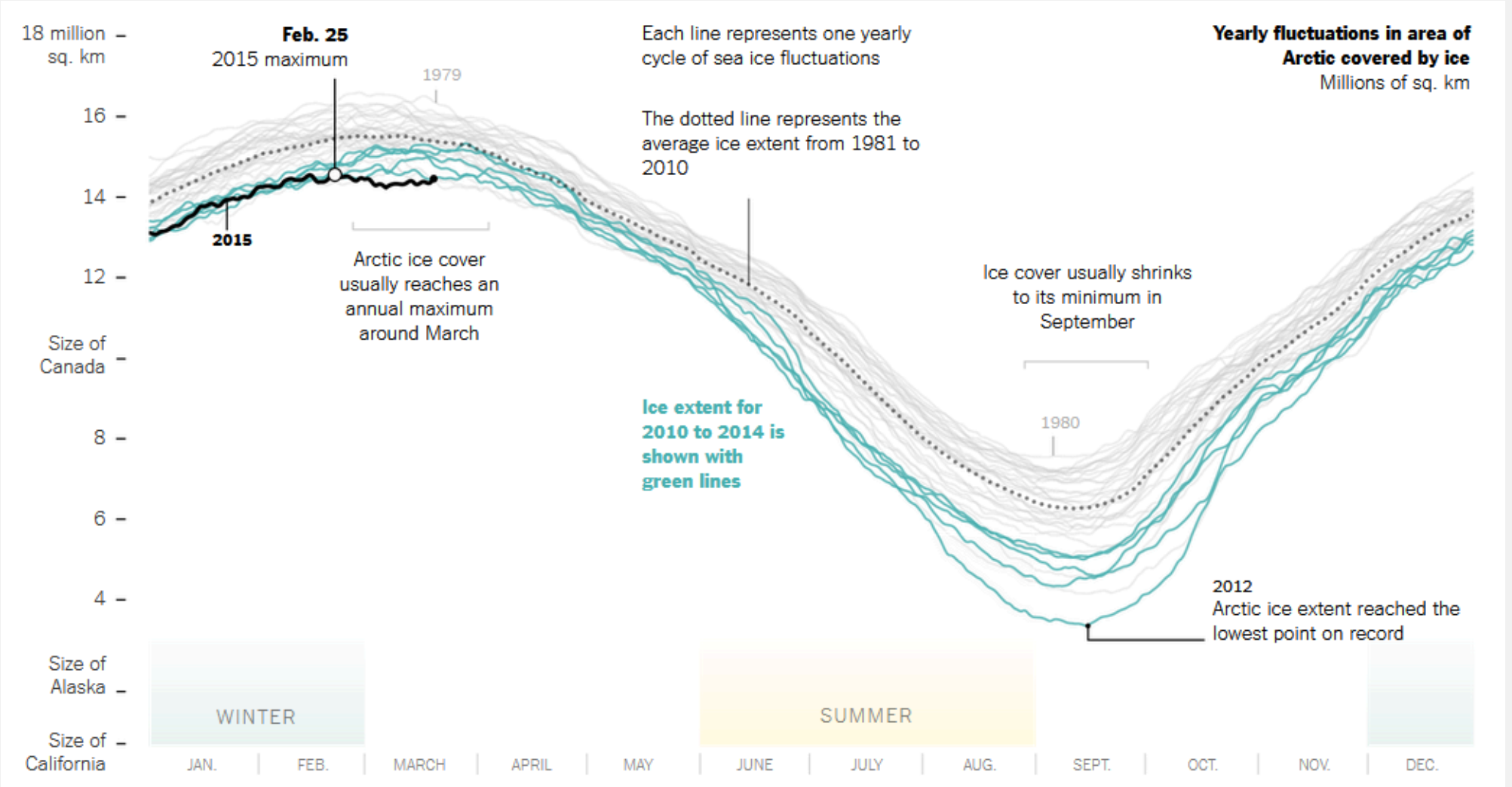
What is this lecture series?

Scientific workflows: Tools and Tips 🛠️

🗓️ Every 3rd Thursday 🕓 4-5 p.m. 📍 Webex

- One topic from the world of scientific workflows
- Material provided online
- If you don't want to miss a lecture
  - Subscribe to the mailing list

Motivation



Annual changes in Arctic sea ice cover by Derek Watkins (New York Times)

What makes a good figure?

- **Correct and transparent**
  Truthful representation of the data, data integrity
- **Useful**
  Supports the main point you want to make
- **Easy to read and understand**
  Accessible for everyone
- **Beautiful**
  Visually interesting and pleasing
- **Appropriate**
  Different outlets have different requirements/freedoms

7 steps for better figures

1: Consider the context

Consider the context

- **Who** is your **audience**?
  Familiarity with the topic
- **What** are **common practices** in your field?
  Established plot types, colors, …
- **Where** do you present your figure?
  Different contexts require different designs

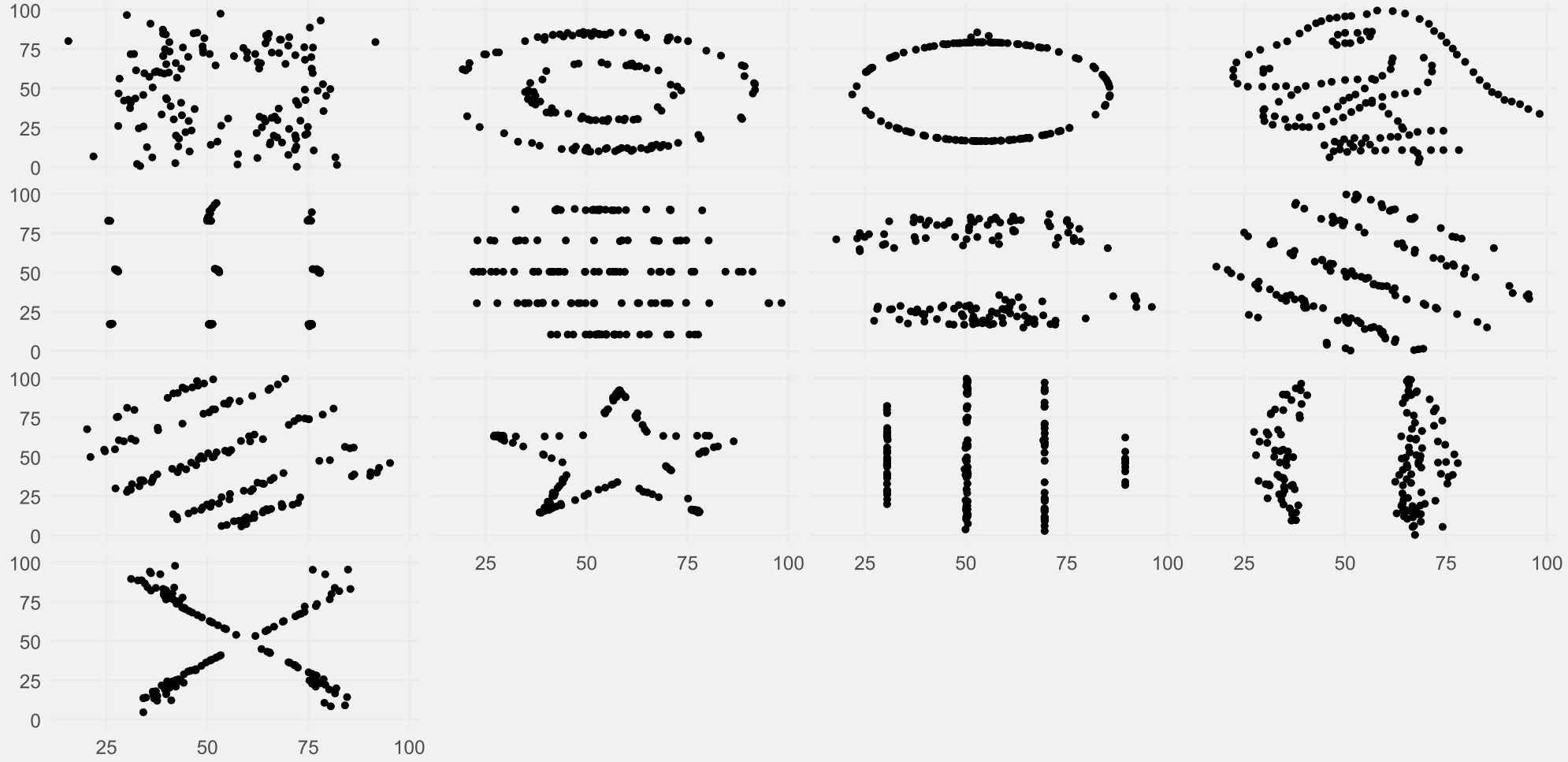
Contexts in science

Context	Things to consider
Paper	- Journal requirements - Usually read on PC but also print (B/W) - More time → Higher complexity
Poster	- More open design choice - Attract people from far - You quickly loose people to other posters - Medium complexity (depending on the event)
Talk	- You can use animations to guide through - Little time → less complex

👣 2: Make your data transparent

Don’t hide data behind summaries

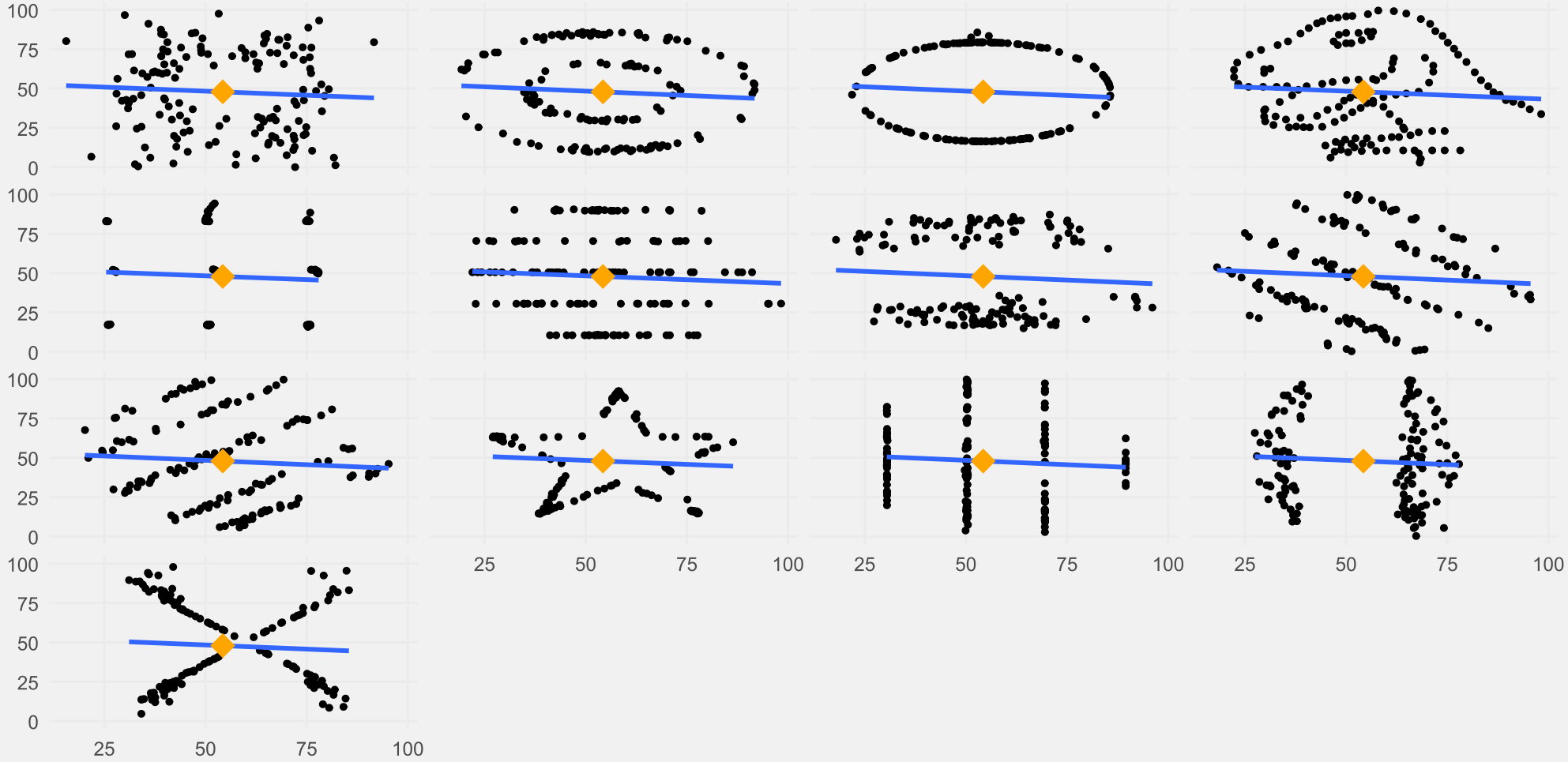
Datasaurus Dozen: Different distributions, same statistics



Data from the `datasauRus` package based on Matejka, J., & Fitzmaurice, G. (2017). Same Stats, Different Graphs.

Don’t hide data behind summaries

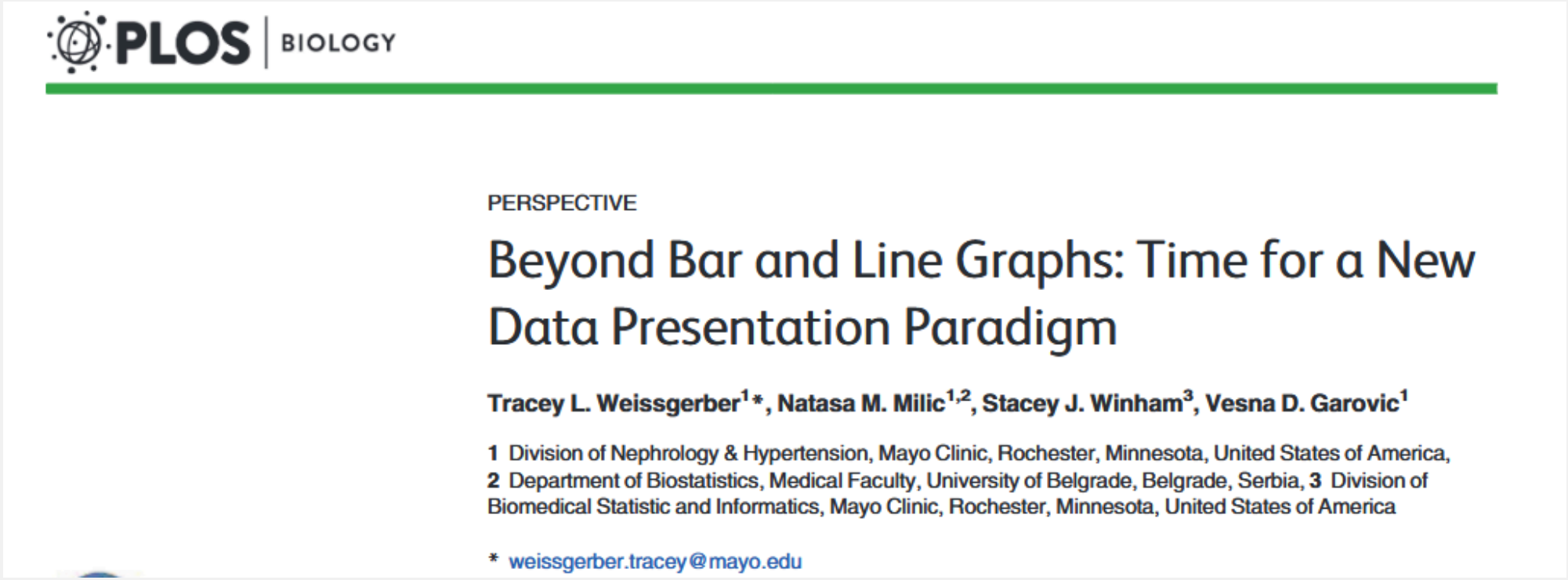
Datasaurus Dozen: Different distributions, same statistics



Data from the `datasauRus` package based on Matejka, J., & Fitzmaurice, G. (2017). Same Stats, Different Graphs.

Don’t hide data behind summaries

Bar graphs hide a lot of information about the data.



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]

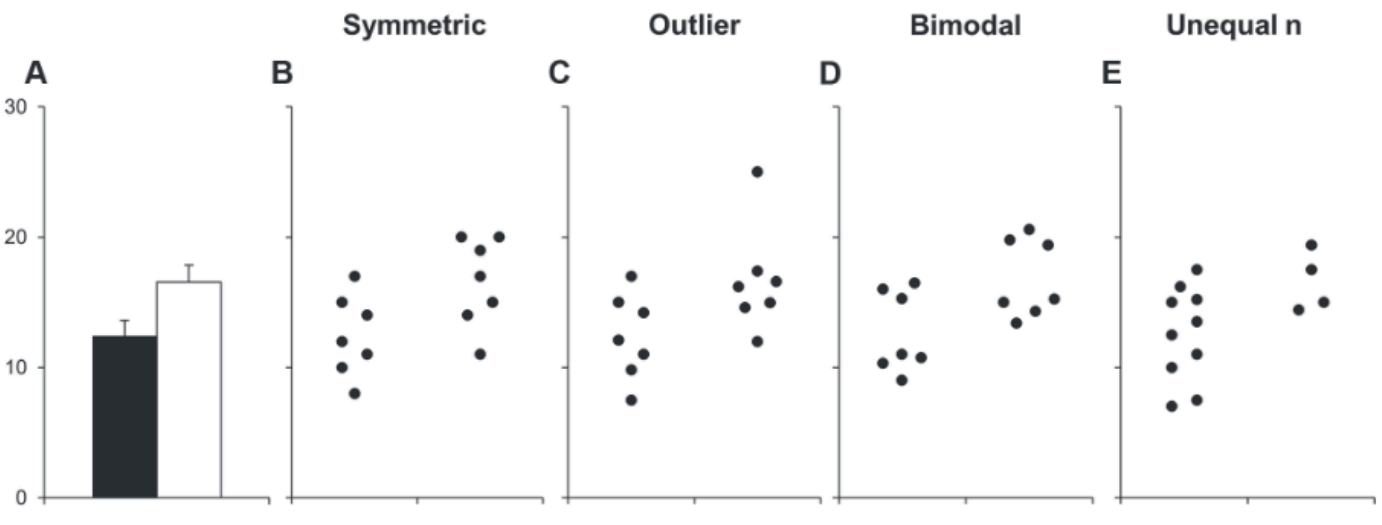
1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America, **2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States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

* weissgerber.tracey@mayo.edu

https://doi.org/10.1371/journal.pbio.1002128

Don’t hide data behind summaries

Same bar plot - different data & statistical test results

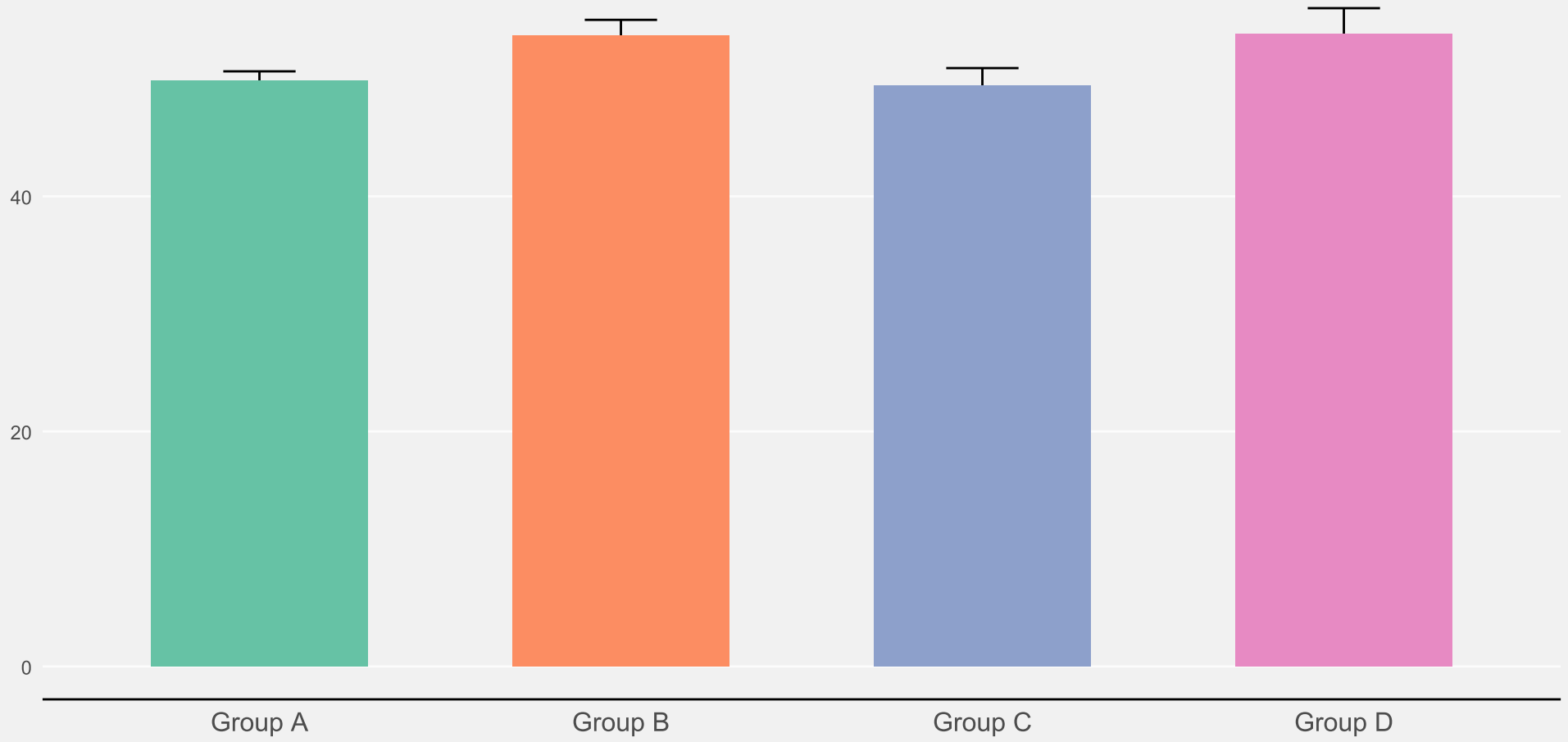


Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Figure 1 from Weissgerber et al. 2015

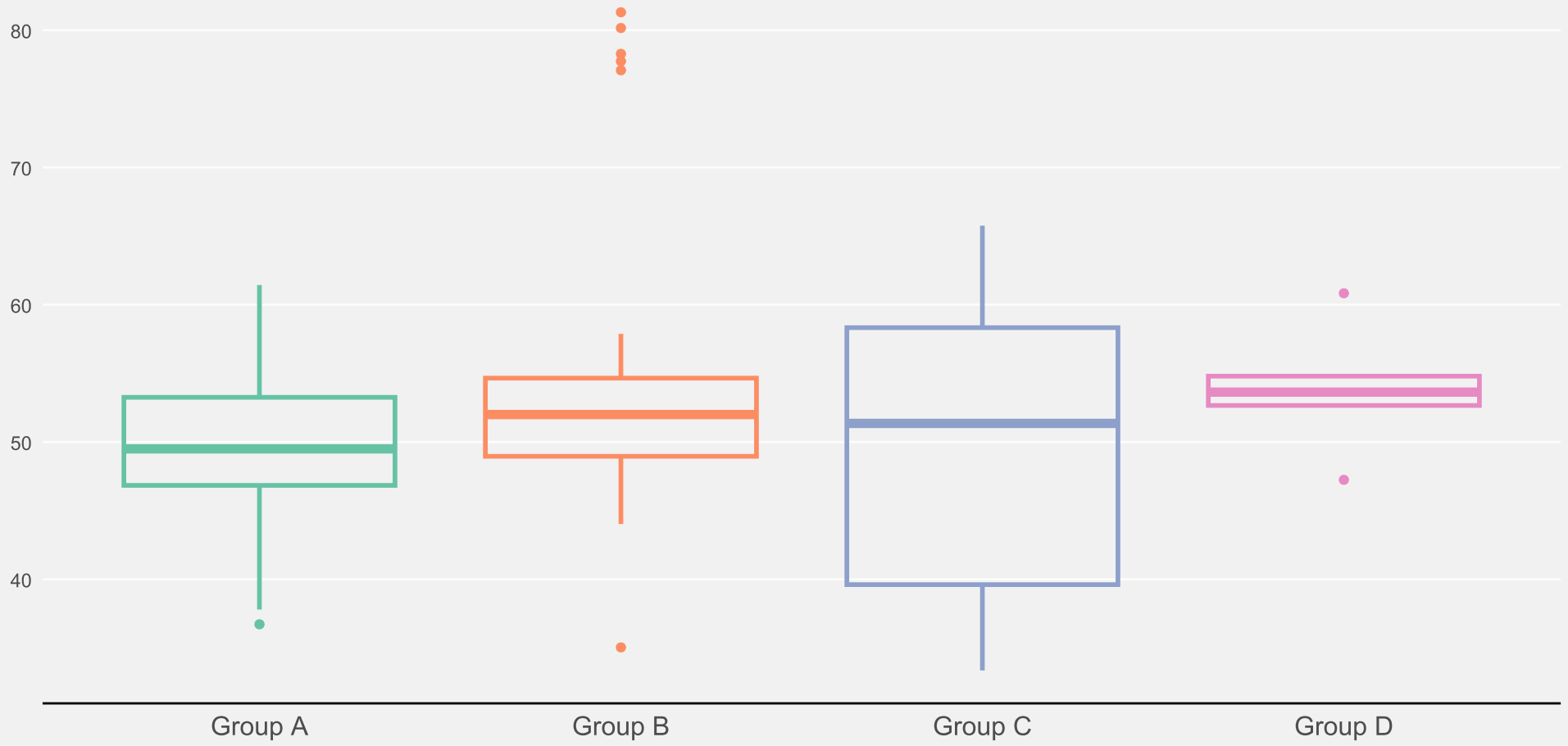
Alternatives to bar plots

Bar plots only show mean ± SE/SD.



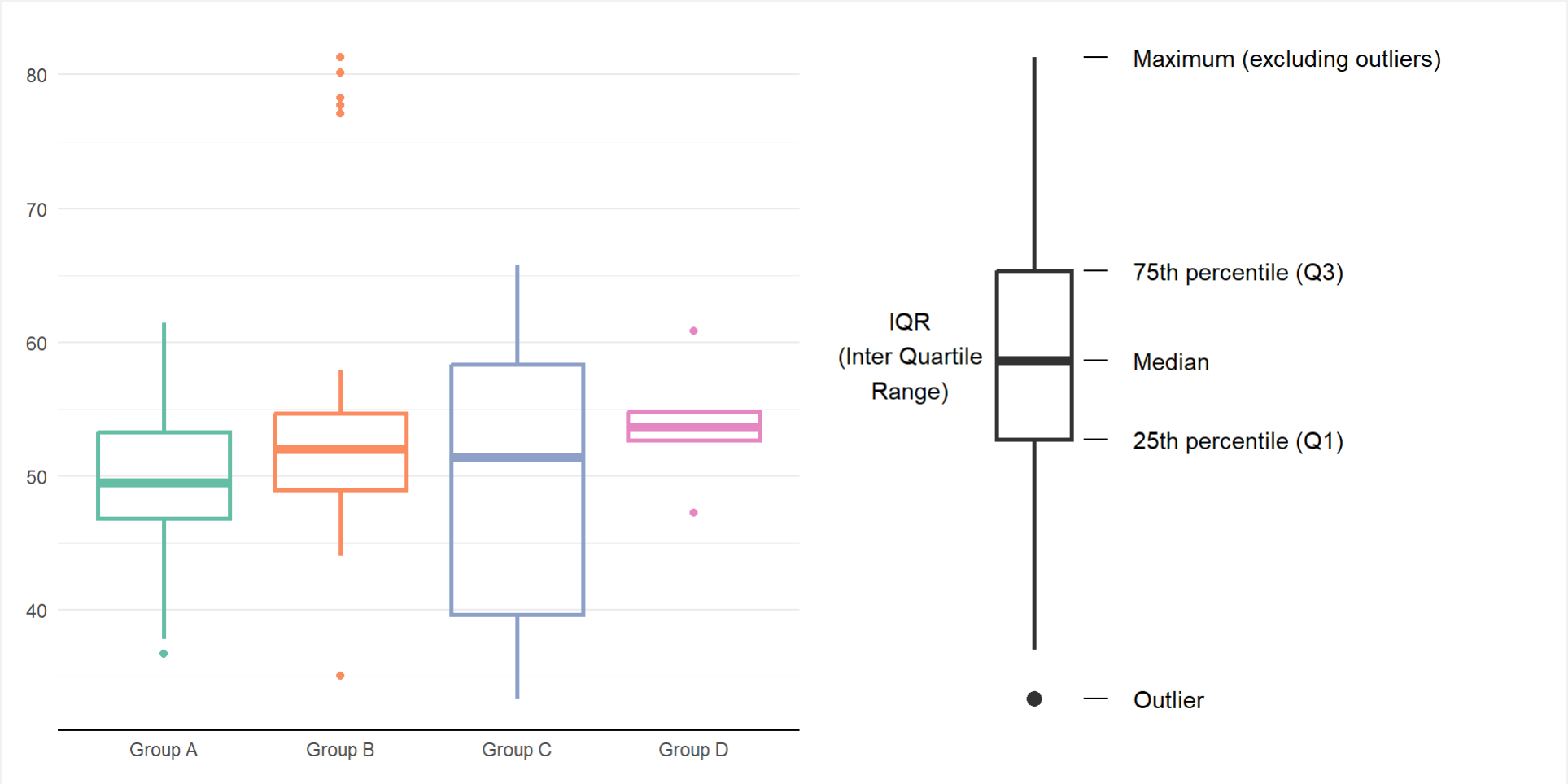
Alternatives to bar plots

A box plot is already better (shows more of the distribution)



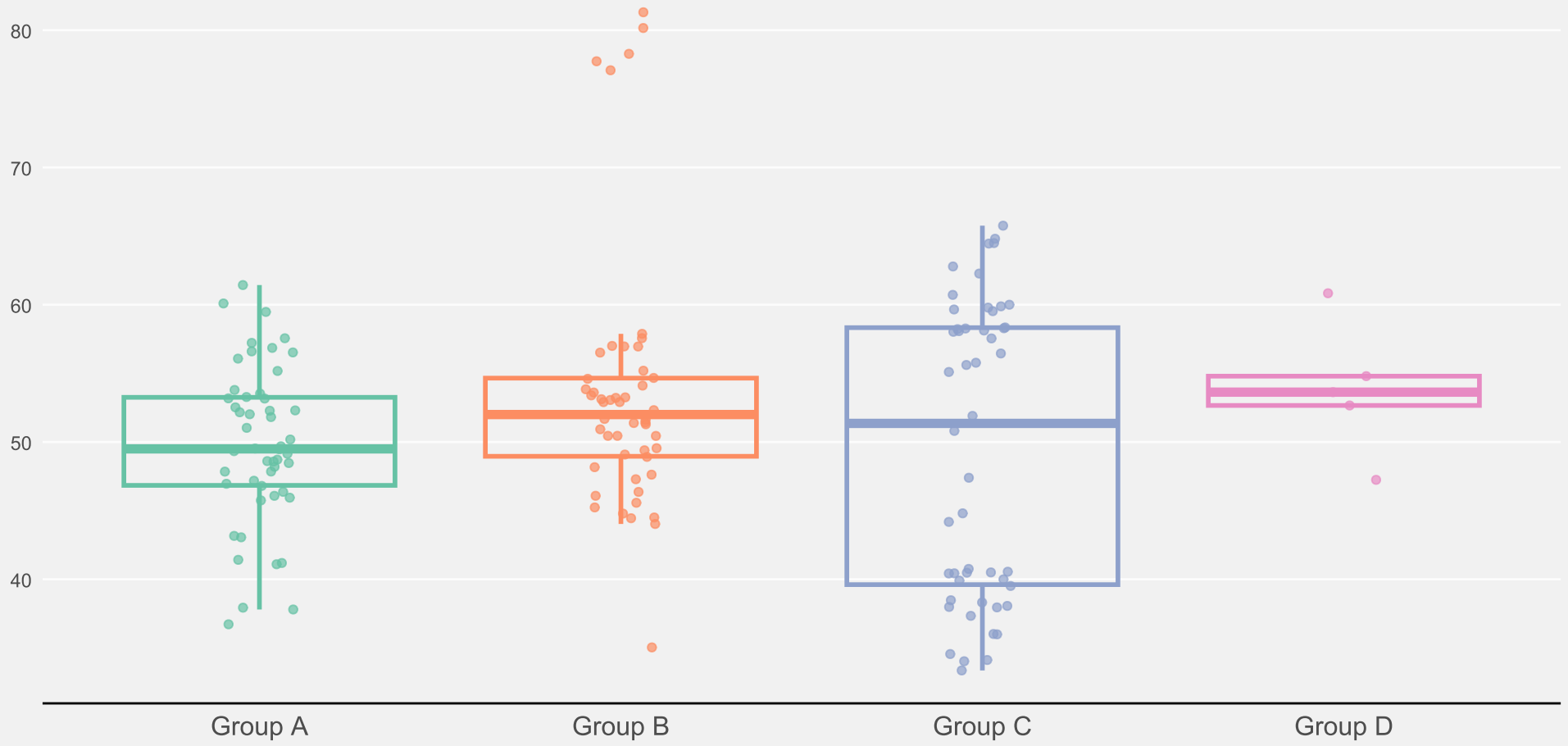
Alternatives to bar plots

A box plot is already better (shows more of the distribution)



Alternatives to bar plots

Add raw data points to increase the information content of the plot

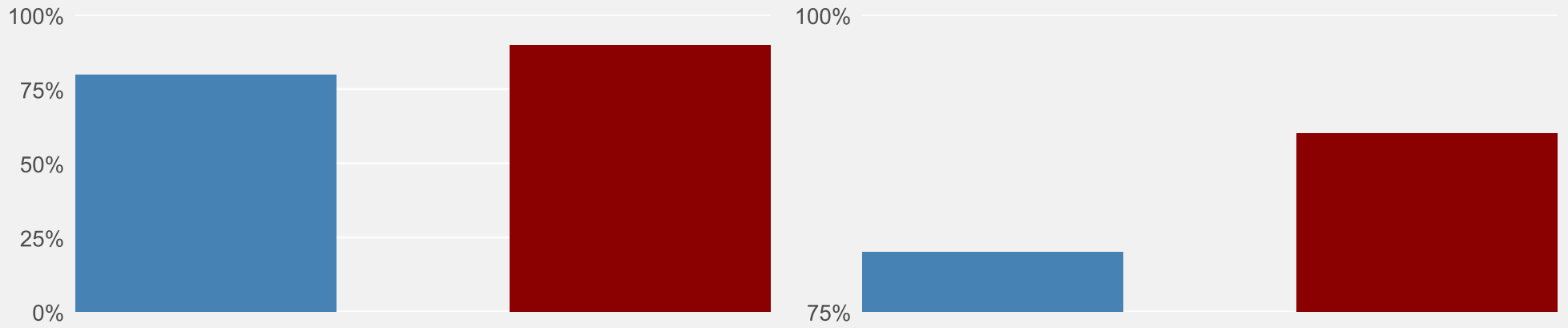


Alternatives to bar plots

Raincloud plots show raw data, summary stats, and distribution

Principle of proportional ink

Sizes of shaded areas should be **proportional** to the **data values** they represent

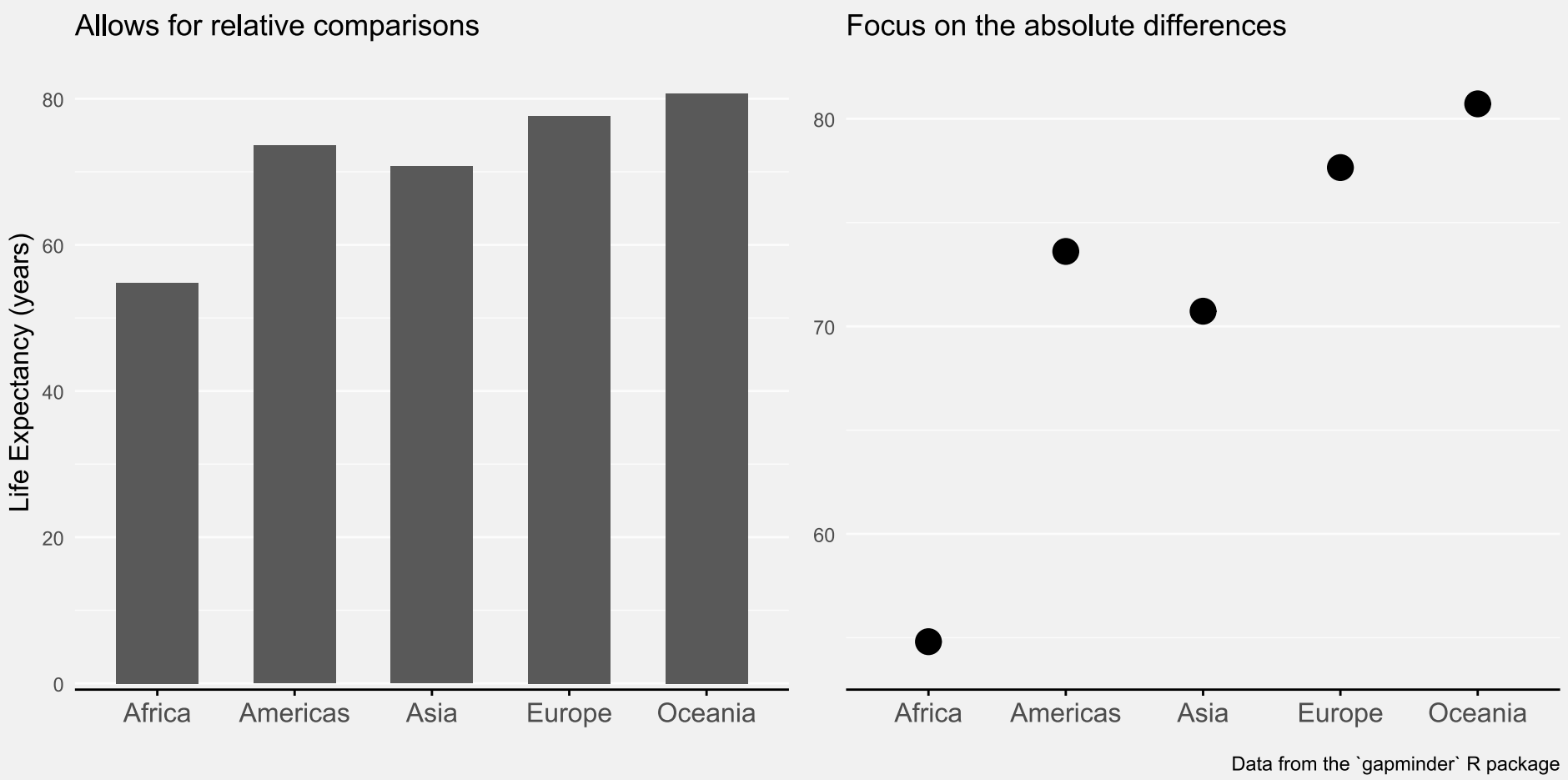


Here: bar length does not represent **relative data proportions** anymore

Always start bars at 0!

Principle of proportional ink

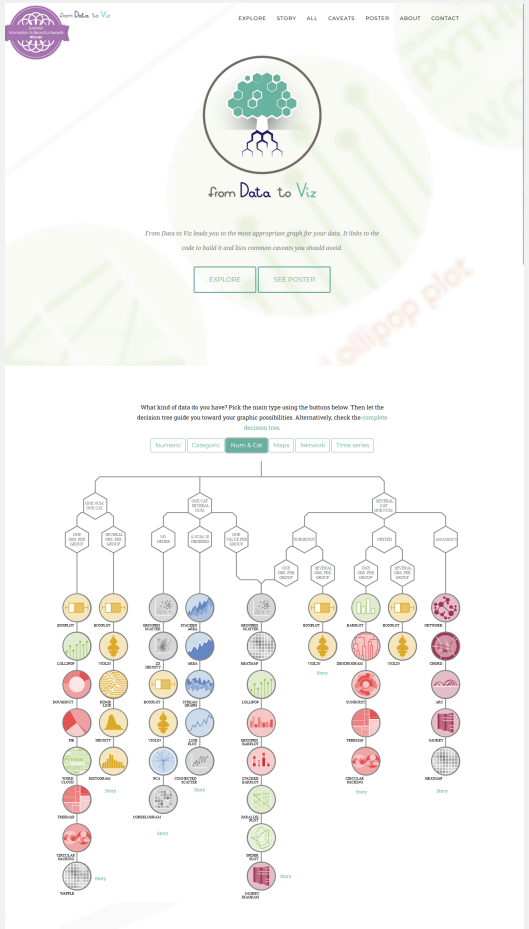
Other plot types don't have to start at 0.



👣 3: Choose the right chart type

Choose the right chart type

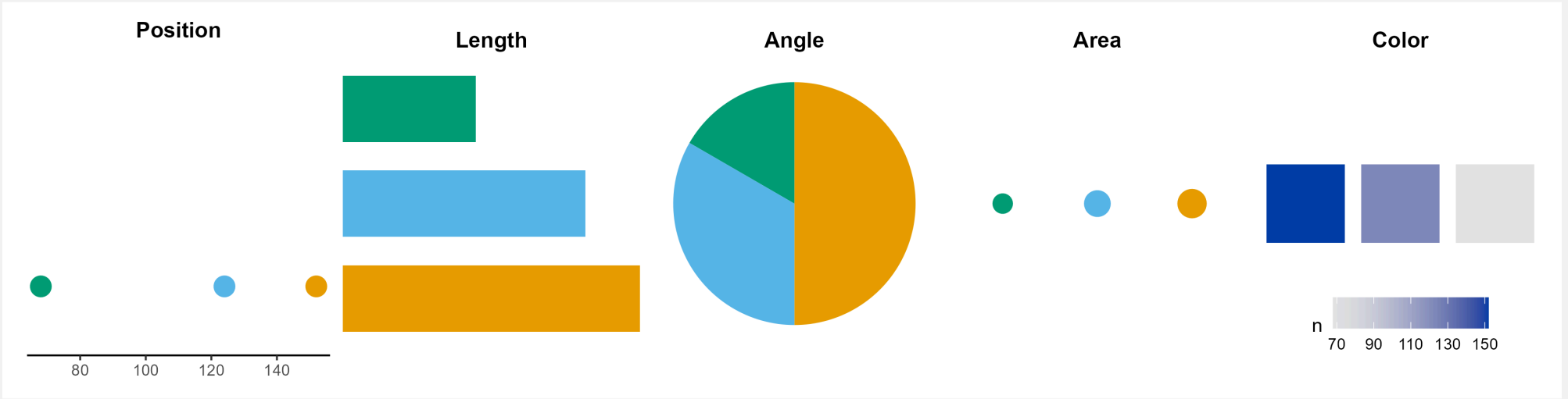
There are so many chart types - and cool tools to explore them



From data vo viz

The dataviz project

Different channels - different accuracy



- Accuracy of judgement decreases from left to right
- More accurate judgements vs. more generic judgements

Different channels - different accuracy

- Combine multiple channels for more accuracy
- Add numbers to increase accuracy of judgement

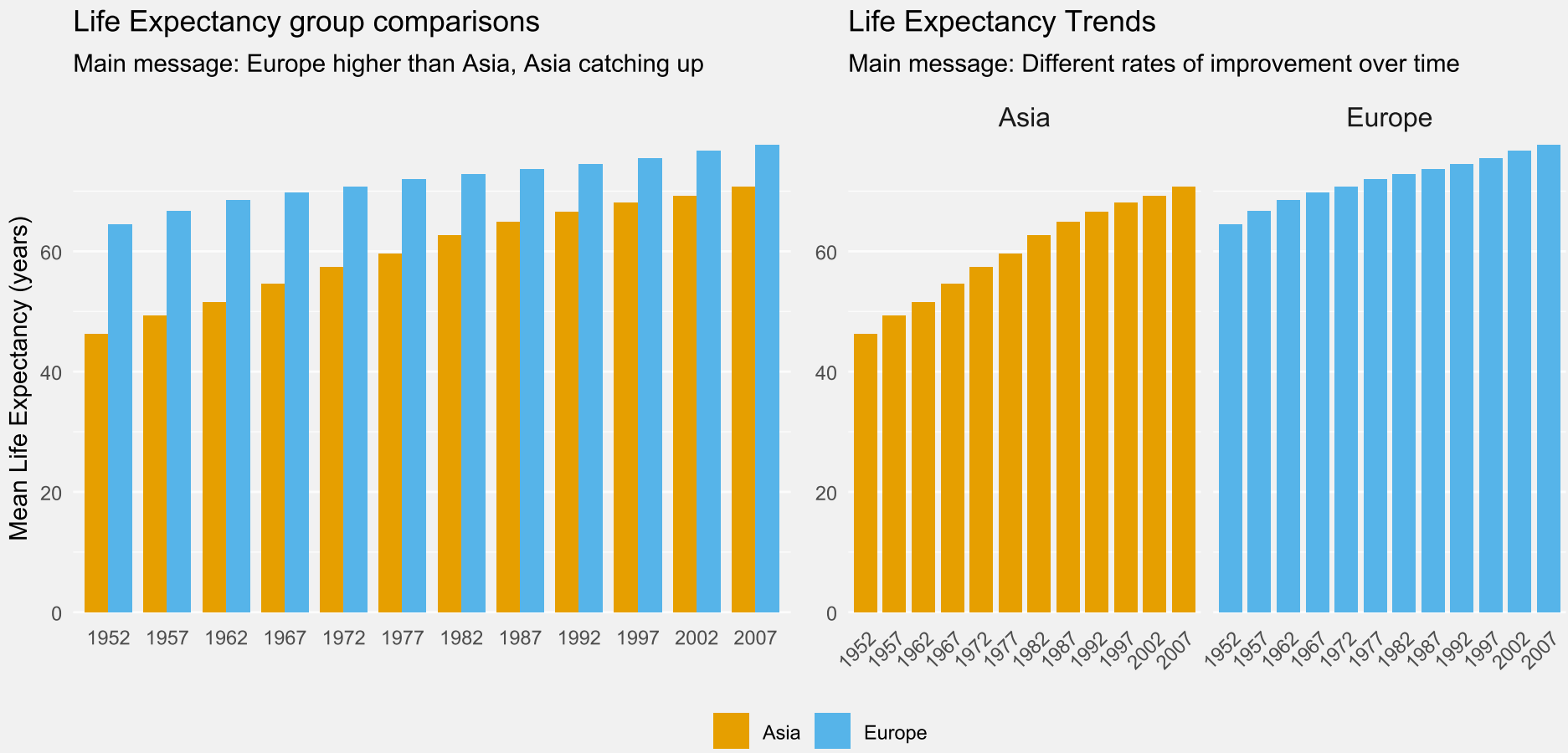
👣 4. Focus on the core message

Focus on the core message

- The readers attention is limited: **be concise**
- Think about the **main message** you want to convey
  - Which variables do you need?
  - What can you omit?

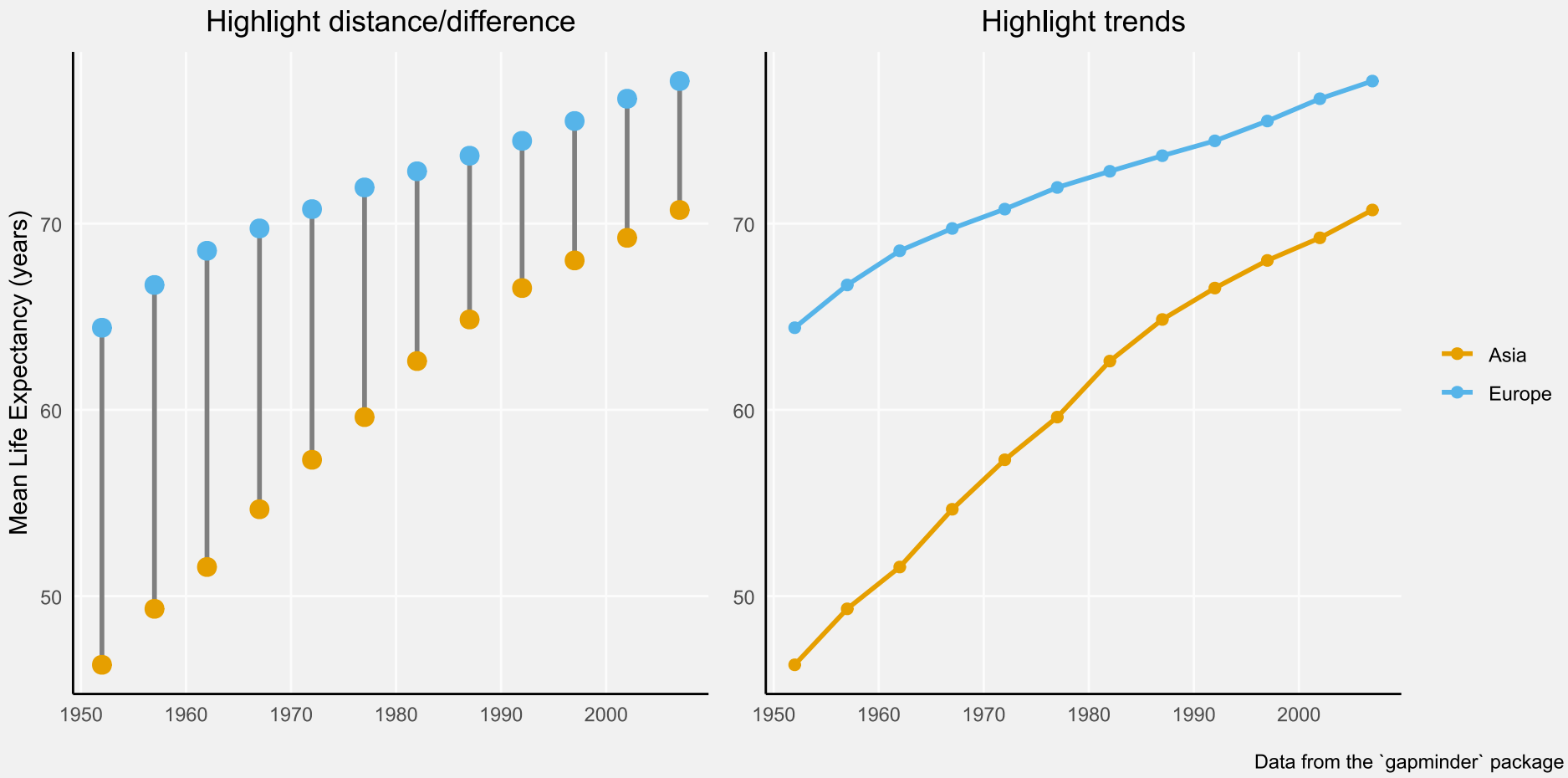
Arrange your plot

Arrange your plot so that it's **easy to extract** the main message



Choose a good plot type

Different plot types tell different stories



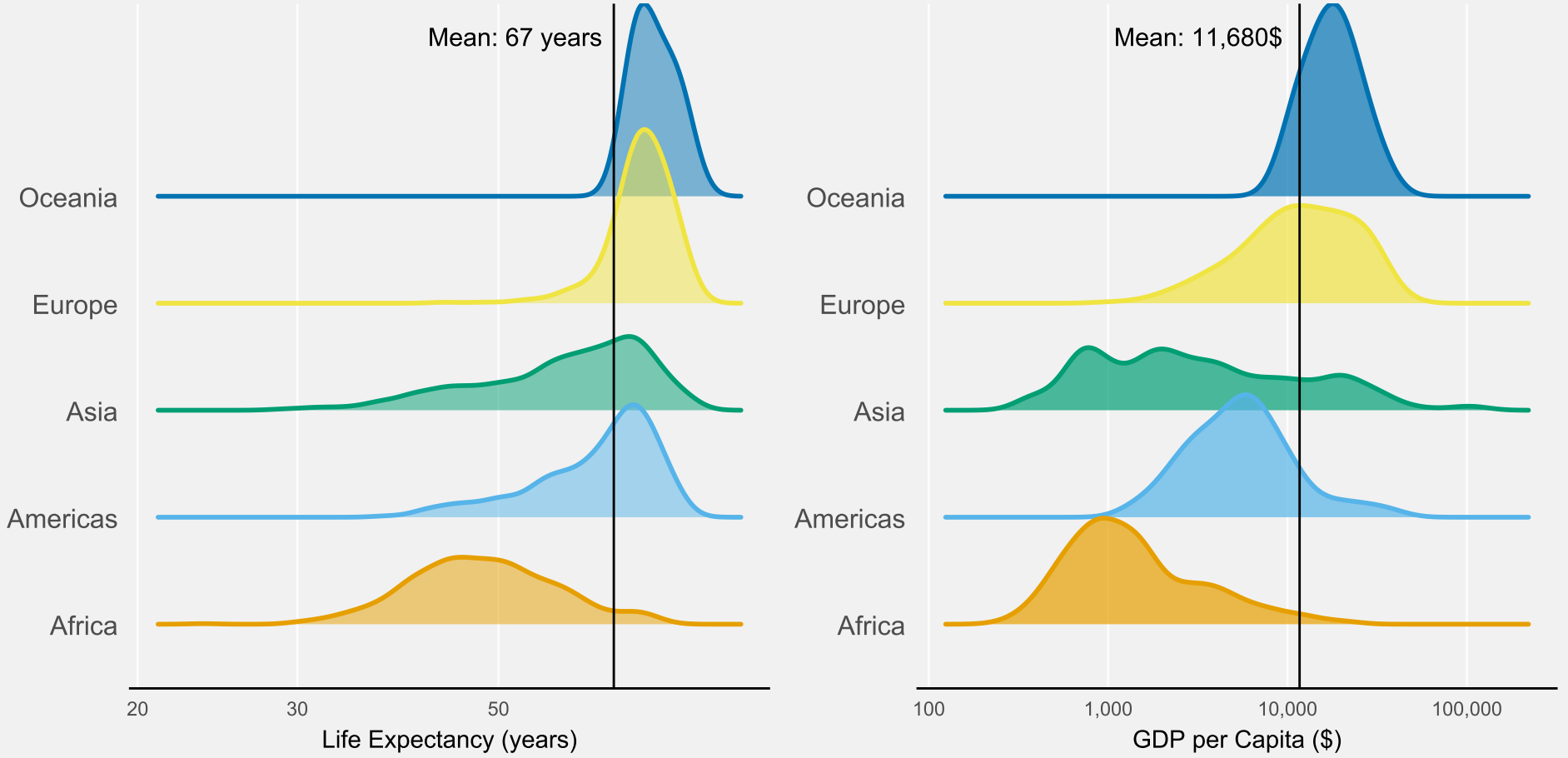
Keep it simple

Don't overcomplicate your figures and bury your message

What is the main message here?

Keep it simple

Message: Life expectancy and GDP differences in the world



5. Consider the trip

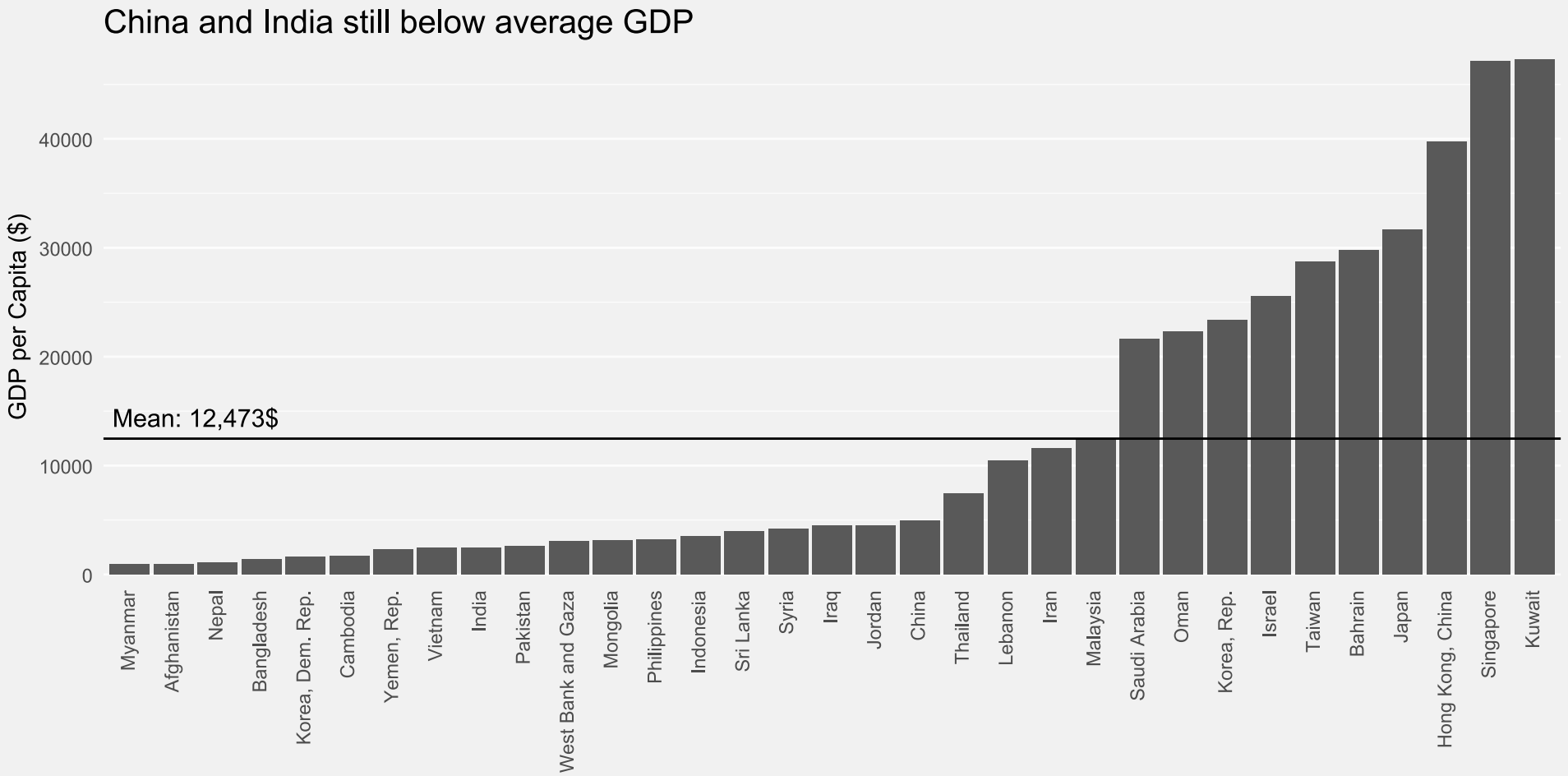
Reading a figure is a timely experience

- We look at elements **step by step** before we come back to understand the **figure as a whole**
- Put yourself in the readers shoes
  - What will they look at first, second, etc.?
  - **How many steps** does it take to understand all the elements?

Goal: Make the trip (for the eyes and the brain) as short as possible

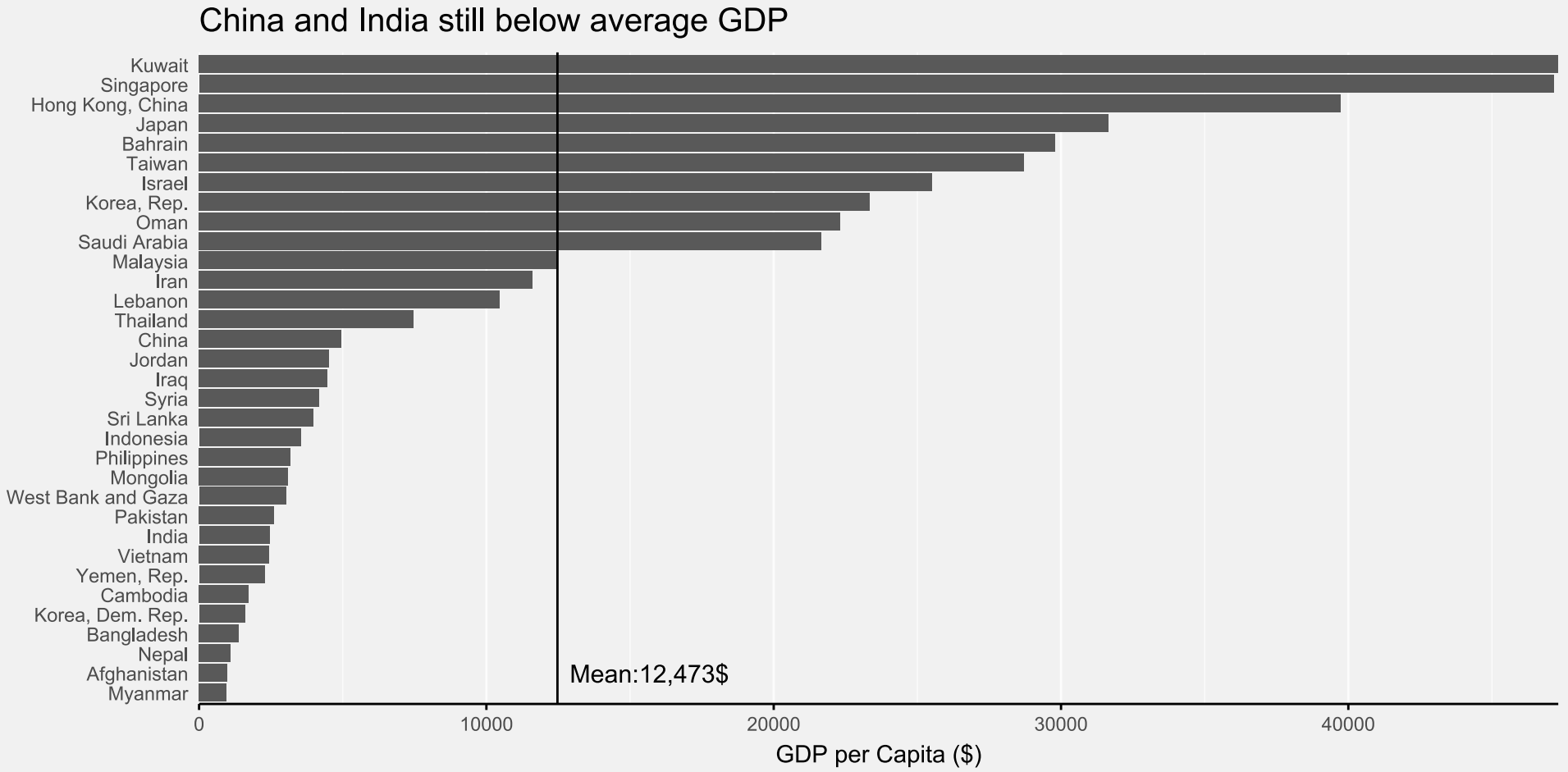
Consider the trip

A story about the GDP China and India



Rotate your plot

Reading labels upside down is a neck rotation - very annoying



Highlight the main message

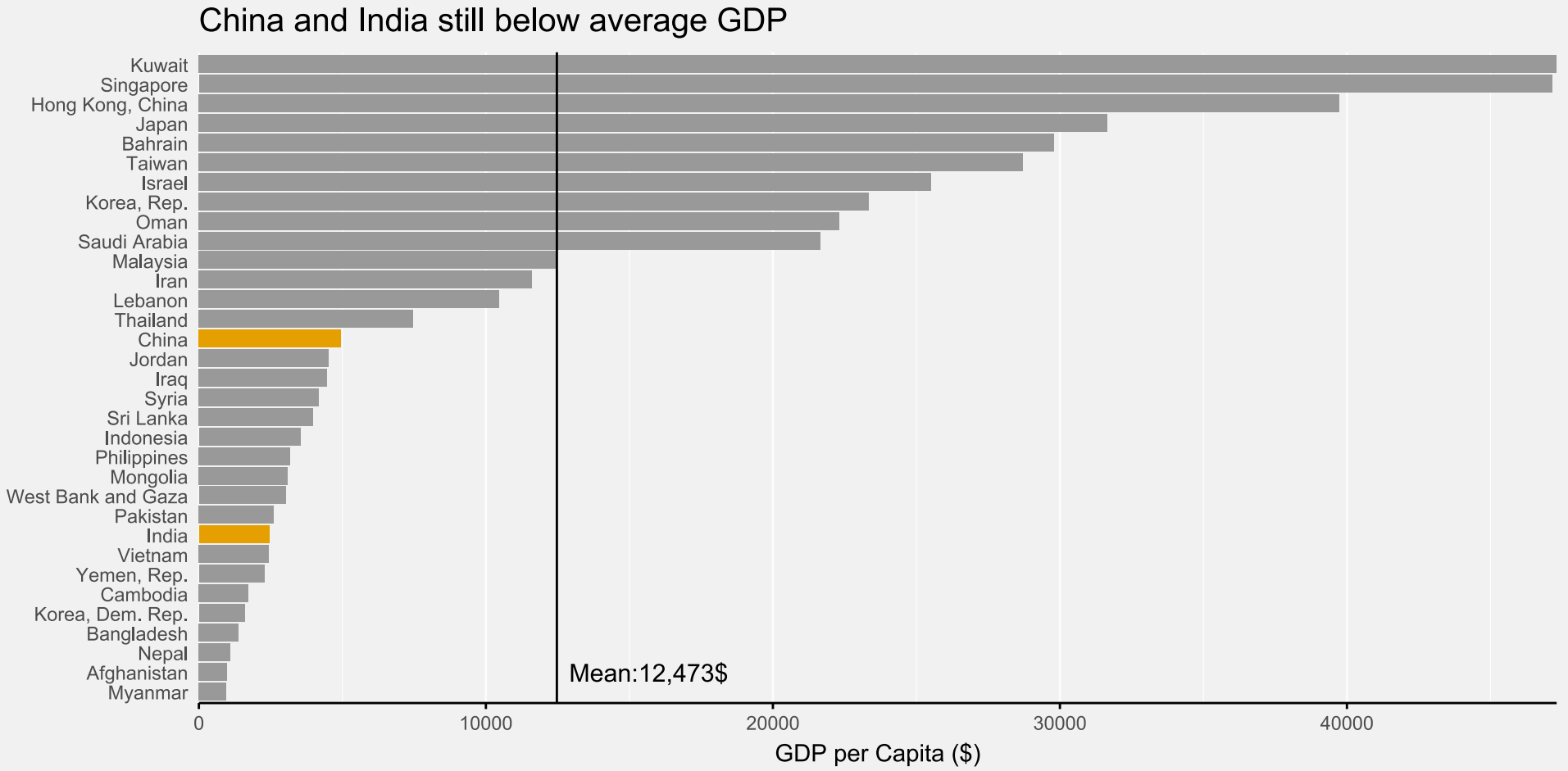
Use highlighting and de-emphasizing

Effective visualization helps us understand data quickly. Patterns emerge naturally, while colors enhance meaning. Good design choices and proper emphasis make insights accessible to everyone.

- Make use of **pre-attentive focus** (Things that *pop* out)
- Possible highlights: color, size, shape, arrows, …

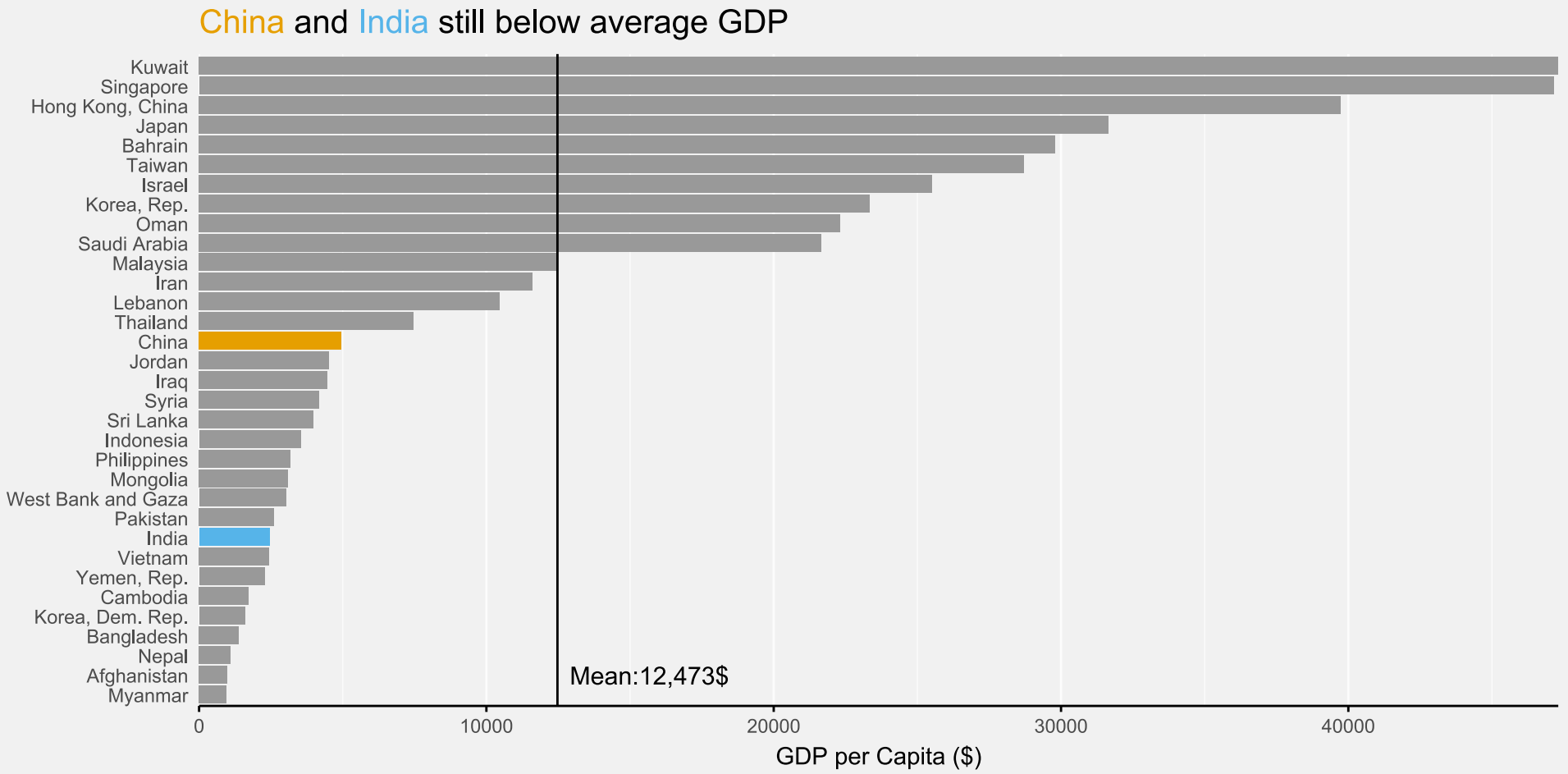
Highlight the main message

Highlight focus contries, de-emphasize all others



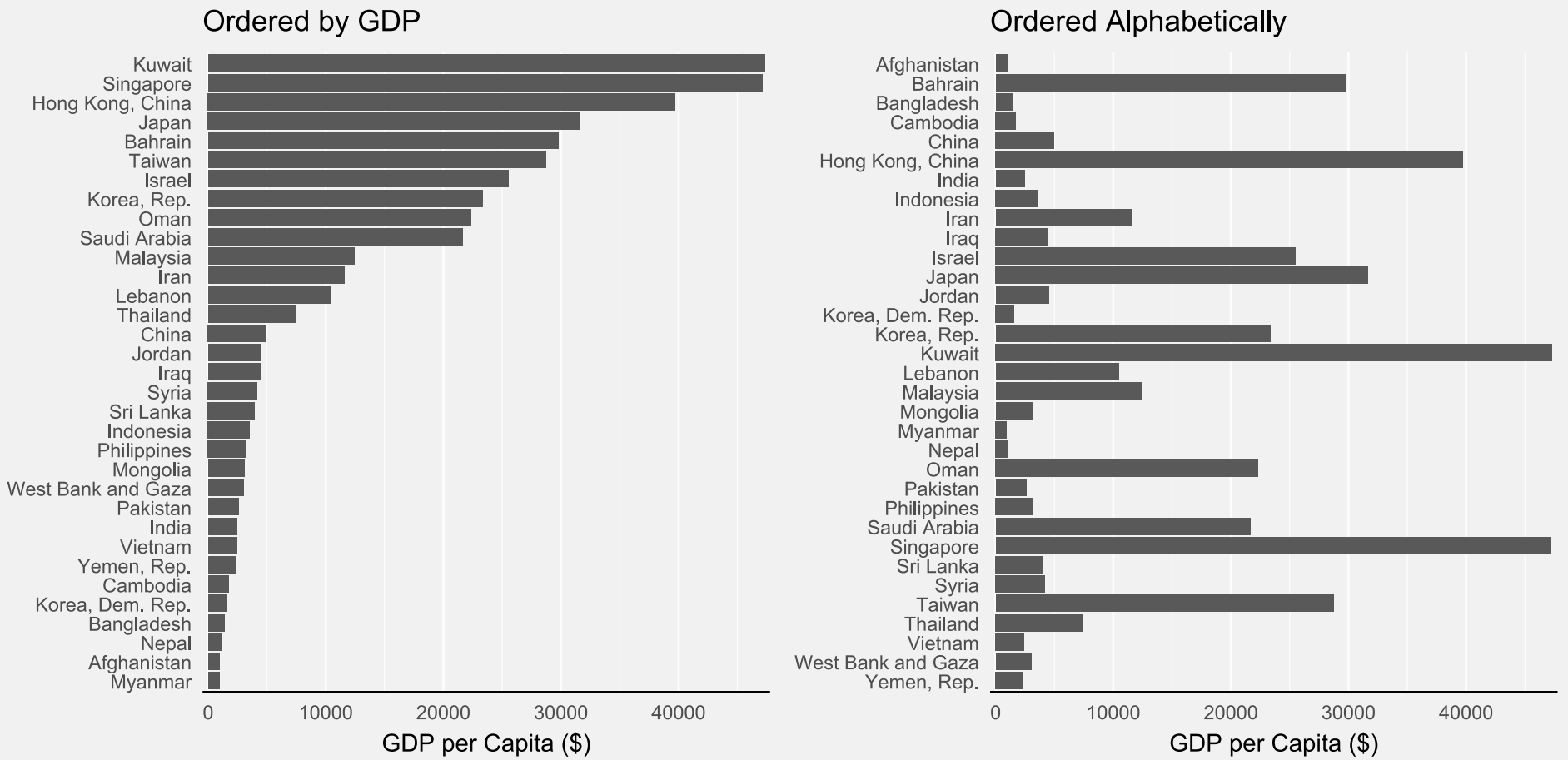
Highlight the main message

Highlight focus contries, de-emphasize all others



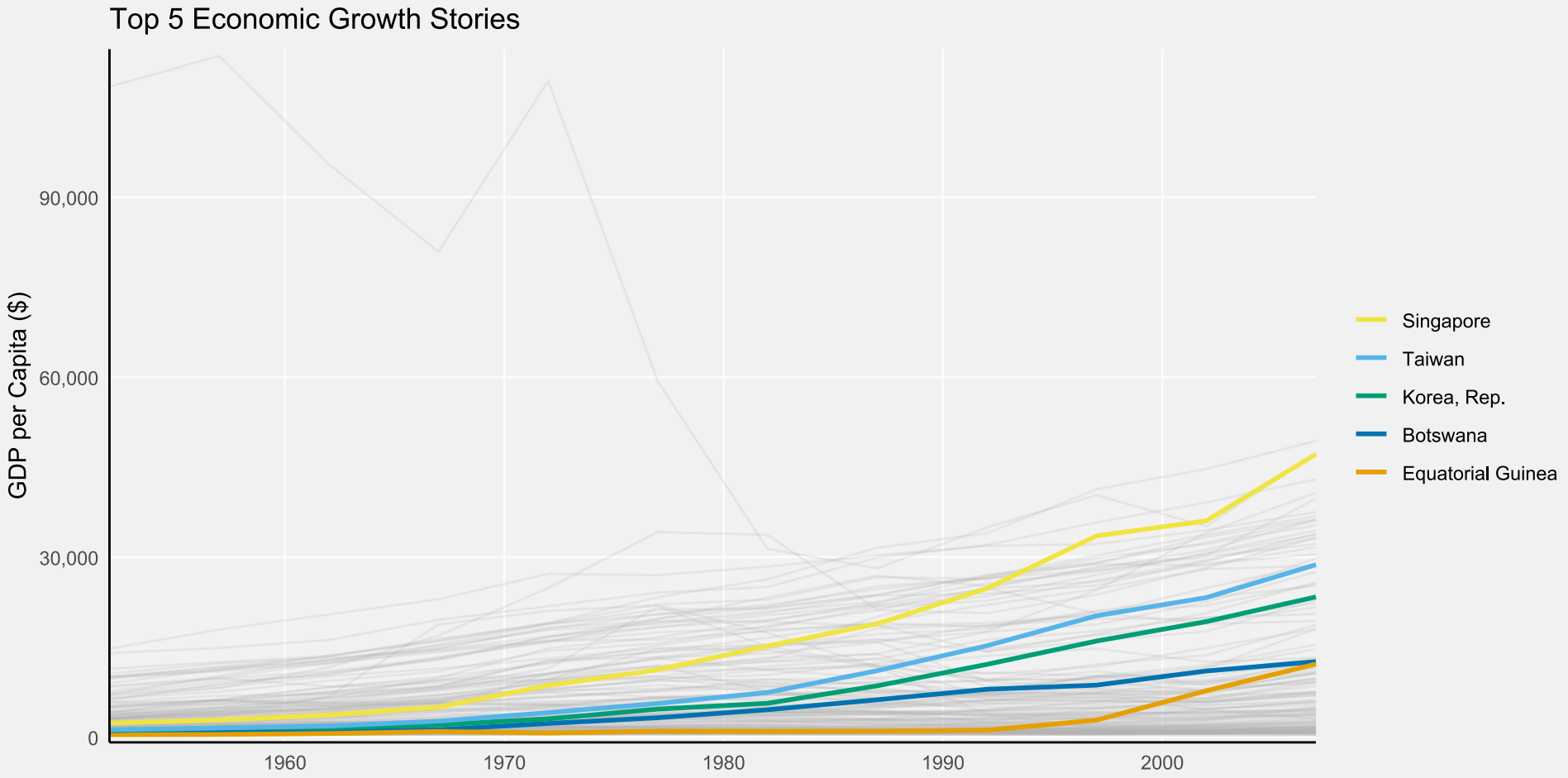
Order your data

Order categories consciously not automatically.



Order your data

Order categories consciously not automatically.



6. Less is more

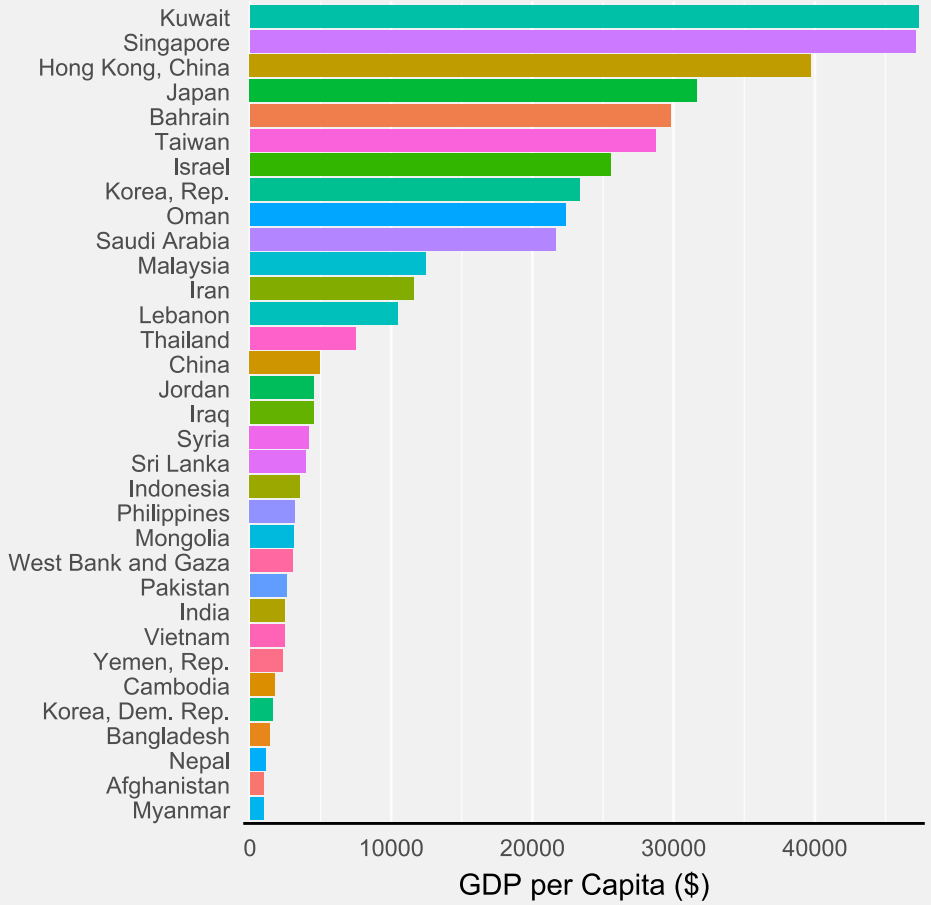
The importance of differences

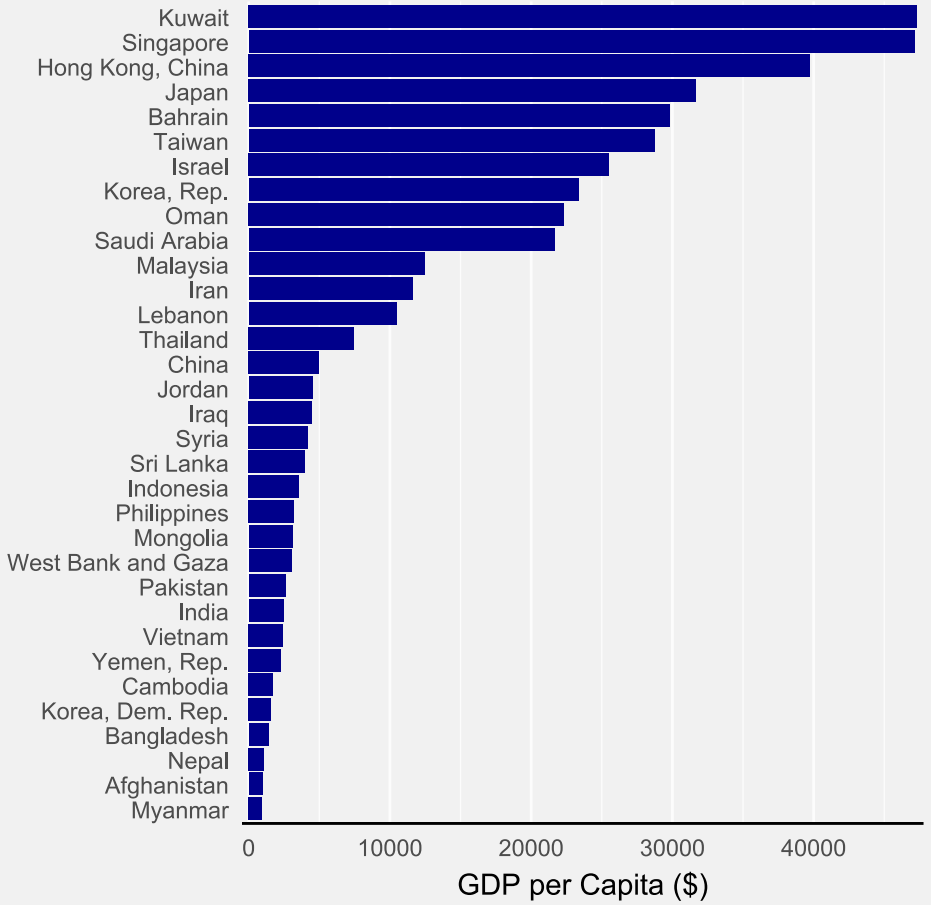
Effective visualization helps *us* understand data *quickly.* Patterns emerge naturally, *while* colors enhance meaning. *Good* design choices *and* proper emphasis make *insights* accessible to *everyone.*

Effective visualization helps us understand data quickly. Patterns emerge naturally, while colors enhance meaning. Good design choices and proper emphasis make insights accessible to everyone.

The importance of differences

Use differences to **communicate not to decorate**

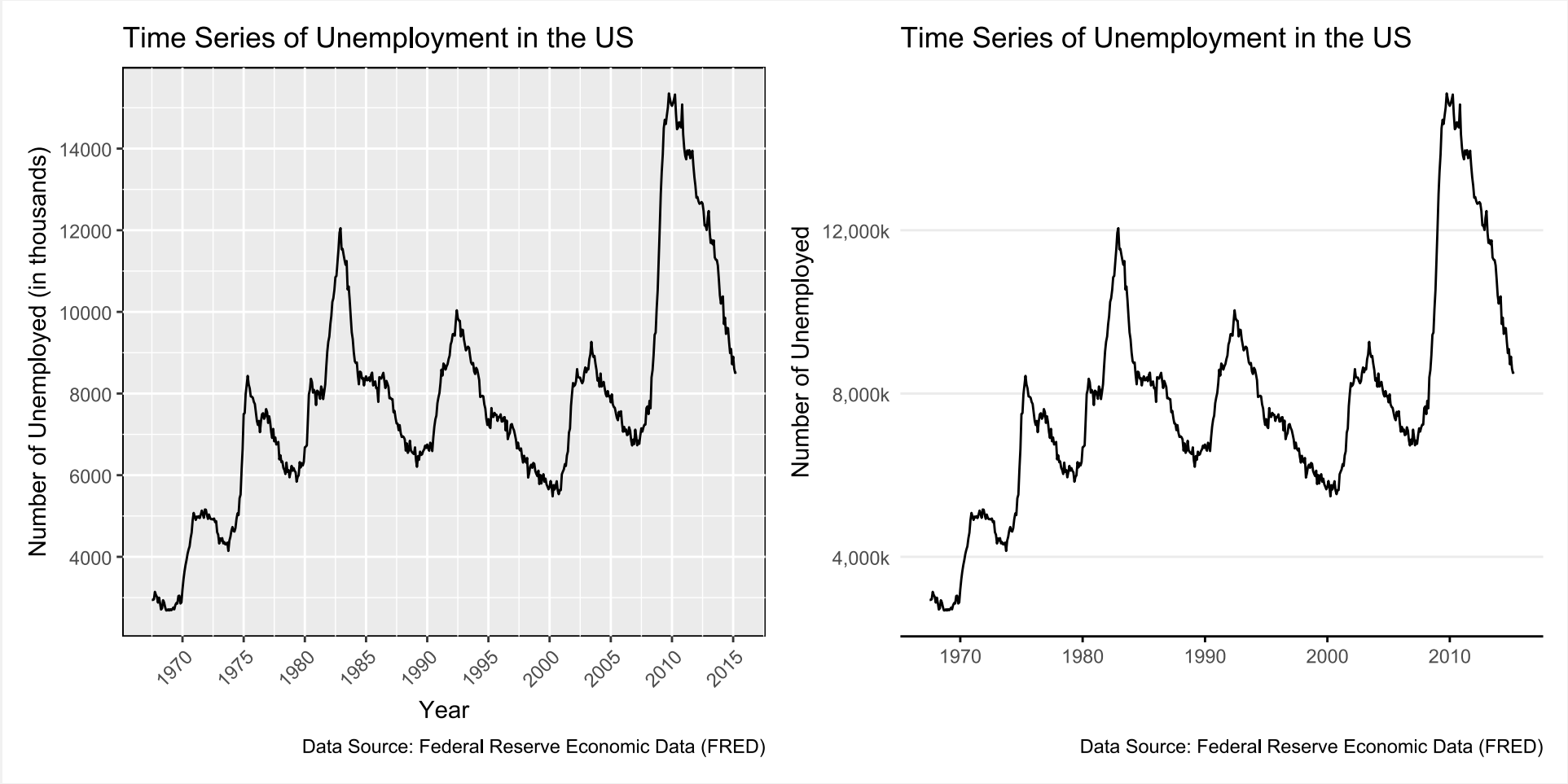




Declutter your figure

- Try to maximize the **data/ink ratio**
- This is to an extent a matter of taste
- Remove redundant figure elements
  Excessive grid lines, boxes, duplicate text …
- But keep elements important for reading

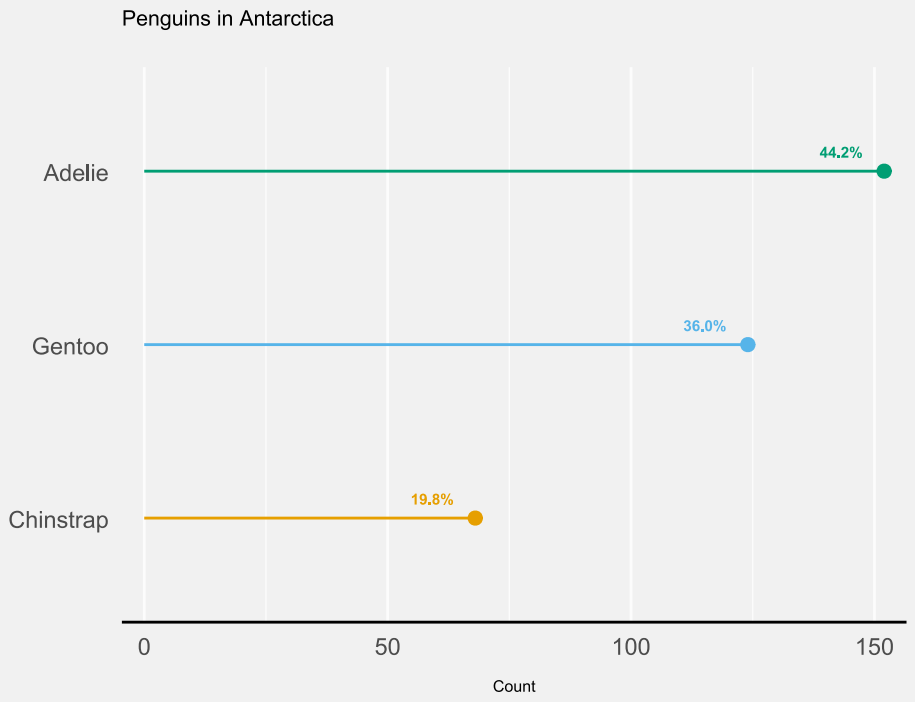
Declutter your graphs

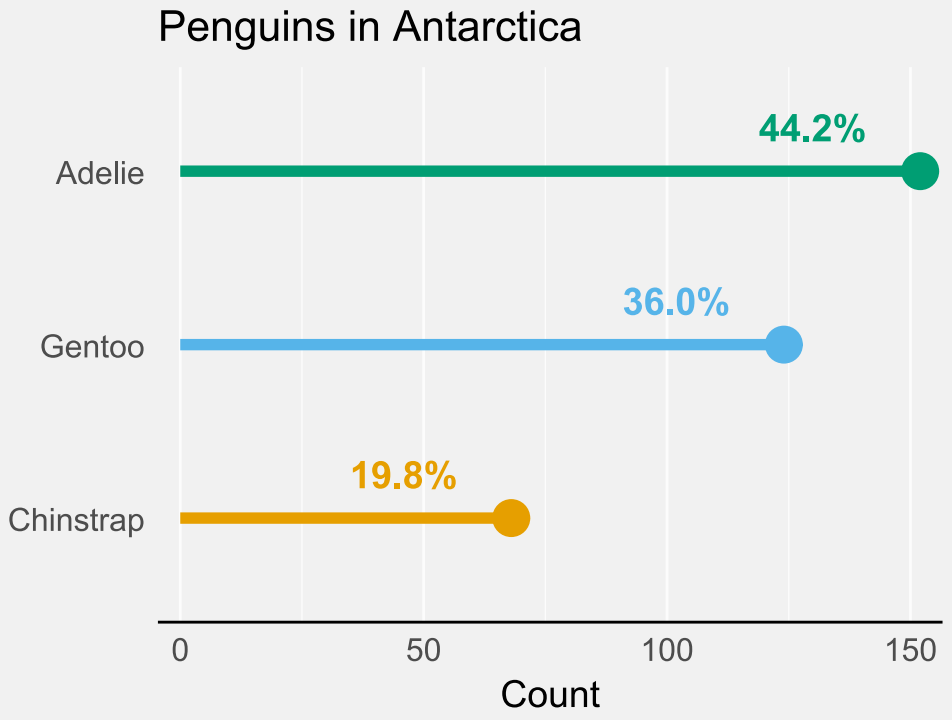


7: Make it accessible

Element size

- Make sure your elements are big enough
  Text size, Linewidth, Point size
- Depends on the context





Contrast

Make sure that the contrast is high enough

Blogpost on colors by Lisa Charlotte Muth (Datawrapper)

- Use tools to check contrast, e.g. https://snook.ca/technical/colour_contrast

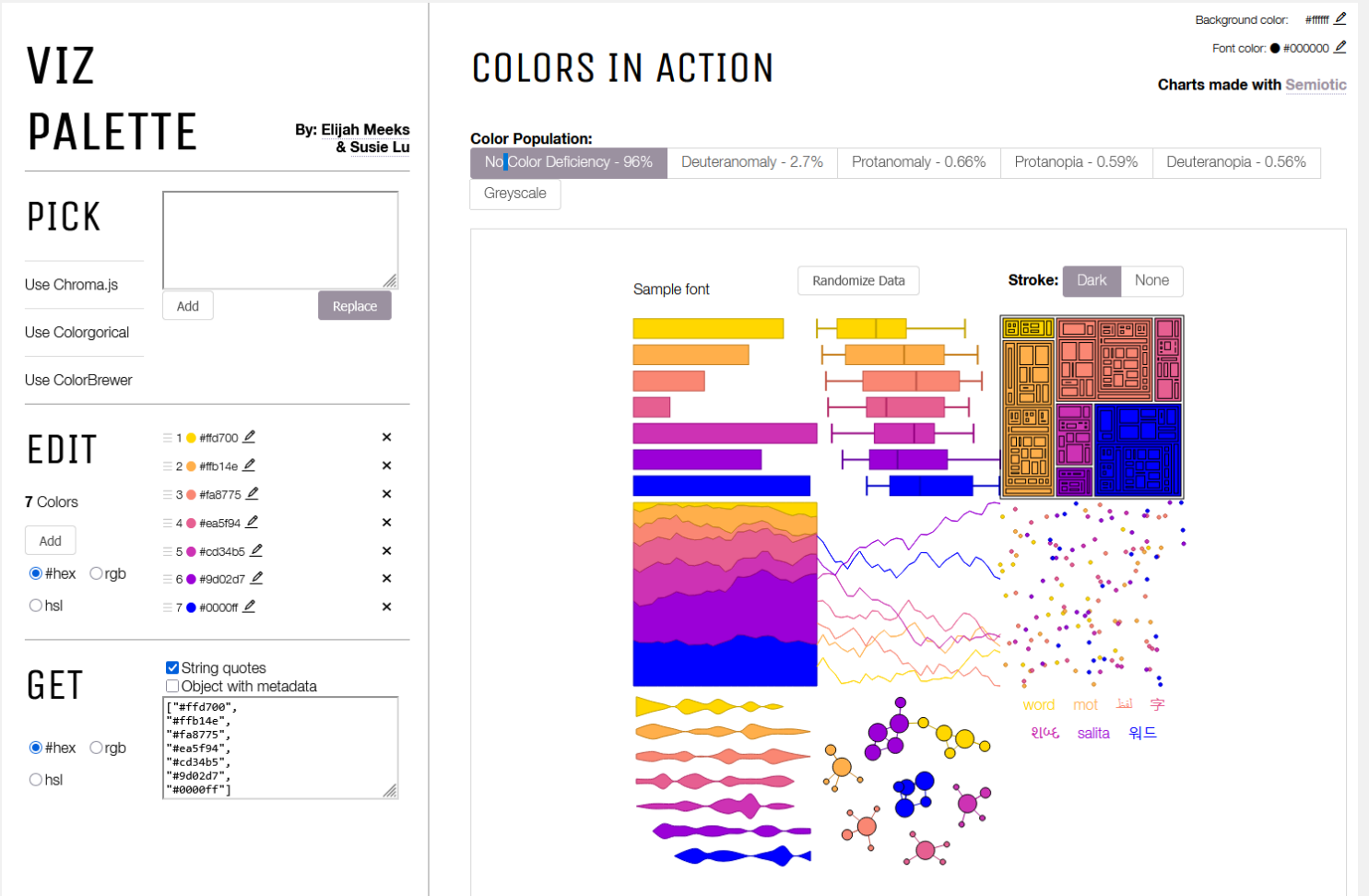
Color

Use logical/intuitive colors

Blogpost on colors by Lisa Charlotte Muth (Datawrapper)

Color

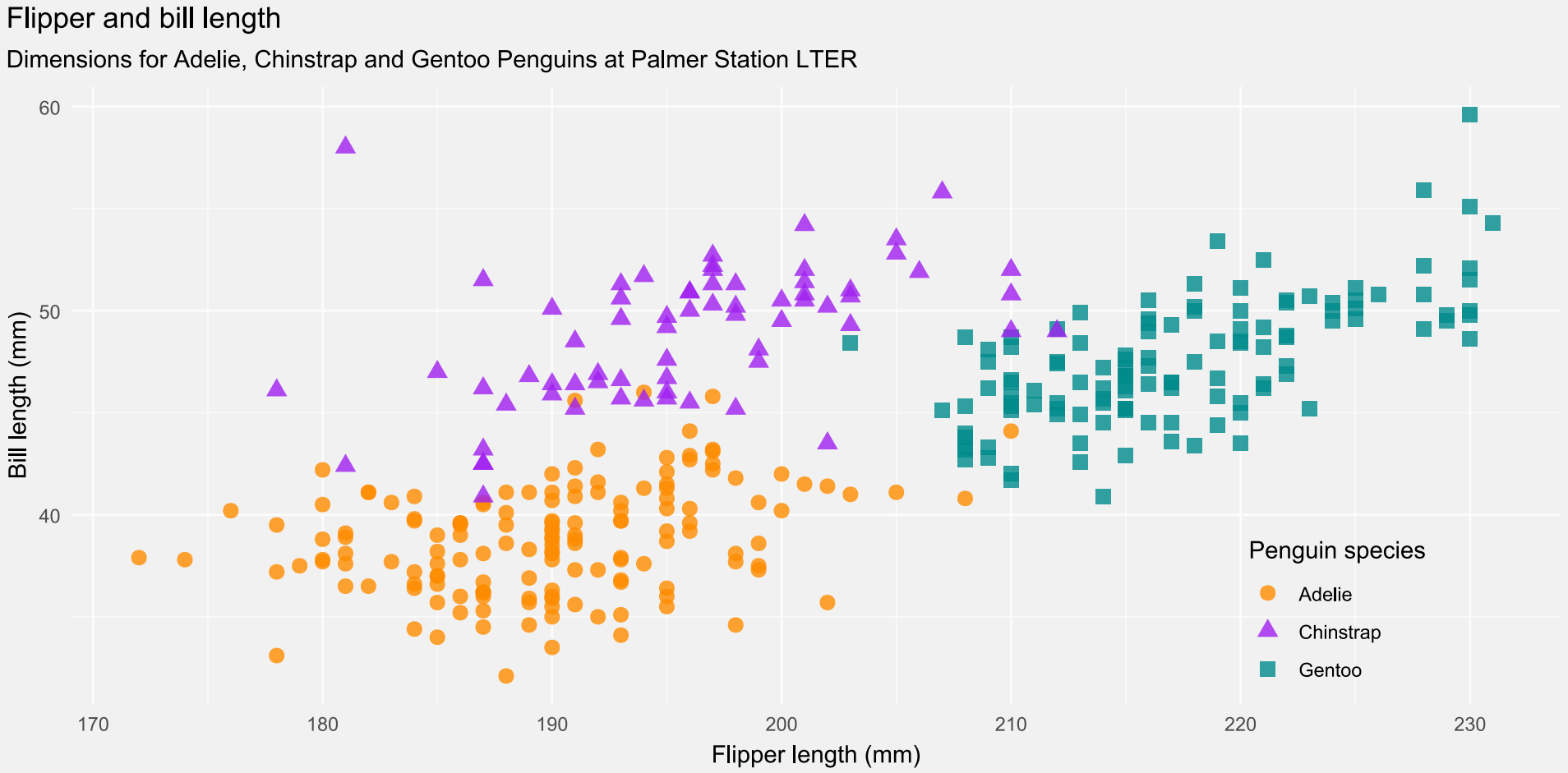
Choose colorblind friendly palettes (if in doubt: test!).



Use the Viz Palette tool to test for color blindness

Add redundancy

Redundancy increases the chance that everyone can see the difference!



Summary

1. Consider the **context**
2. Make your data **transparent**
3. Choose the **right chart type**
4. Focus on the **core message**
5. Consider the **trip**
6. **Less is more**
7. Make it **accessible**

Start analyzing these points in yours and other people's plots.

Next lecture

Topic t.b.a.

📅 15th May 🕒 3 - 4 p.m. 📍 Webex

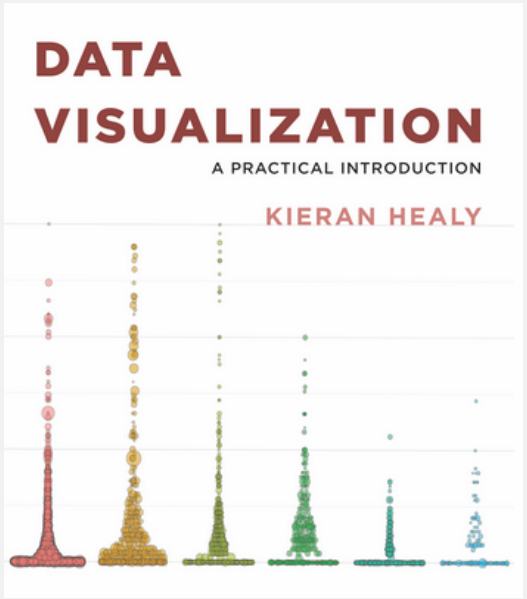
🔔 Subscribe to the mailing list

📧 For topic suggestions and/or feedback send me an email

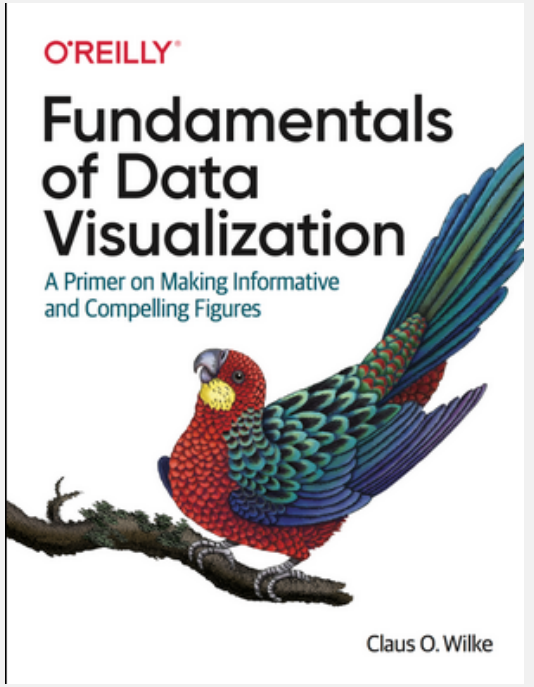
Thank you for your attention :)

Questions?

References



Healy, K. (2018). Data Visualization: A Practical Introduction. Princeton University Press.



Wilke, C. O. (2019). Fundamentals of Data Visualization. O'Reilly Media.