

# Clear figures, stronger stories

Scientific workflows: Tools and Tips 

Dr. Selina Baldauf

2025-04-17

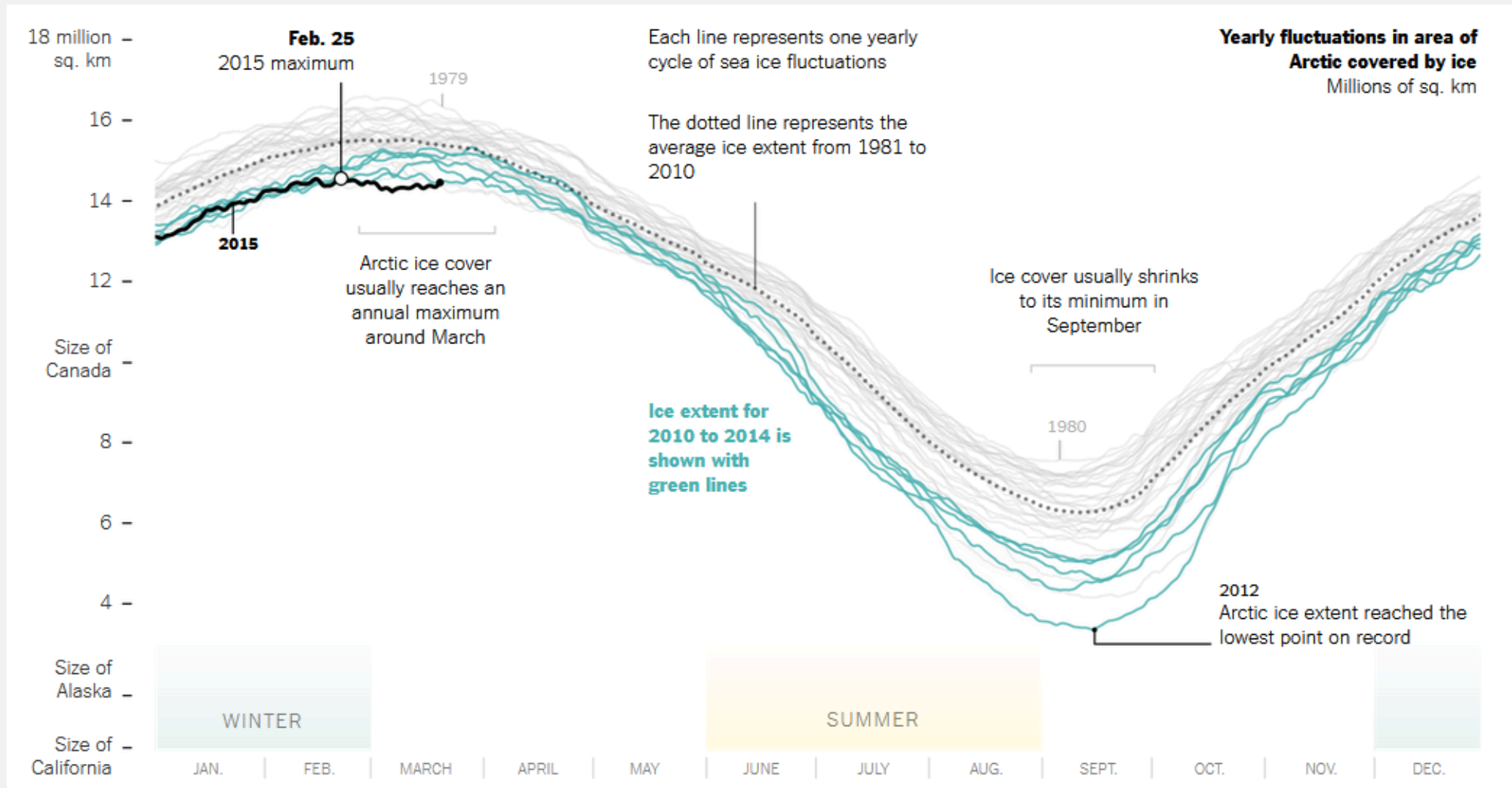
# What is this lecture series?

## 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
- 

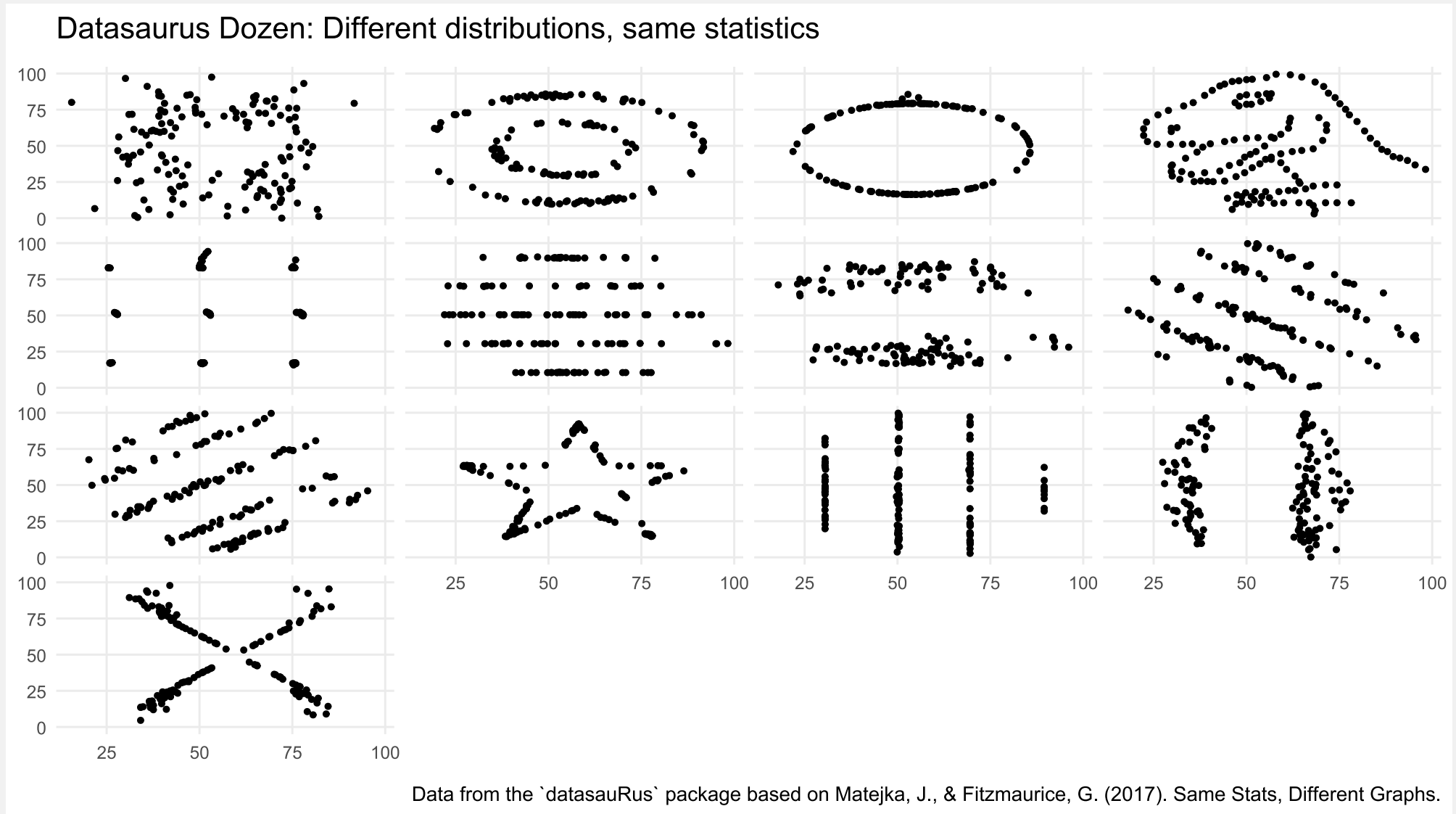
- Poster
- More open design choice
  - Attract people from far
  - You quickly loose people to other posters
  - Medium complexity (depending on the event)
- 

- Talk
- You can use animations to guide through
  - Little time → less complex

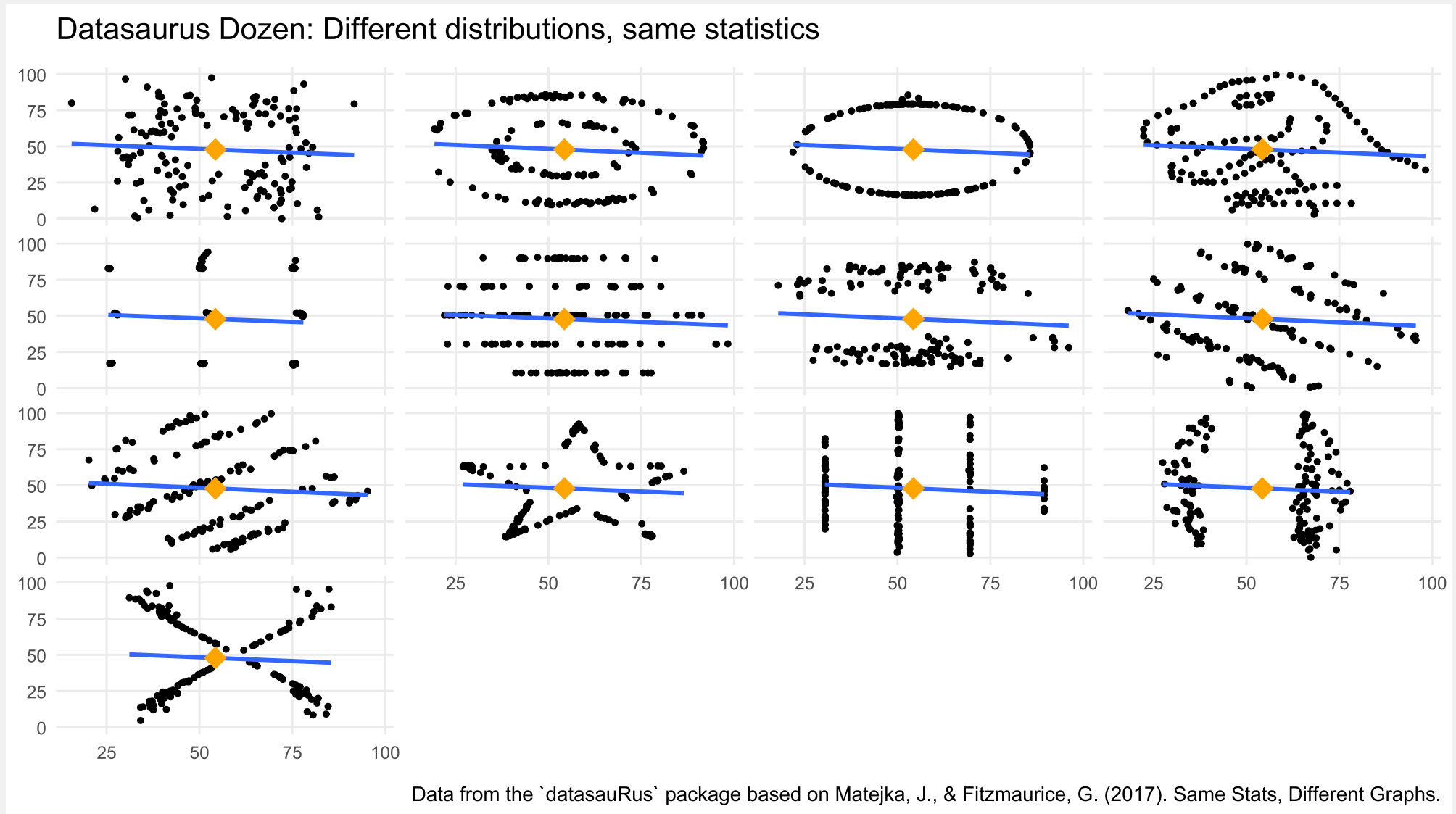


# 2: Make your data transparent

# Don't hide data behind summaries




# Don't hide data behind summaries



# Don't hide data behind summaries

Bar graphs hide a lot of information about the data.



---

PERSPECTIVE

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

**Tracey L. Weissgerber<sup>1\*</sup>, Natasa M. Milic<sup>1,2</sup>, Stacey J. Winham<sup>3</sup>, Vesna D. Garovic<sup>1</sup>**

**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](mailto: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

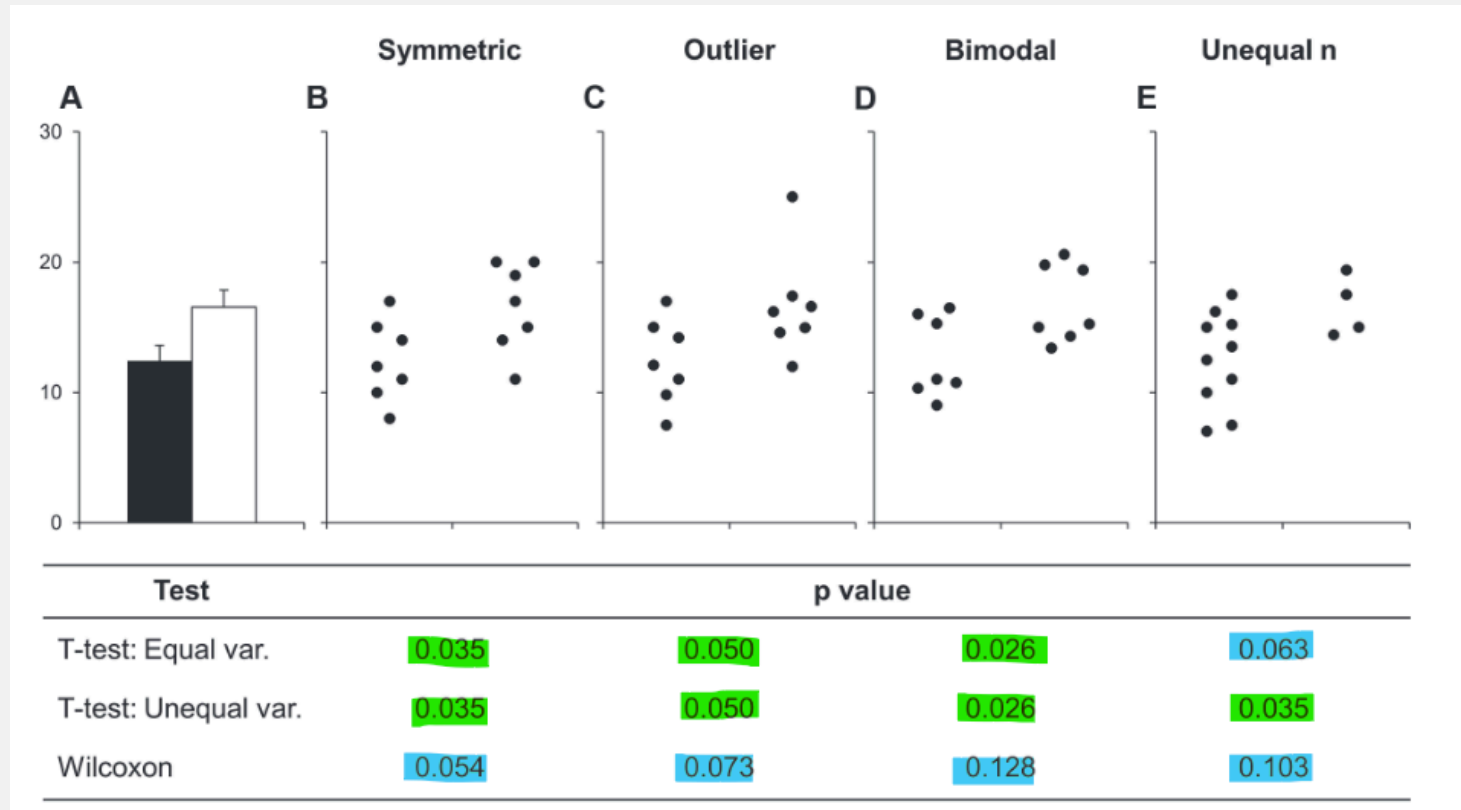
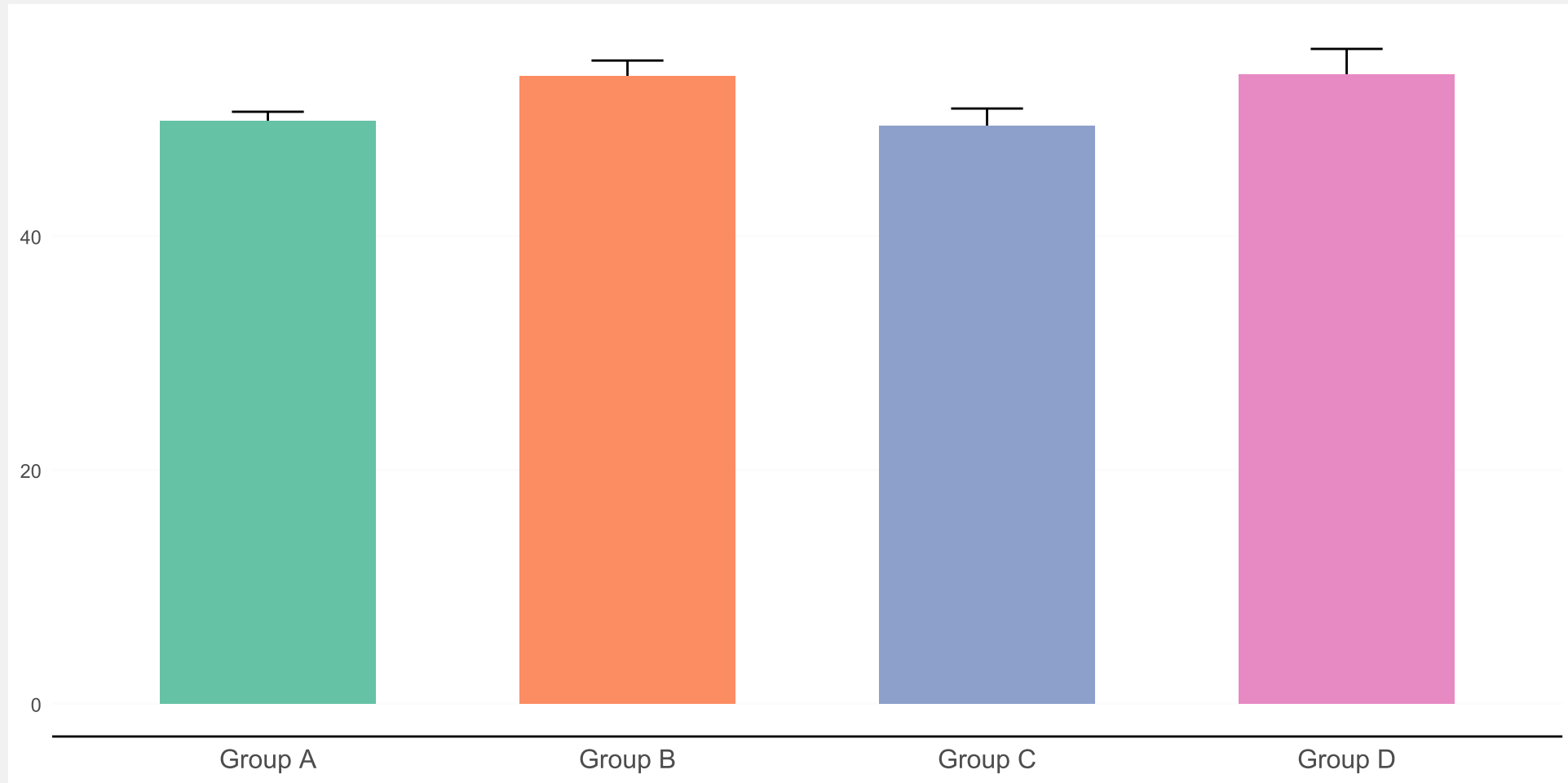


Figure 1 from Weissgerber et al. 2015

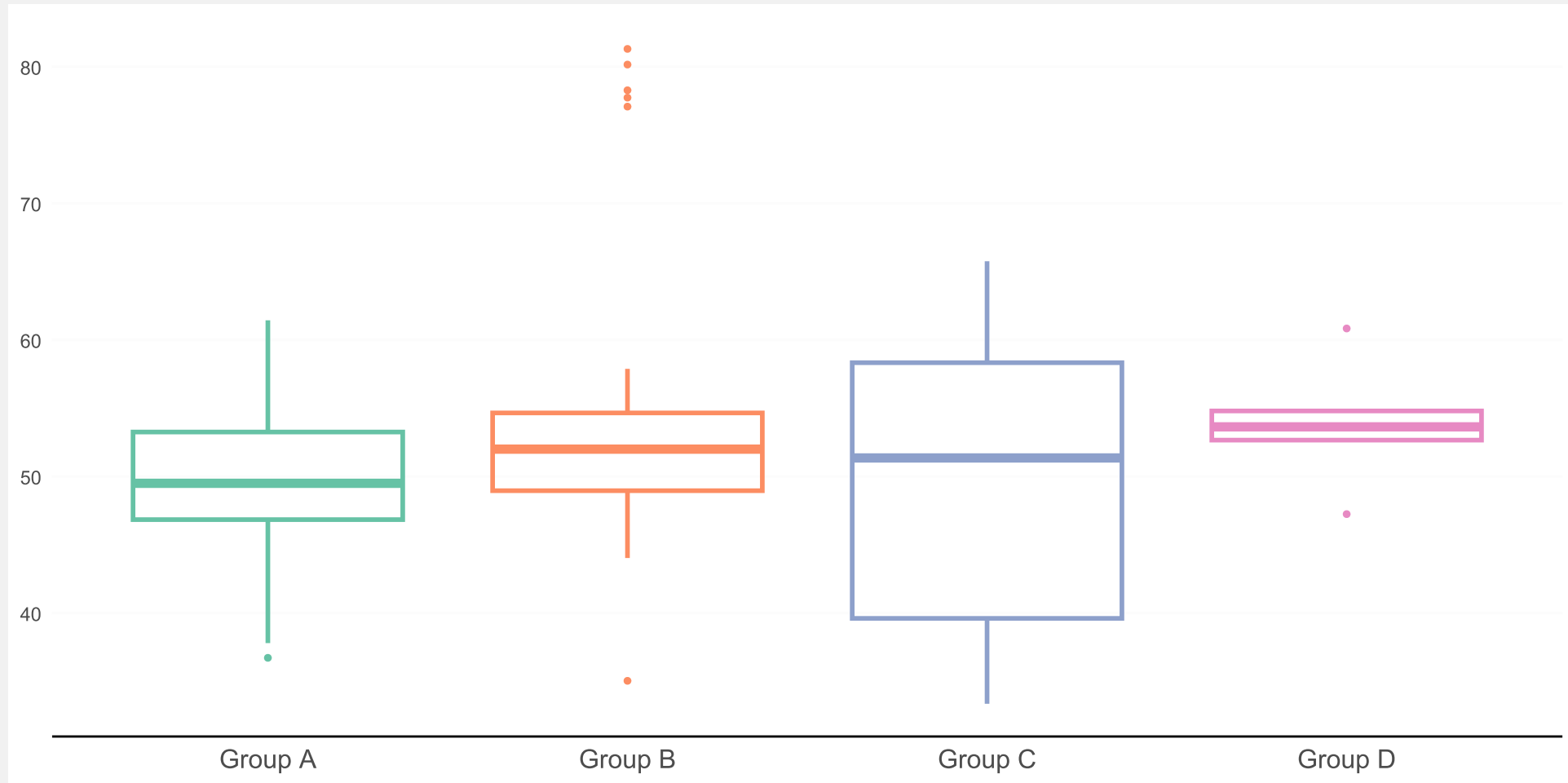
# Alternatives to bar plots

Bar plots only show mean  $\pm$  SE/SD.



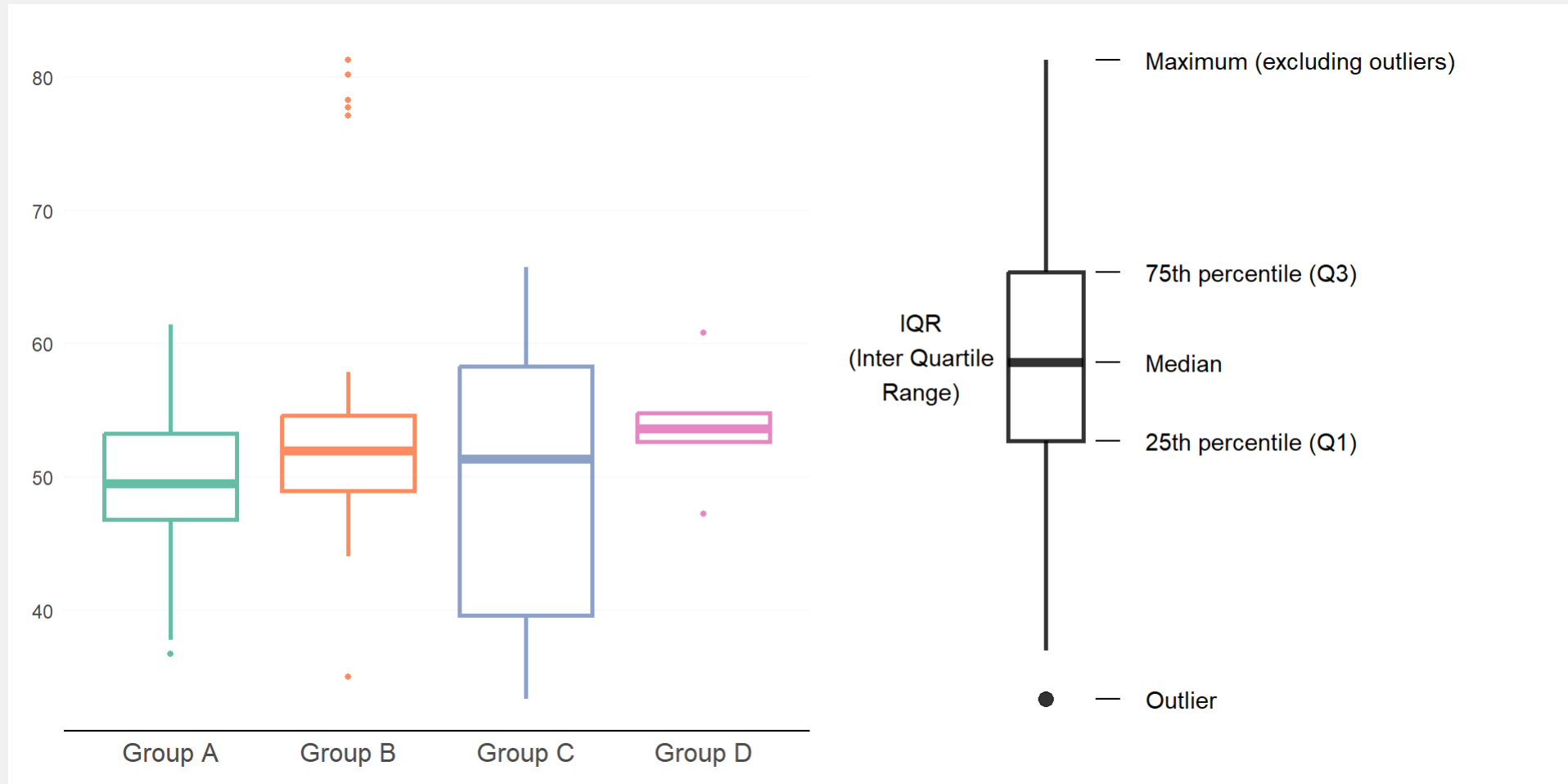
# Alternatives to bar plots

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



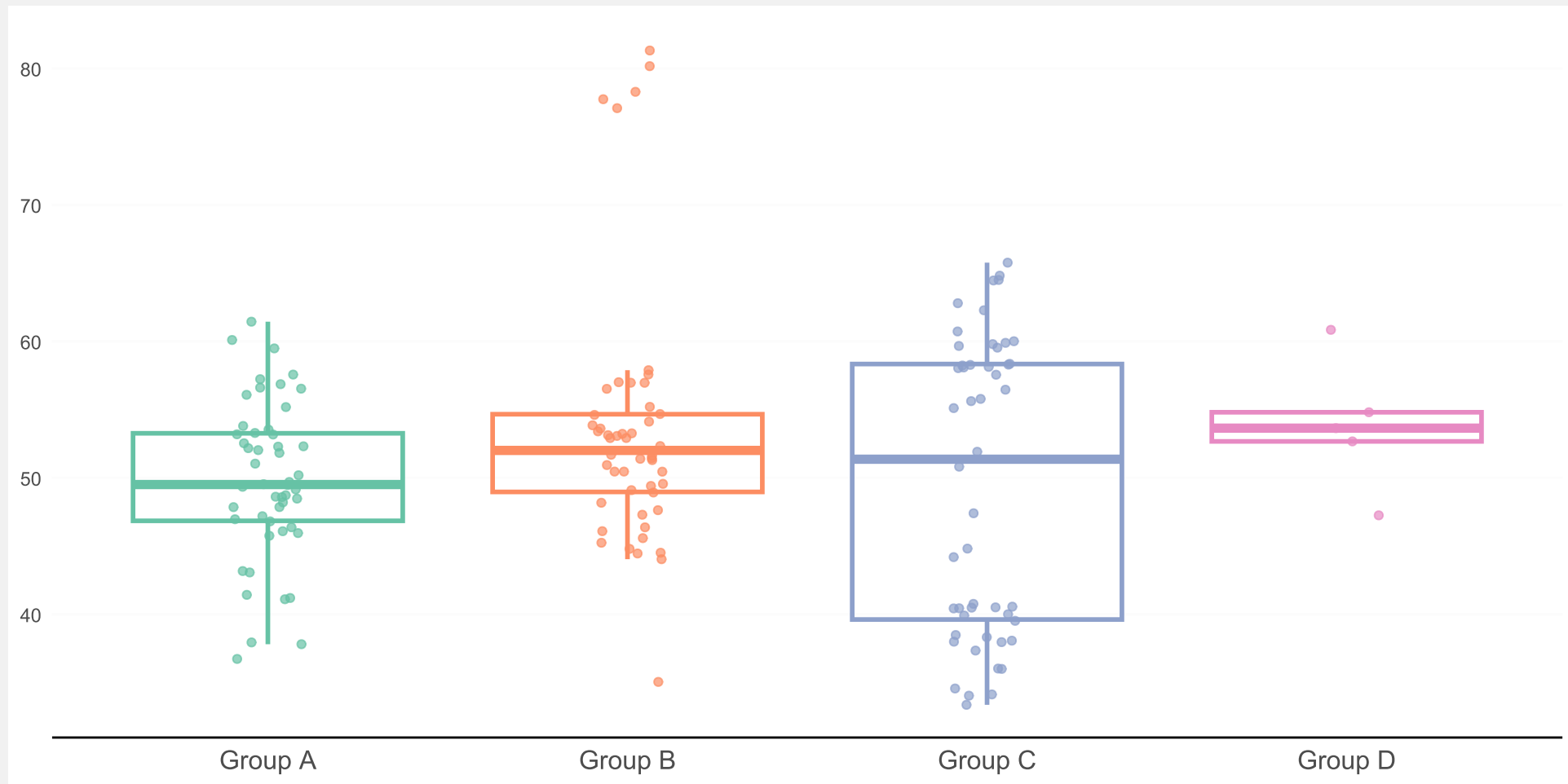
# Alternatives to bar plots

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



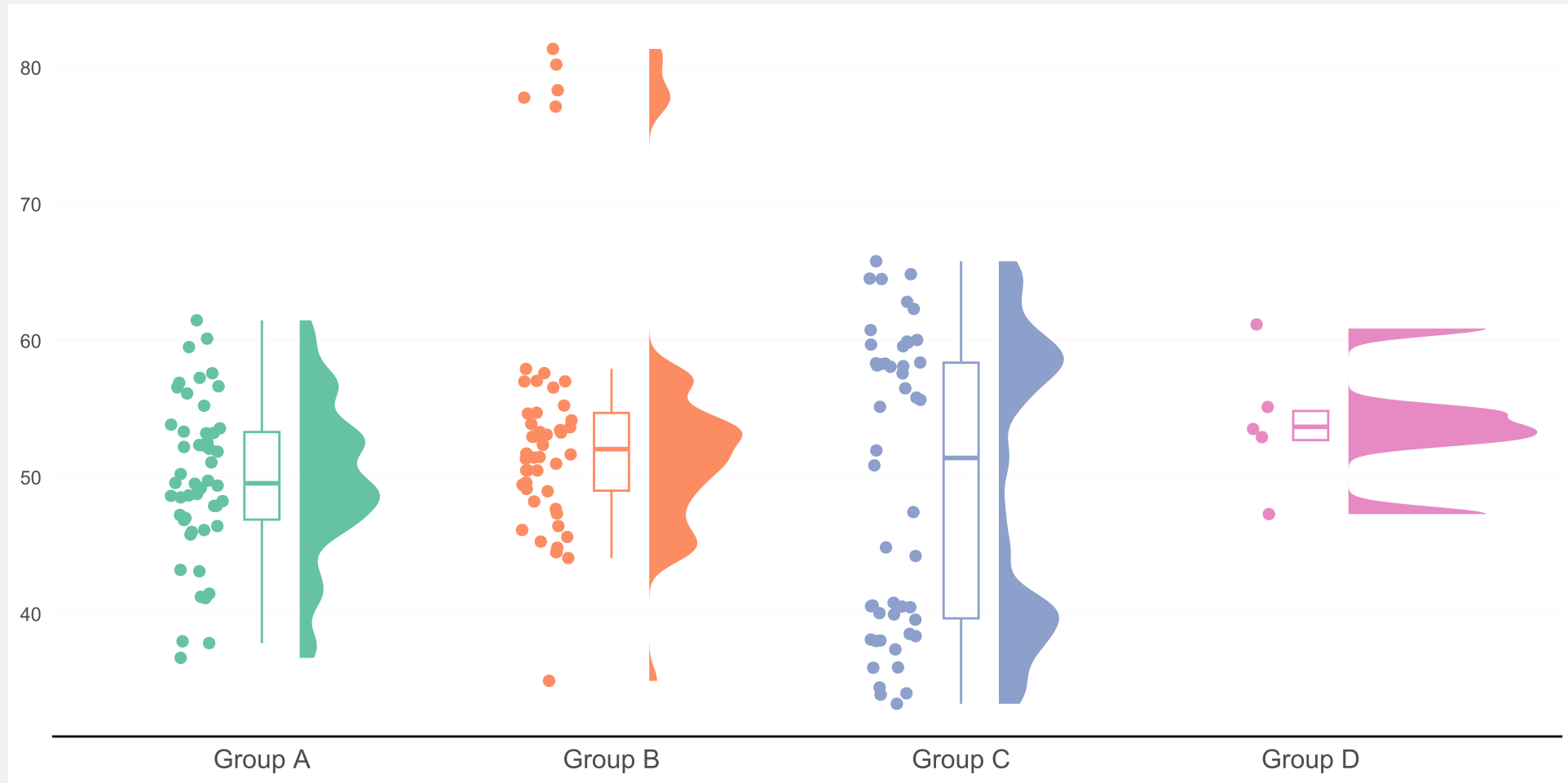
# Alternatives to bar plots

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



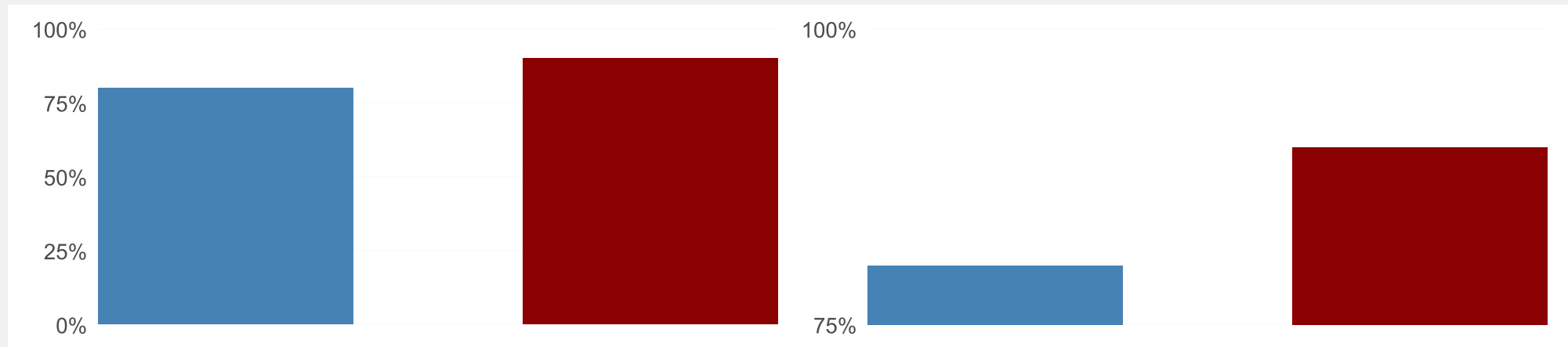
# Alternatives to bar plots

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

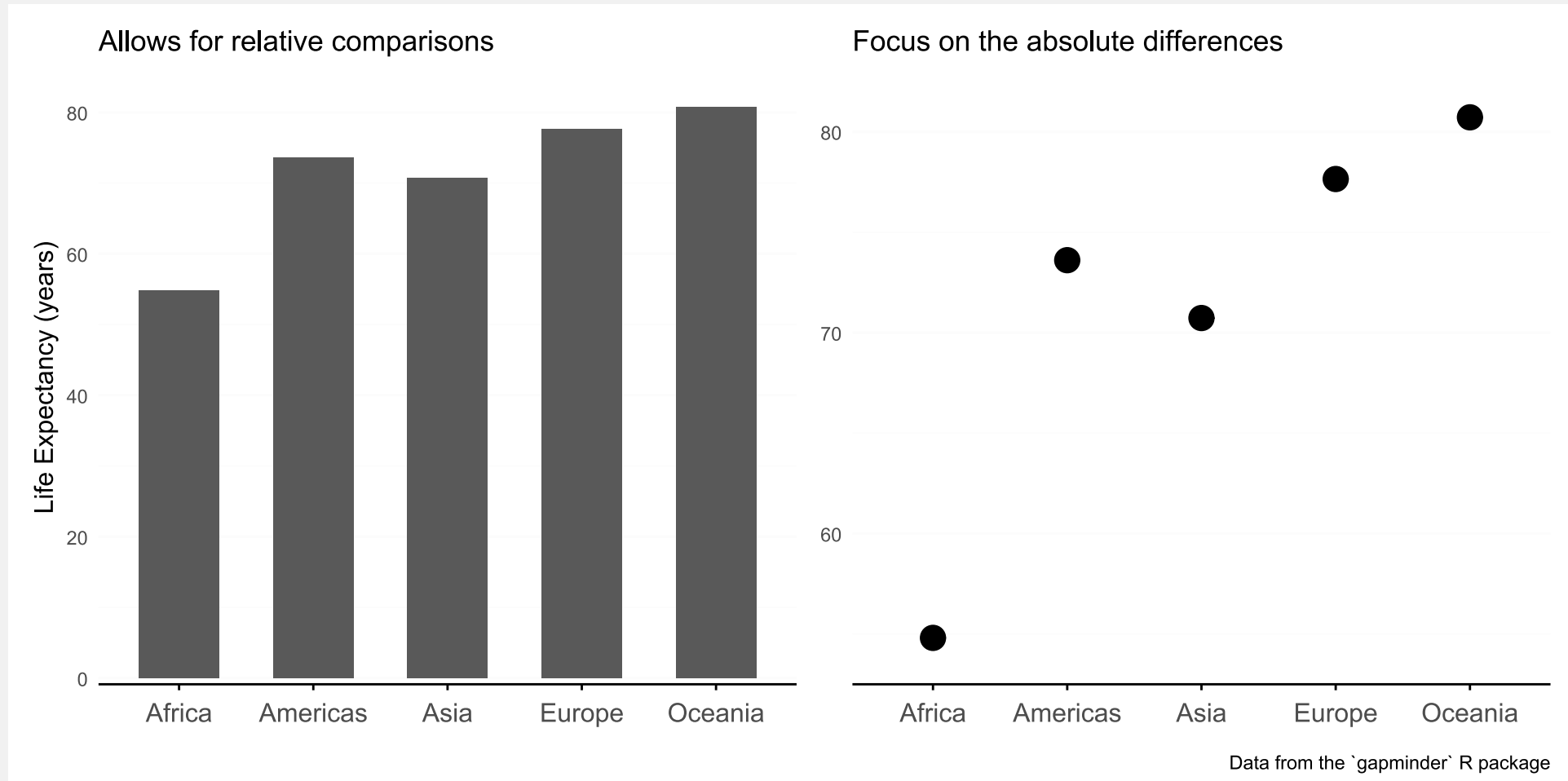


Here: bar length does not represent **relative data proportions** anymore

Always start bars at 0!

# Principle of proportional ink

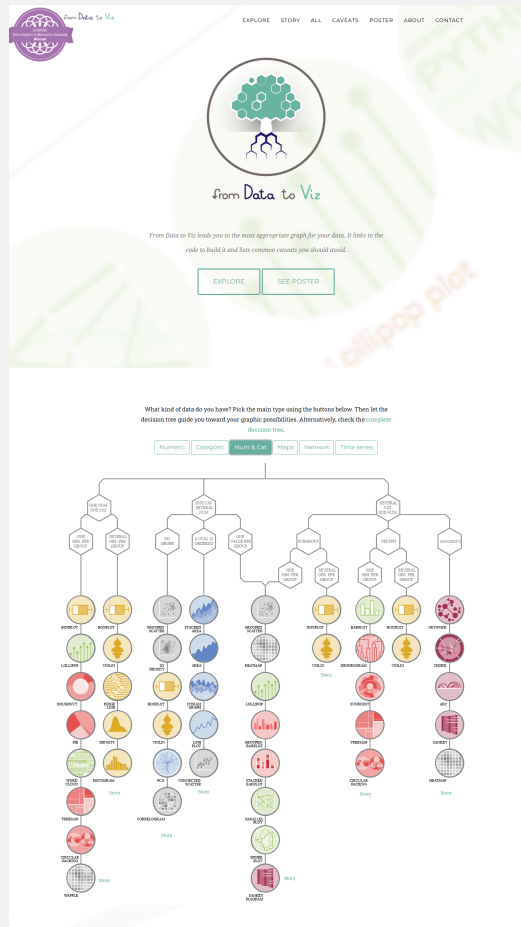
Other plot types don't have to start at 0.



# 3: Choose the right chart type

# Choose the right chart type

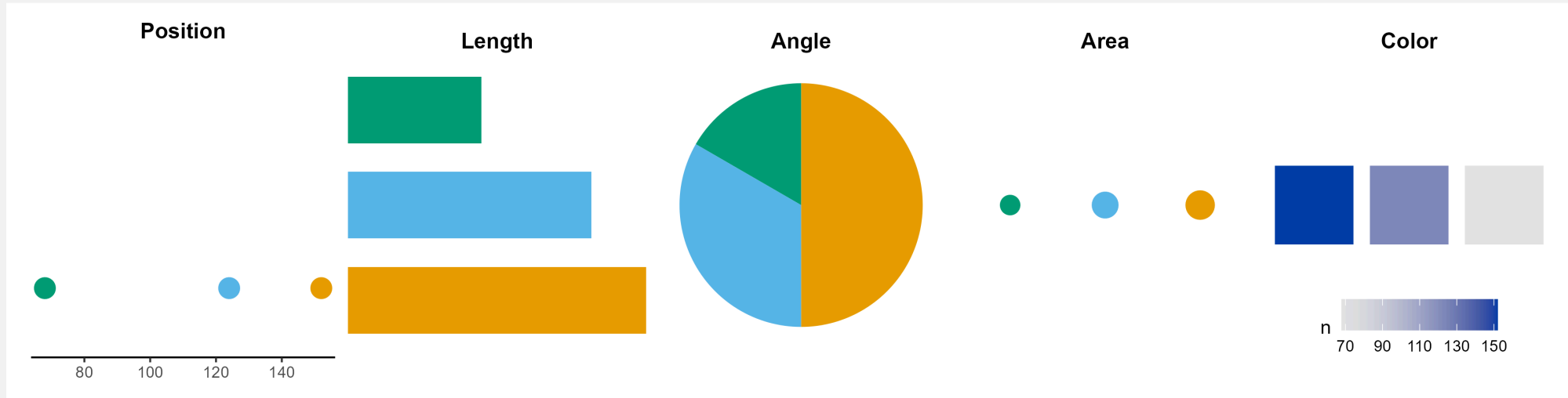
There are so many chart types - and cool tools to explore them



The dataviz project

From data vo viz

# 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

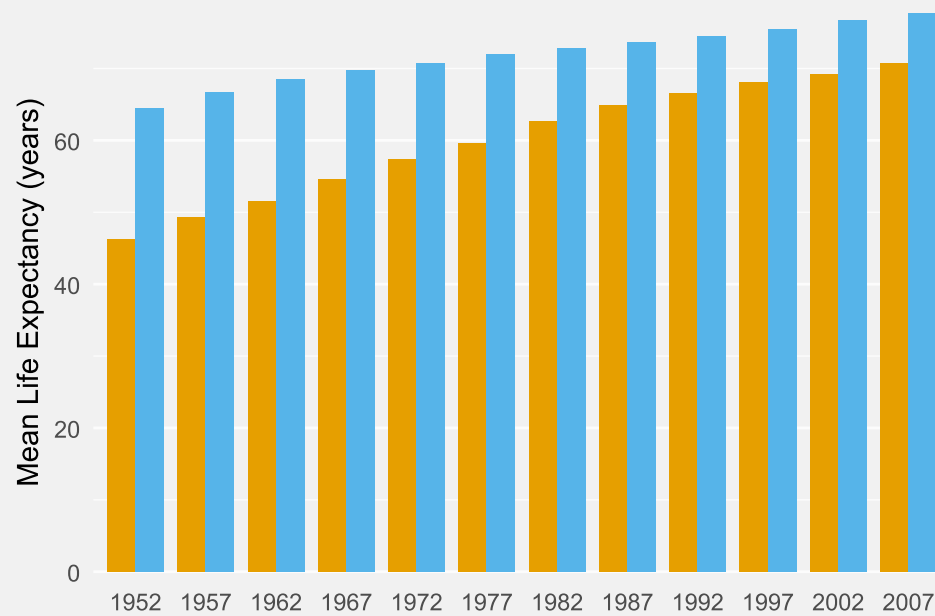
- The readers attention is limited: **be concise**
- Think about the **main message** you want to convey
  - Which variables do you need?
  - What can you omit?

# Arrange your plot

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

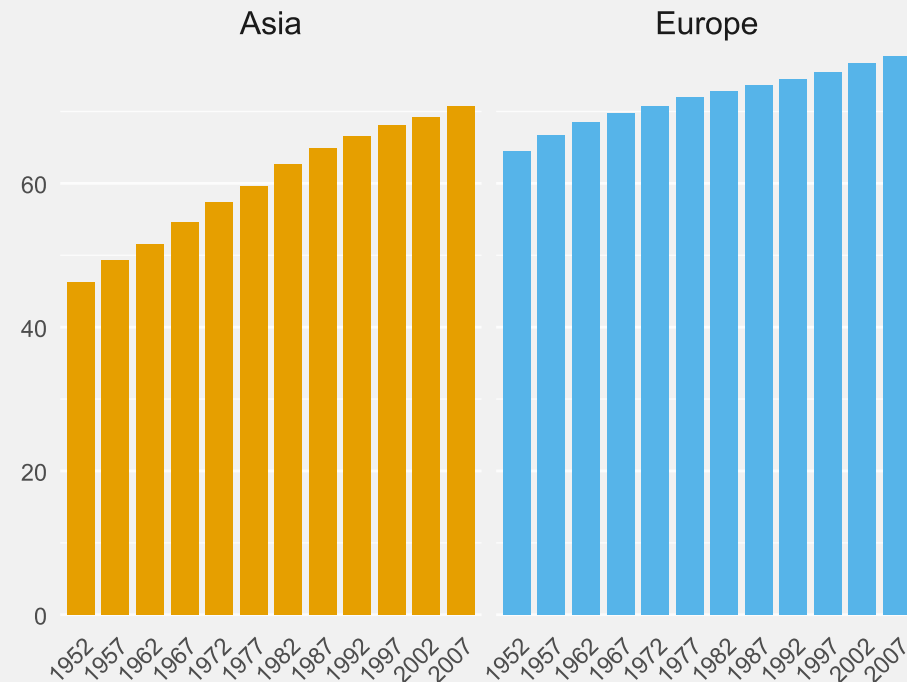
Life Expectancy group comparisons

Main message: Europe higher than Asia, Asia catching up



Life Expectancy Trends

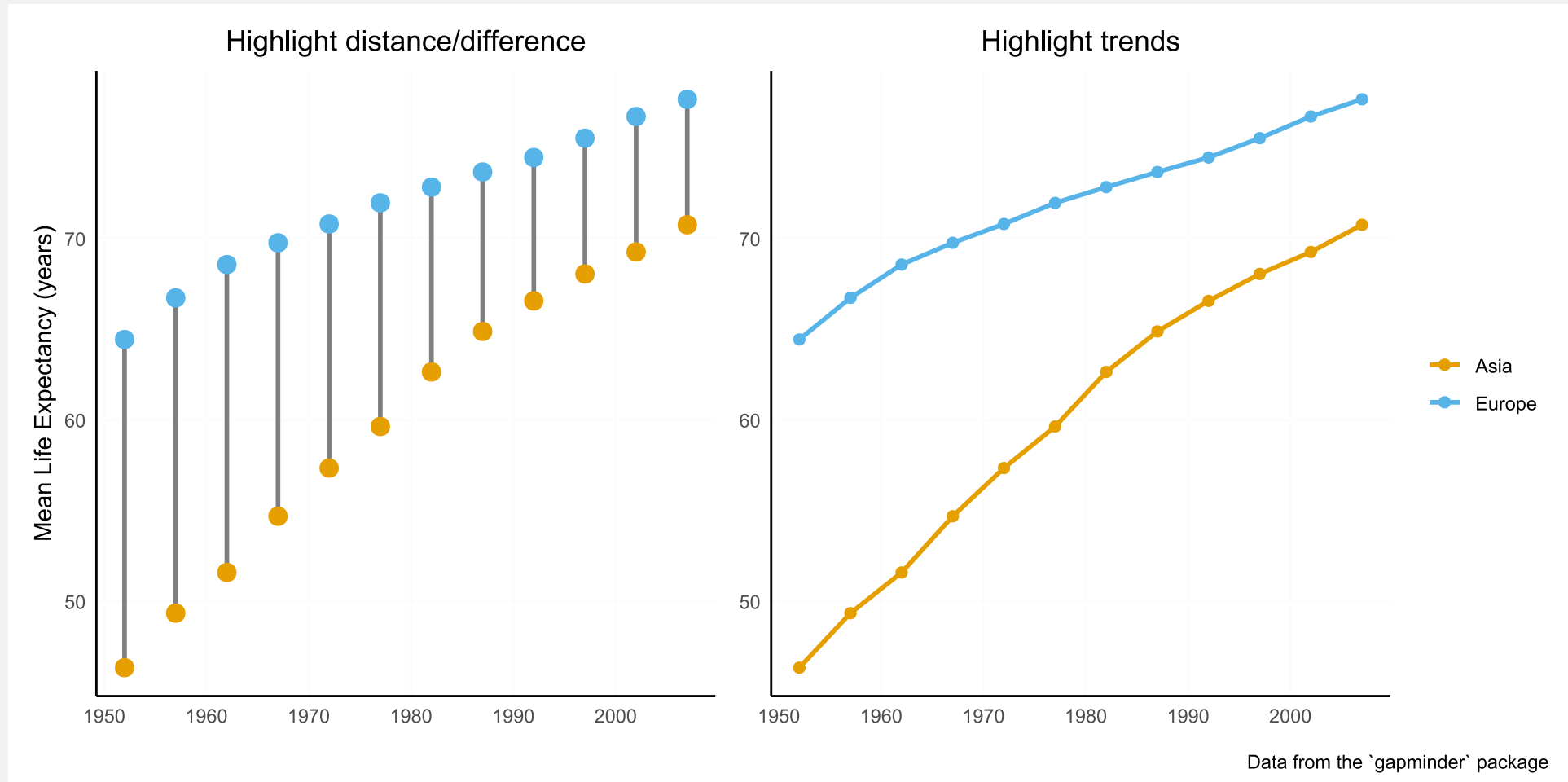
Main message: Different rates of improvement over time



Asia Europe

# Choose a good plot type

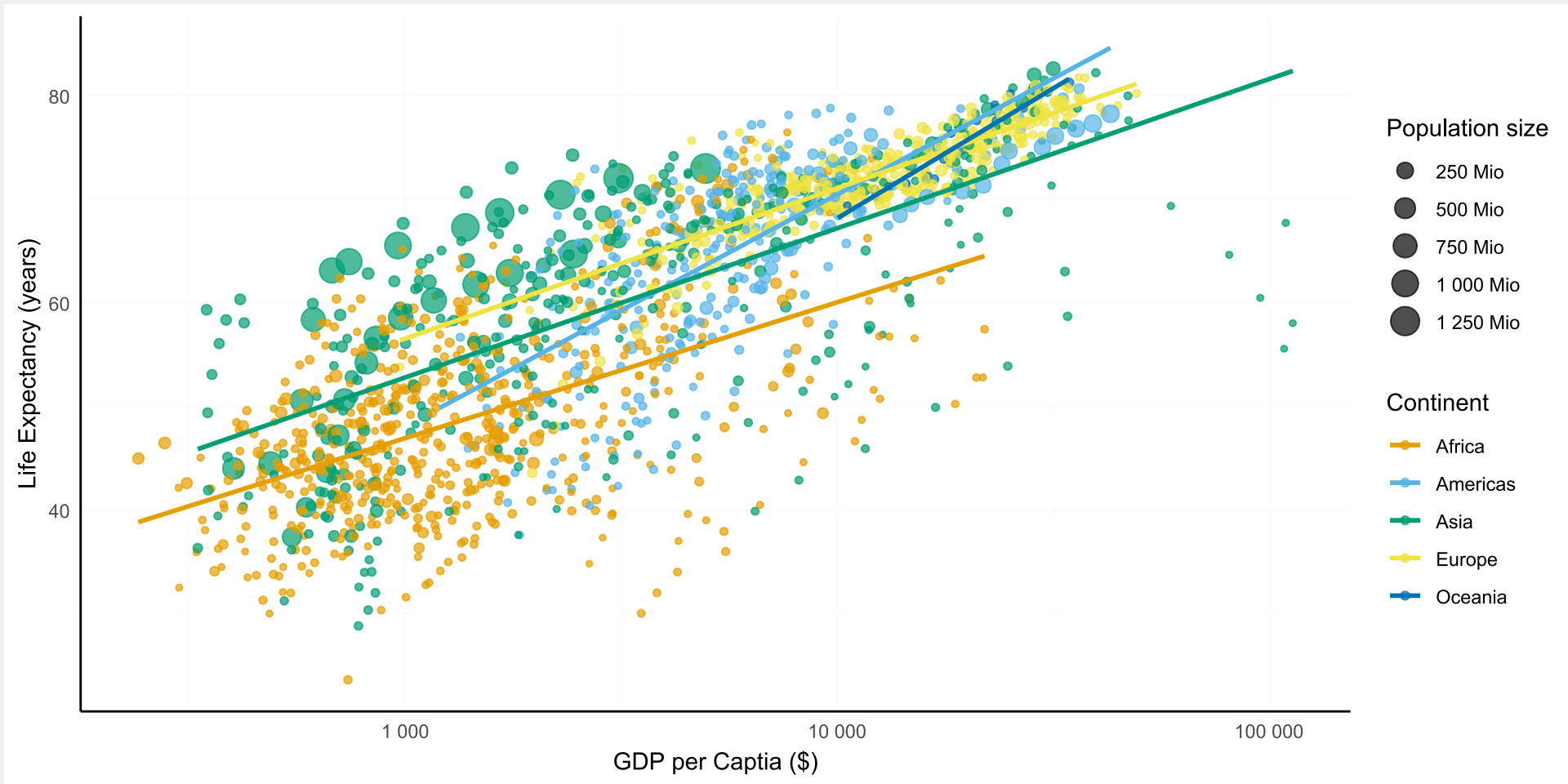
Different plot types tell different stories



# Keep it simple

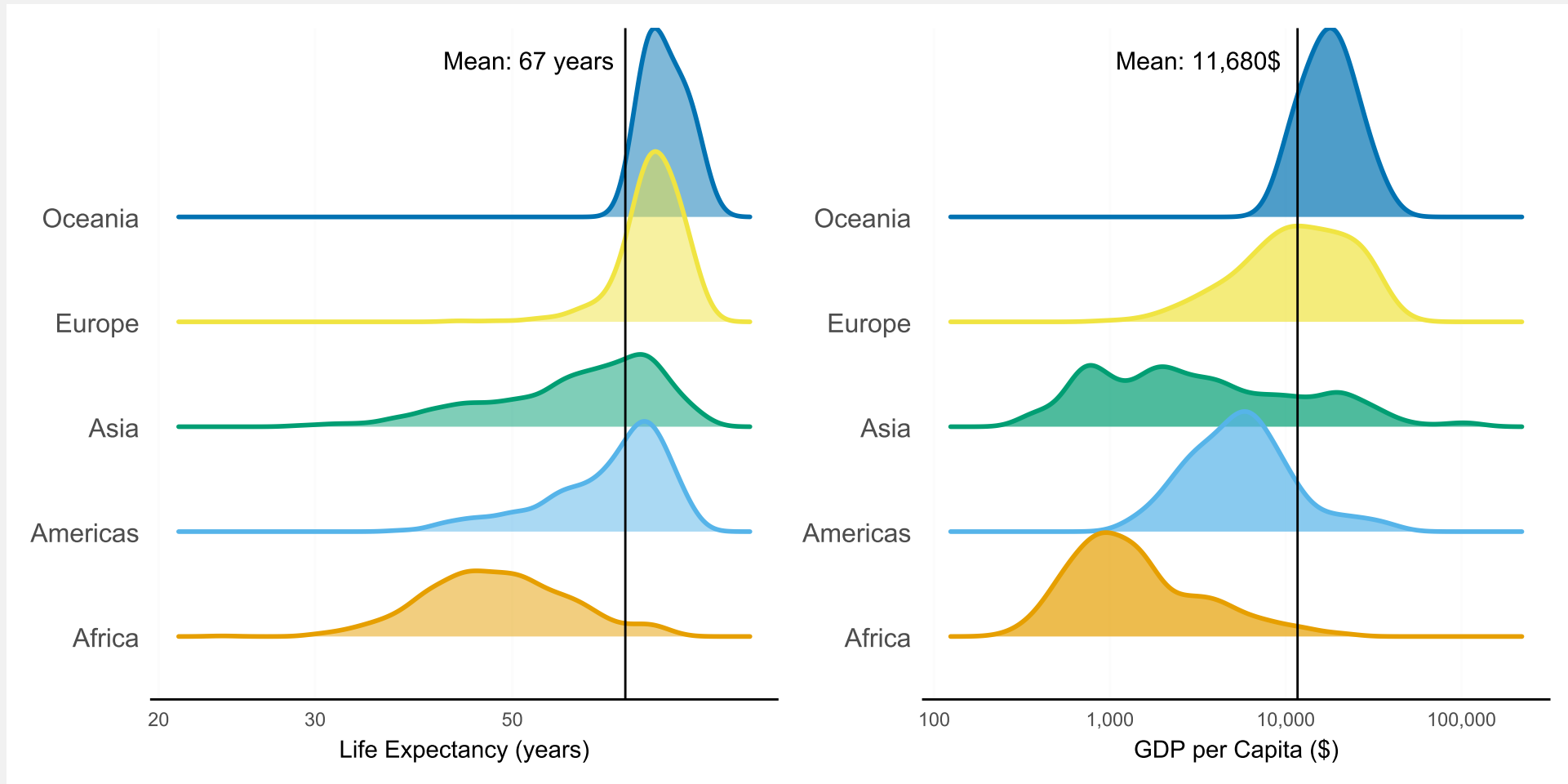
Don't overcomplicate your figures and bury your message

What is the main message here?



# Keep it simple

Message: Life expectancy and GDP differences in the world





# 5. Consider the trip

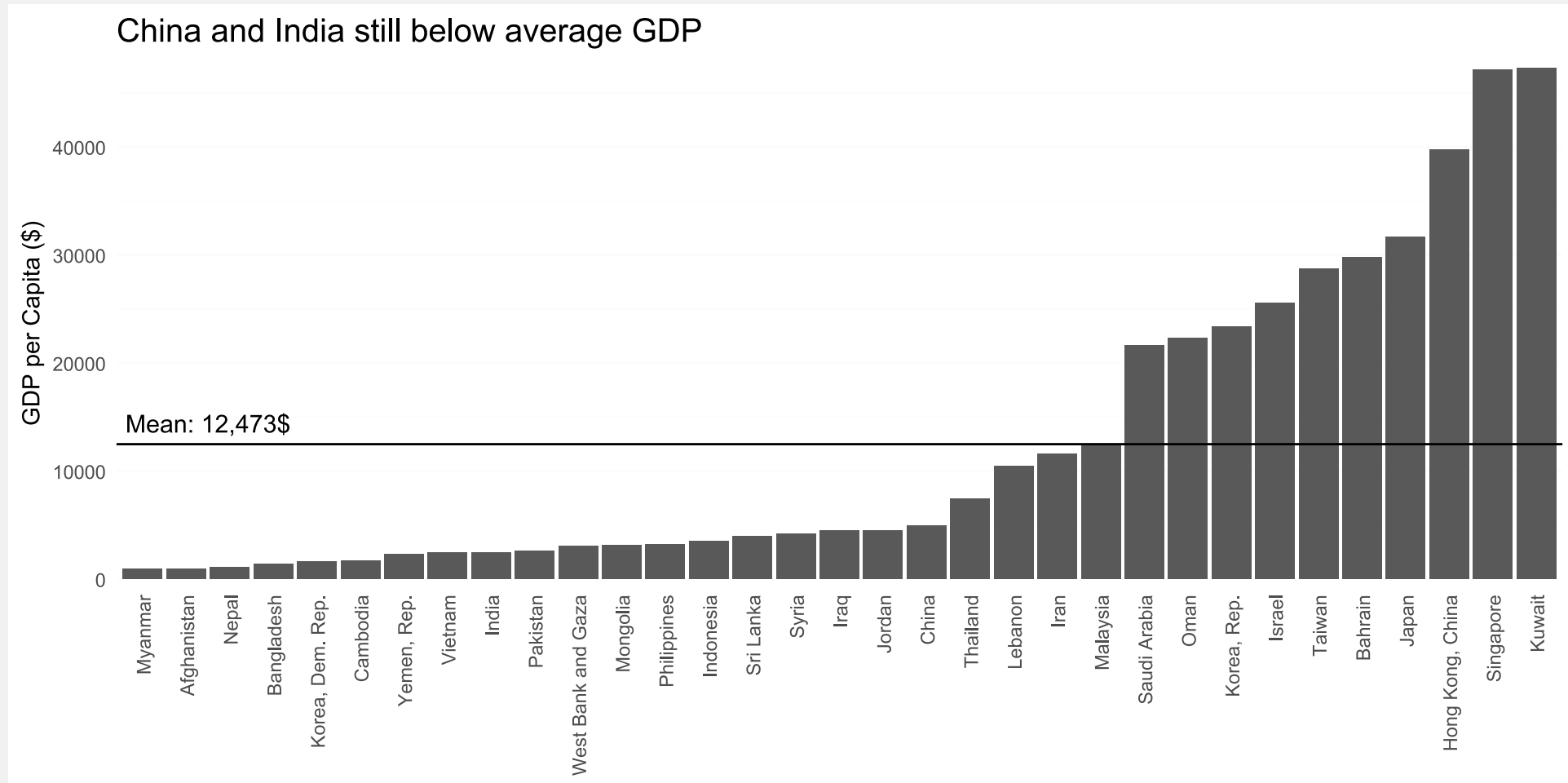
# Reading a figure is a timely experience

- 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

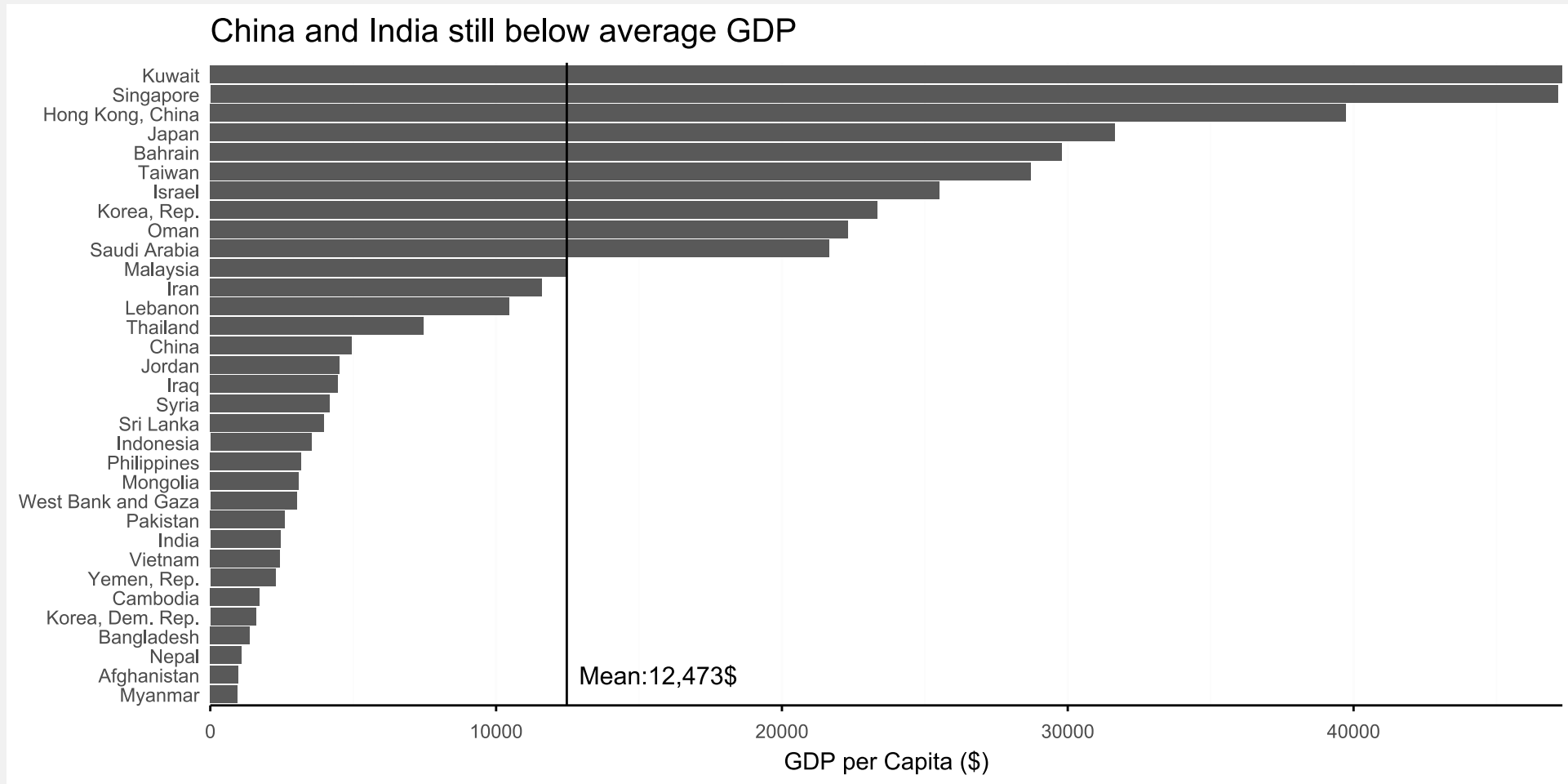
# Consider the trip

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

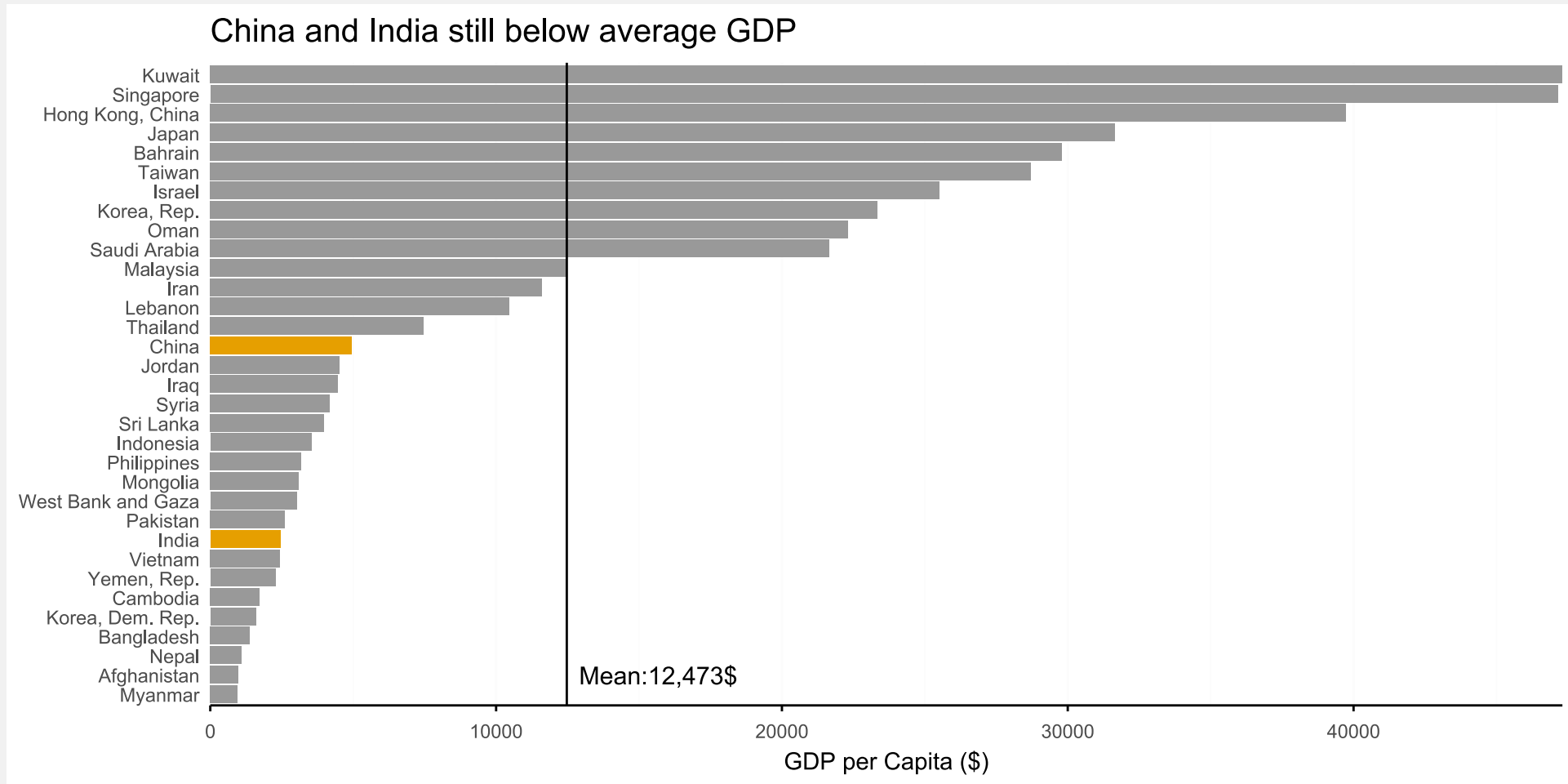
Use **highlighting** and de-emphasizing

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, ...

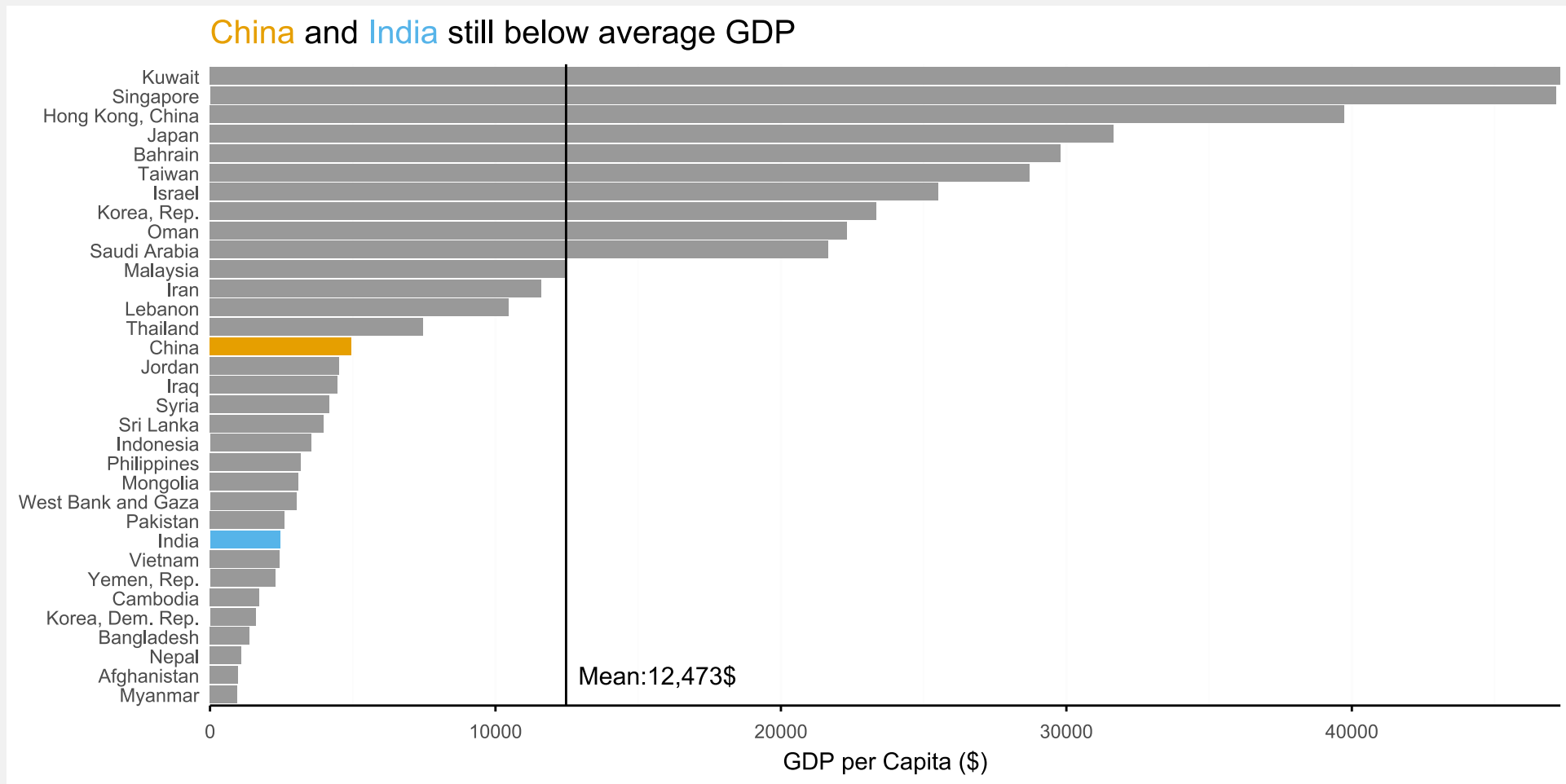
# Highlight the main message

Highlight focus countries, de-emphasize all others



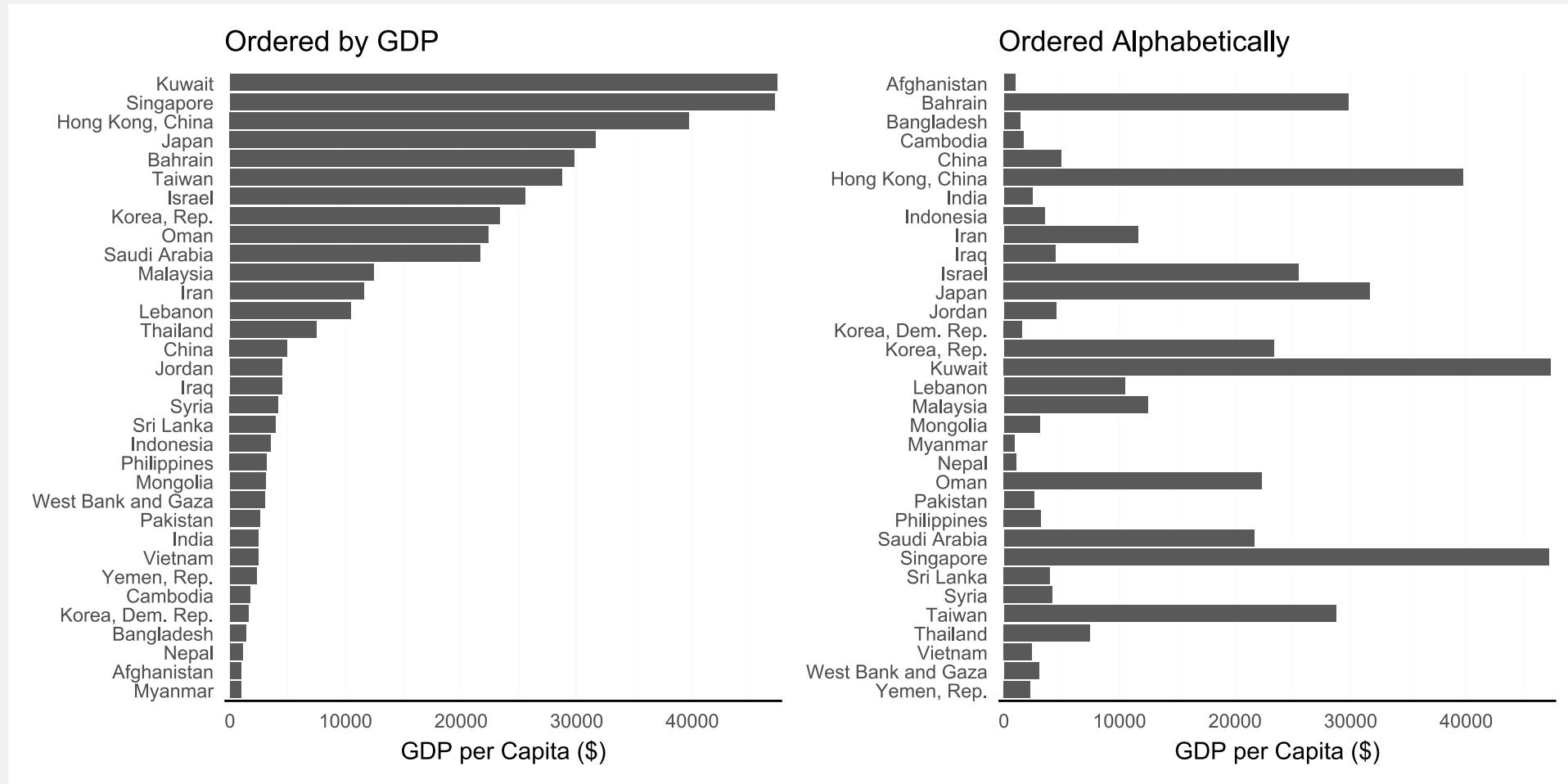
# Highlight the main message

Highlight focus countries, de-emphasize all others



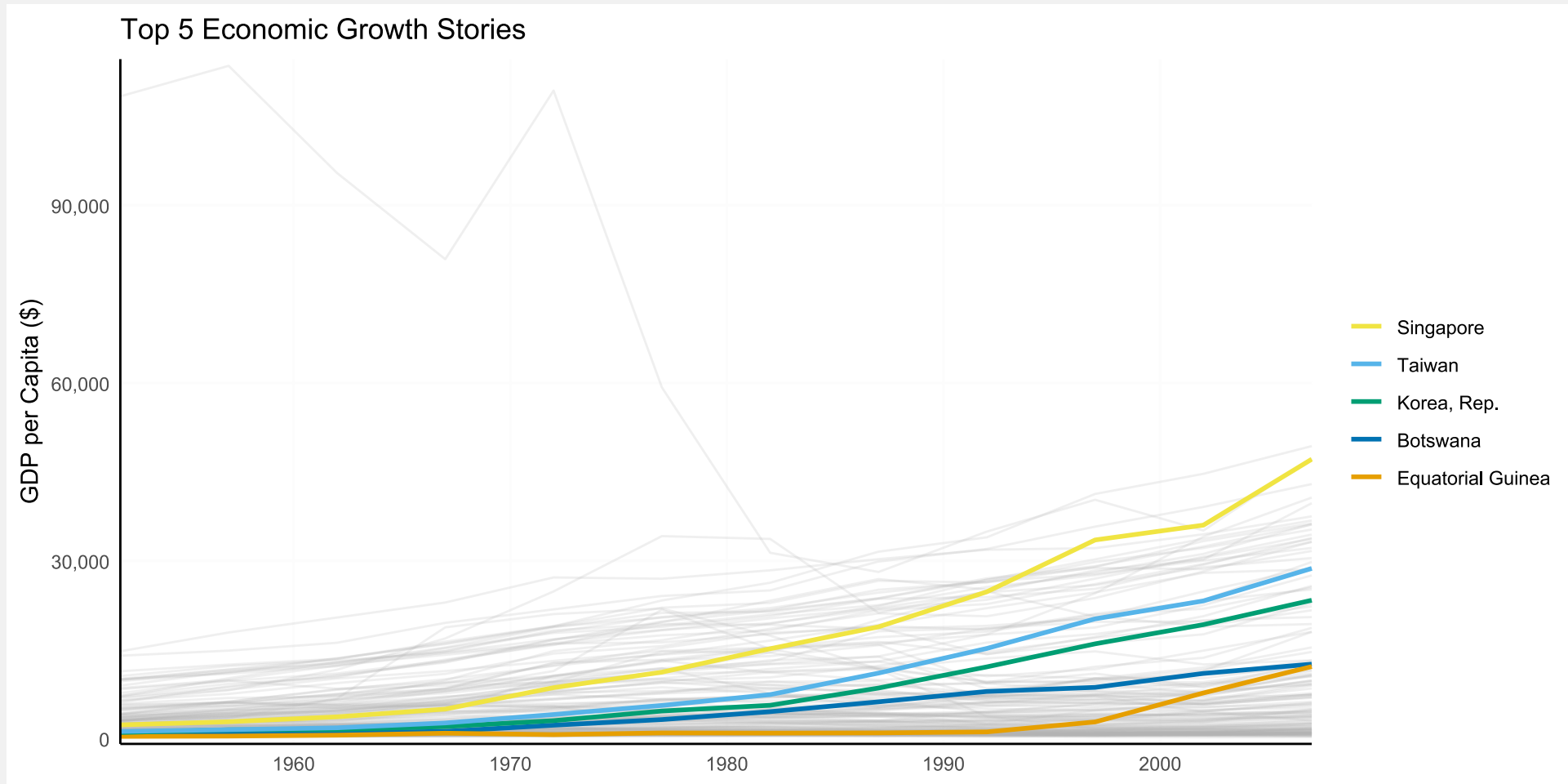
# Order your data

Order categories consciously not automatically.



# Order your data

Order categories consciously not automatically.





# 6. Less is more

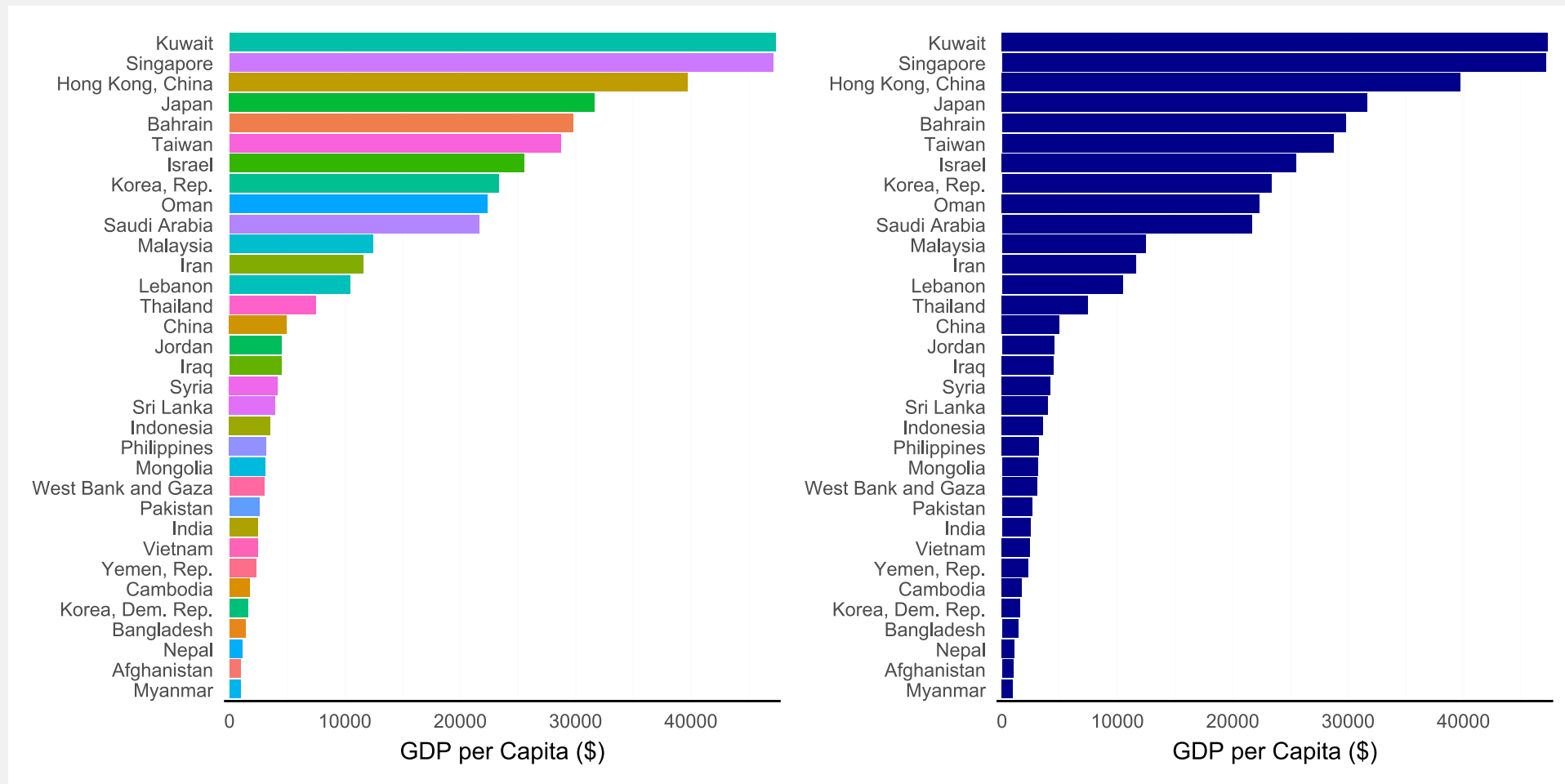
# The importance of differences

*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

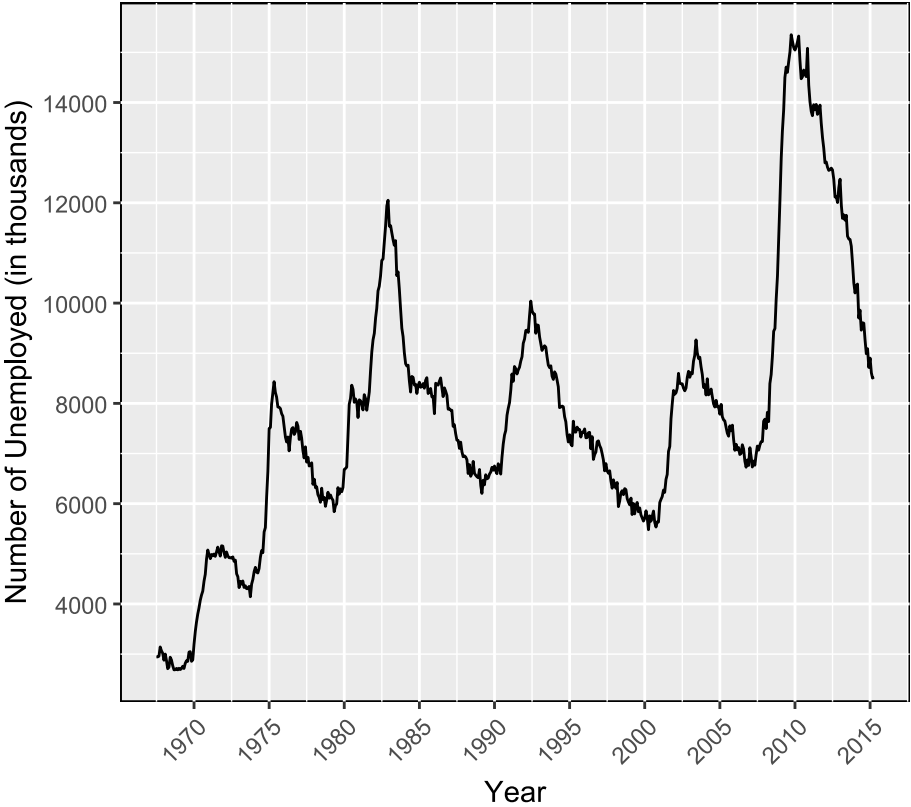


# Declutter your figure

- 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

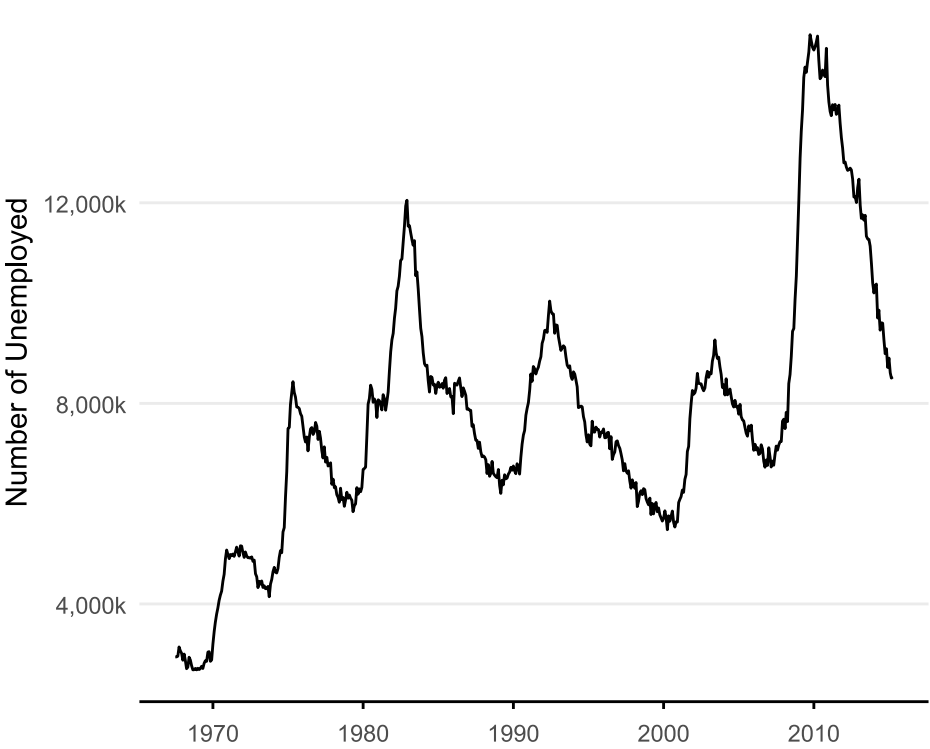
# Declutter your graphs

Time Series of Unemployment in the US



Data Source: Federal Reserve Economic Data (FRED)

Time Series of Unemployment in the US



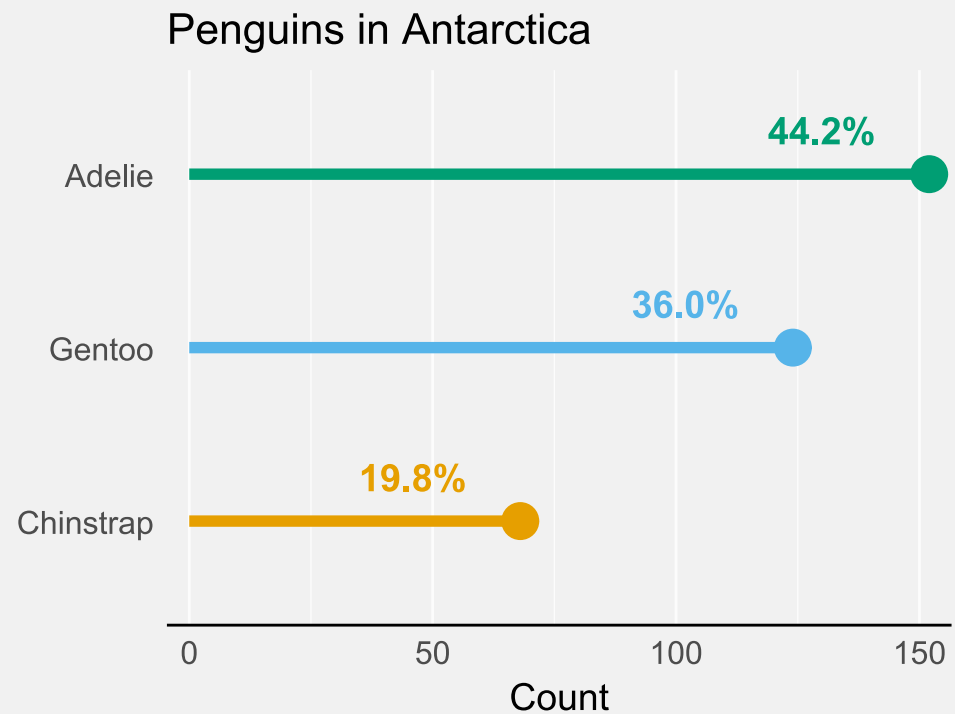
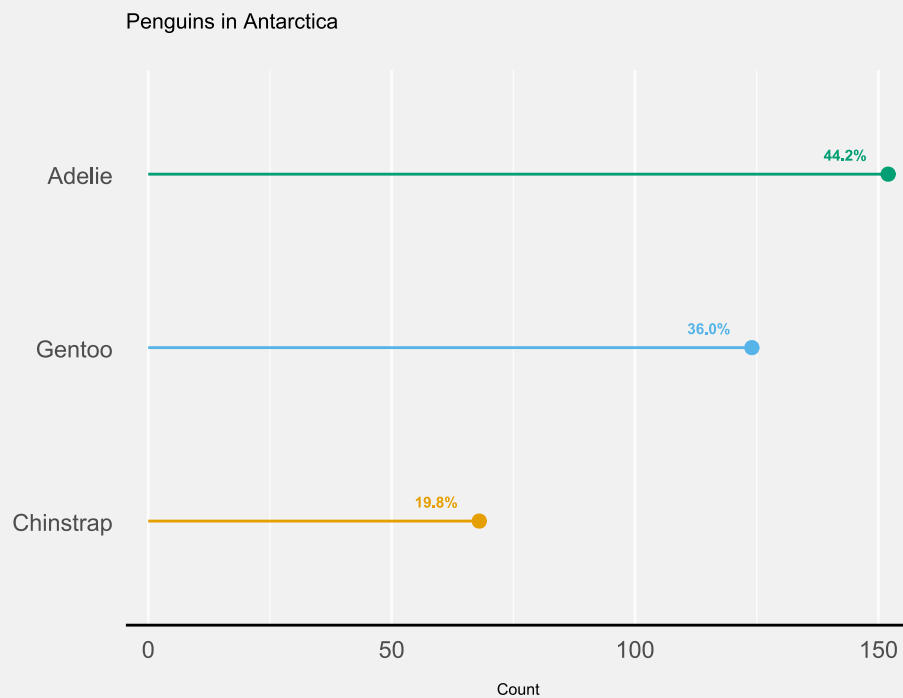
Data Source: Federal Reserve Economic Data (FRED)



# 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

1:1				
1.1:1	You dislike readers.	That's bad.	Nope	Nope
1.5:1	Not ideal.	Not ideal.	That's bad.	That's bad.
3:1	Can be ok.	Can be ok.	Not ideal.	Not ideal.
4.5:1	Safe for large text.	Safe for large text.	Ok.	Ok.
7:1	Safe.	Safe.		

Blogpost on colors by Lisa Charlotte Muth (Datawrapper)

- Use tools to check contrast, e.g.  
[https://snook.ca/technical/colour\\_contrast](https://snook.ca/technical/colour_contrast)

# Color

Use logical/intuitive colors



[Blogpost on colors](#) by Lisa Charlotte Muth (Datawrapper)

# Color

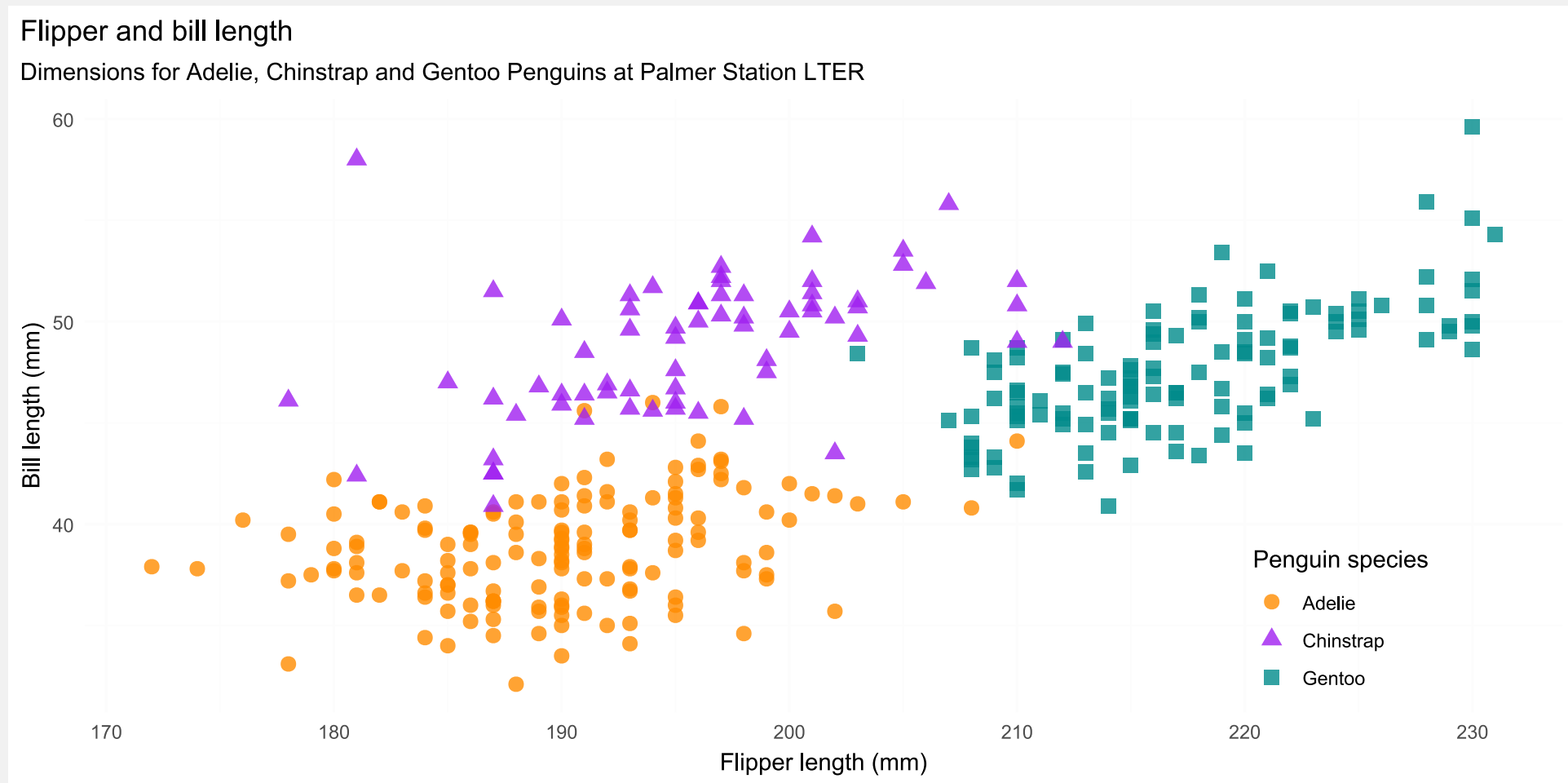
Choose colorblind friendly palettes (if in doubt: test!).

The image shows two web-based tools for color selection and testing. The left tool, 'VIZ PALETTE' by Elijah Meeks & Susie Lu, has sections for 'PICK' (with 'Add' and 'Replace' buttons), 'EDIT' (with a list of 7 colors and their hex codes), and 'GET' (with a text area for color codes). The right tool, 'COLORS IN ACTION', displays 'Color Population' statistics: Color Deficiency - 96%, Deuteranomaly - 2.7%, Protanomaly - 0.66%, Protanopia - 0.59%, and Deuteranopia - 0.56%. It also features a 'Greyscale' option and a 'Randomize Data' button. The main visualization area shows various data representations: a 'Sample font' with colored bars, a 'Randomize Data' box, a 'Stroke' selector (Dark/None), a grid of colored boxes, a scatter plot of colored dots, and a network graph of colored nodes and edges. Text labels like 'word', 'mot', 'لفظ', '字', '인사', 'salita', and '워드' are visible near the scatter plot.

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

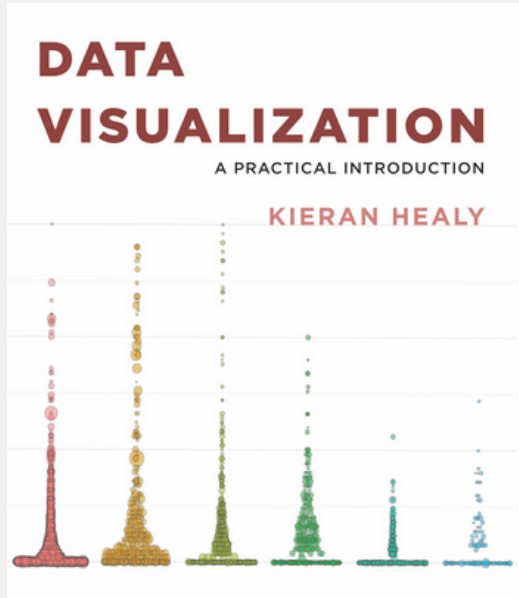
 [Subscribe to the mailing list](#)

 For topic suggestions and/or feedback [send me an email](#)

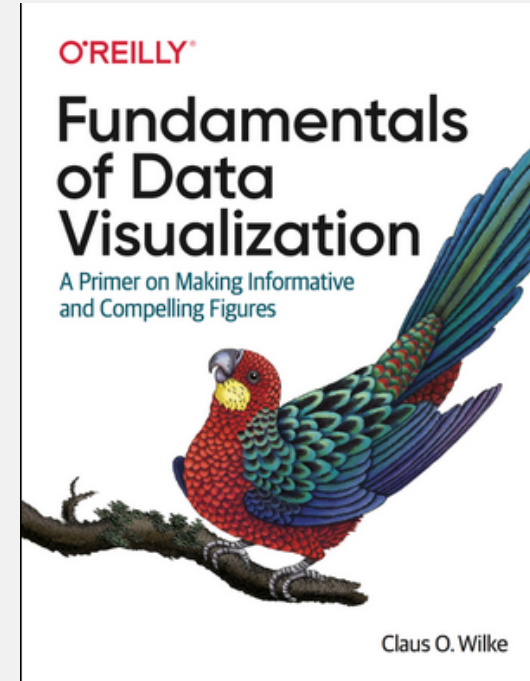
# Thank you for your attention :)

Questions?

# References



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



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