Data Visualization for Economics Research

Lyon Summer School in Empirical Research Methods - ENS de Lyon

Vincent Bagilet

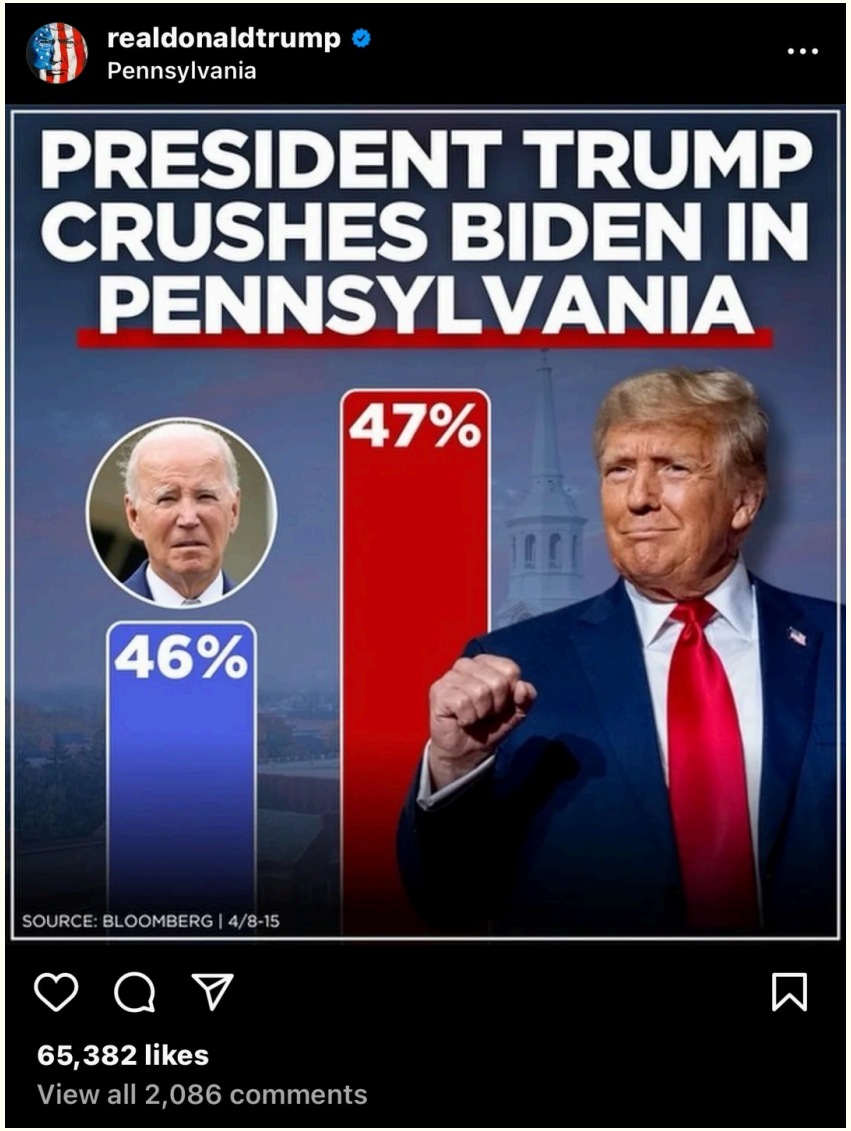
2025-07-01

Introduction

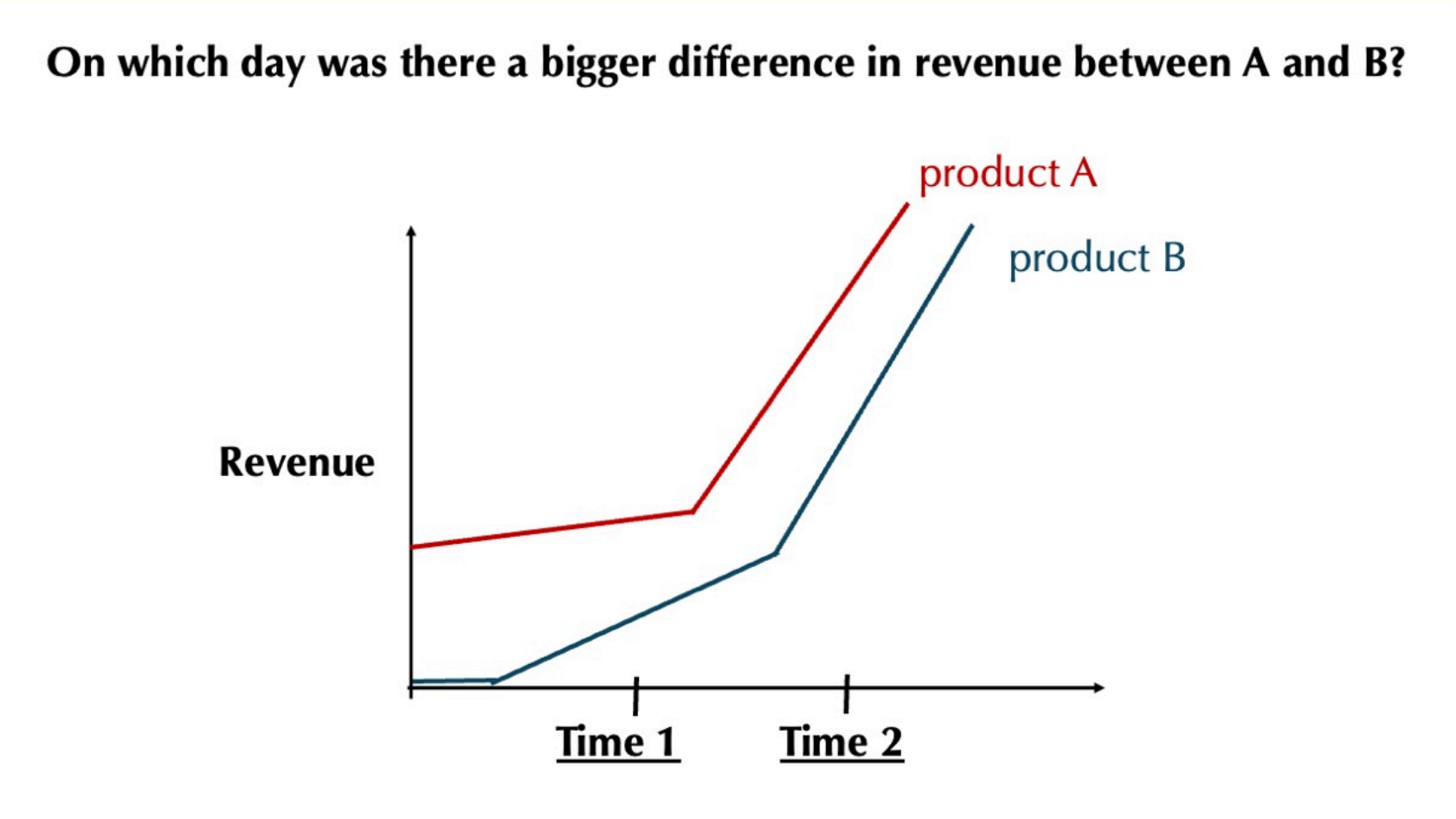
Everywhere but not so simple

- Data viz is **everywhere**
- We work with data, we routinely (need to) visualize it
- **Seems pretty simple**, we all know how to make graphs
- Sure BUT **there are a few things we need to think about** when visualizing data
- Once you pay attention to data viz, it is fun, instructive and satisfying!

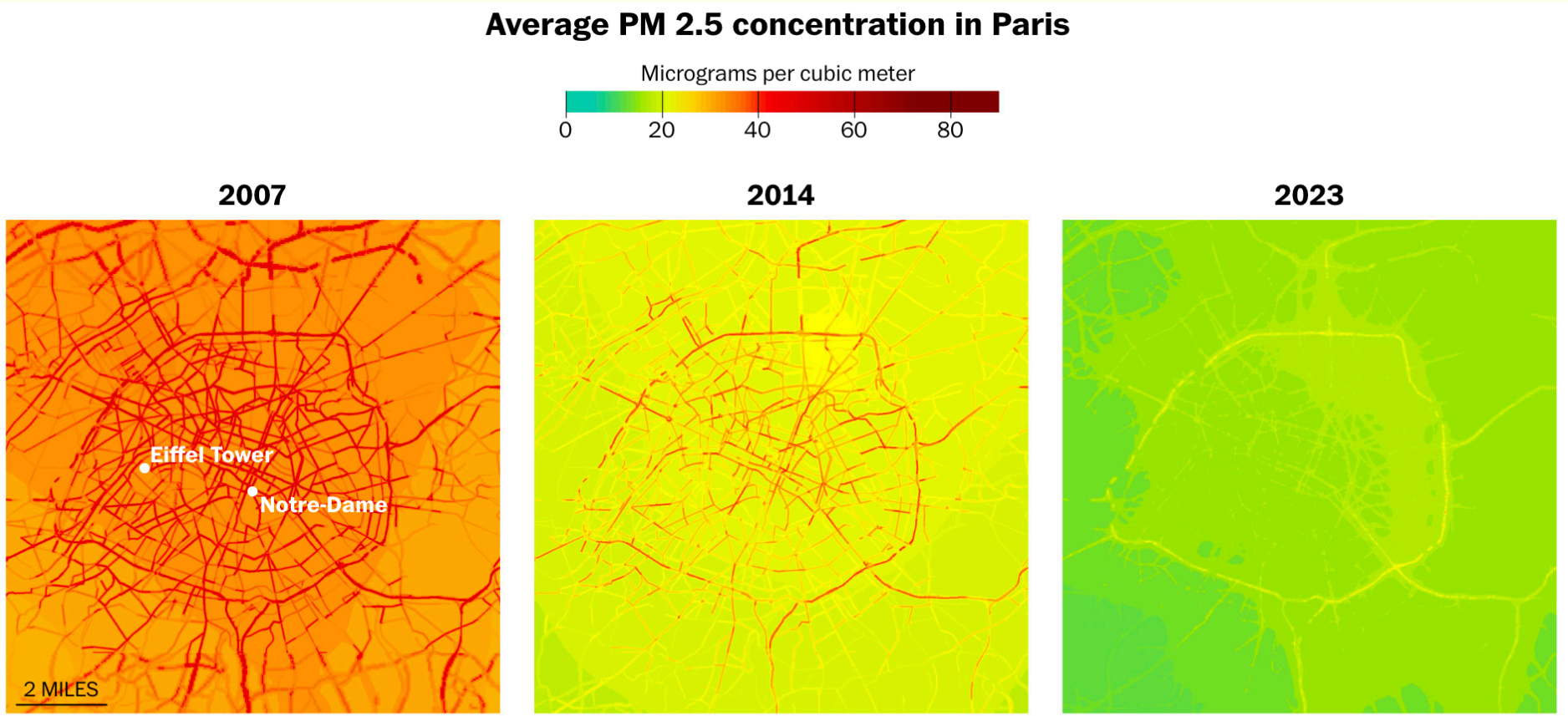
Data viz can be obviously deceptive



...or difficult to interpret

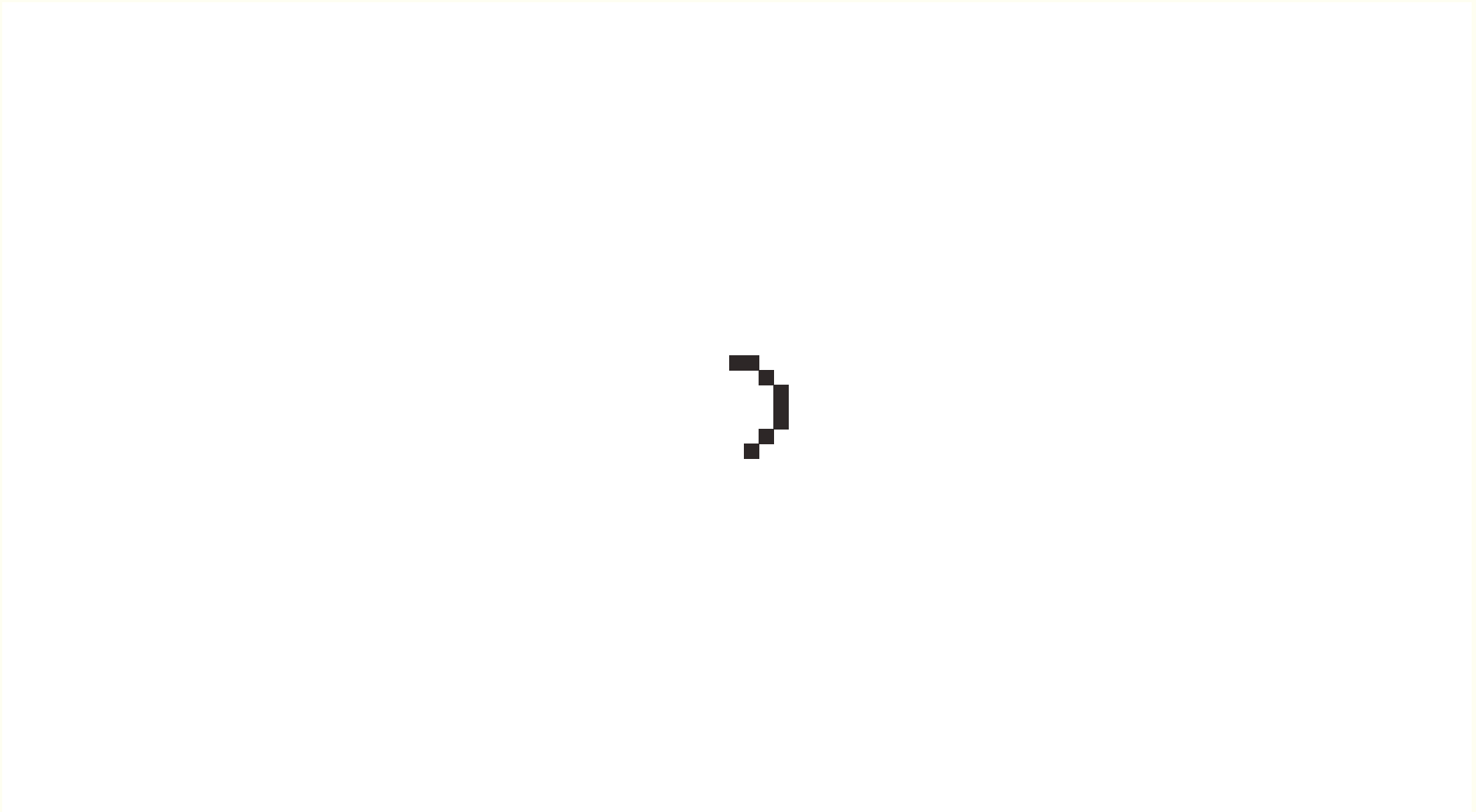


They can be memorable and insightfull



Source: Airparif

...and also beautiful



Outline

1. Usefulness and importance

2. Key data viz principles

3. Building a graph

4. Data viz for research in economics

Usefulness and importance

Types of graphs

Graphs to explore

- Analyze
- Confirm

Graphs to explain

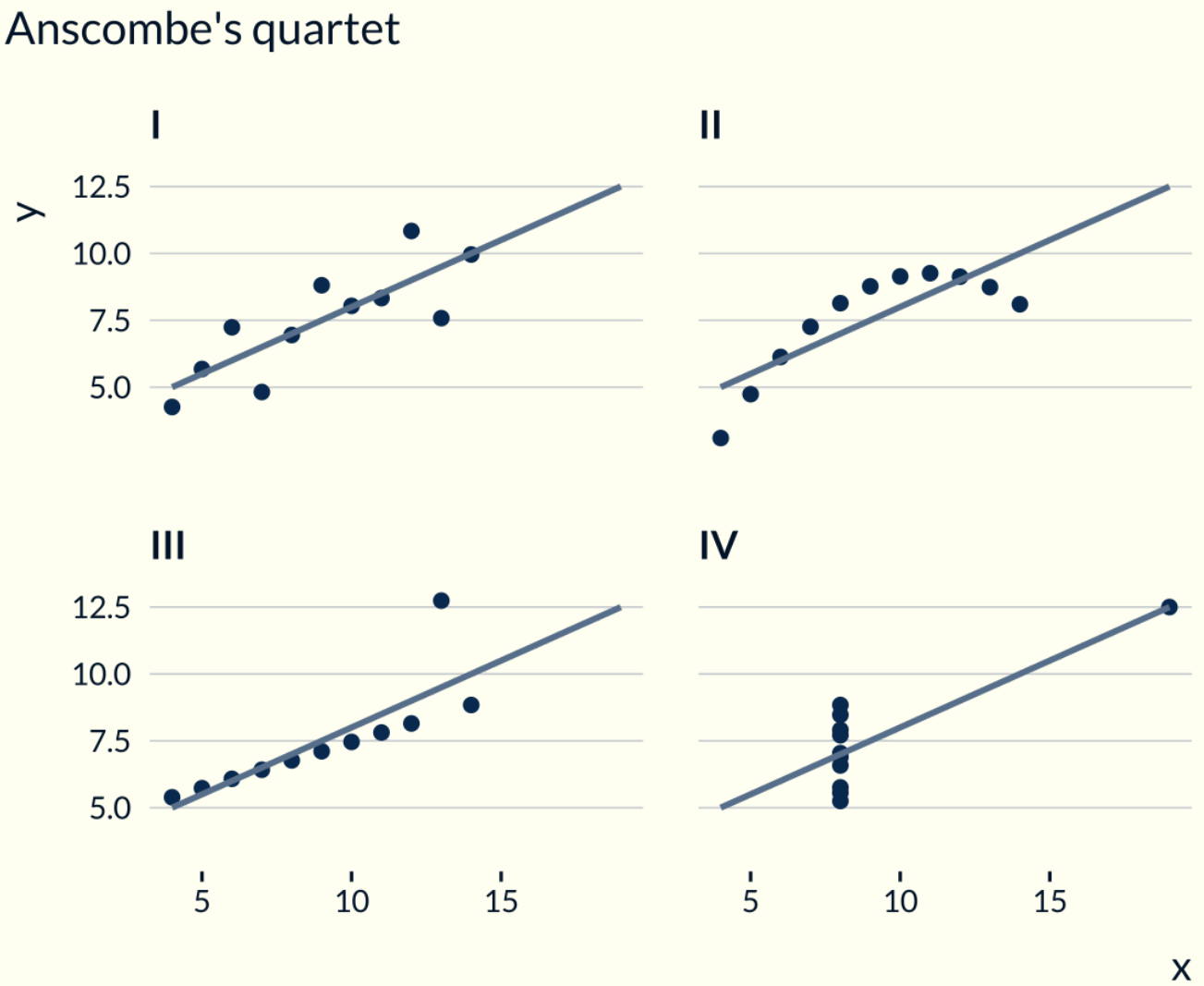
- Inform
- Convince

They have **different goals and audiences**

Explore: make sense of your data

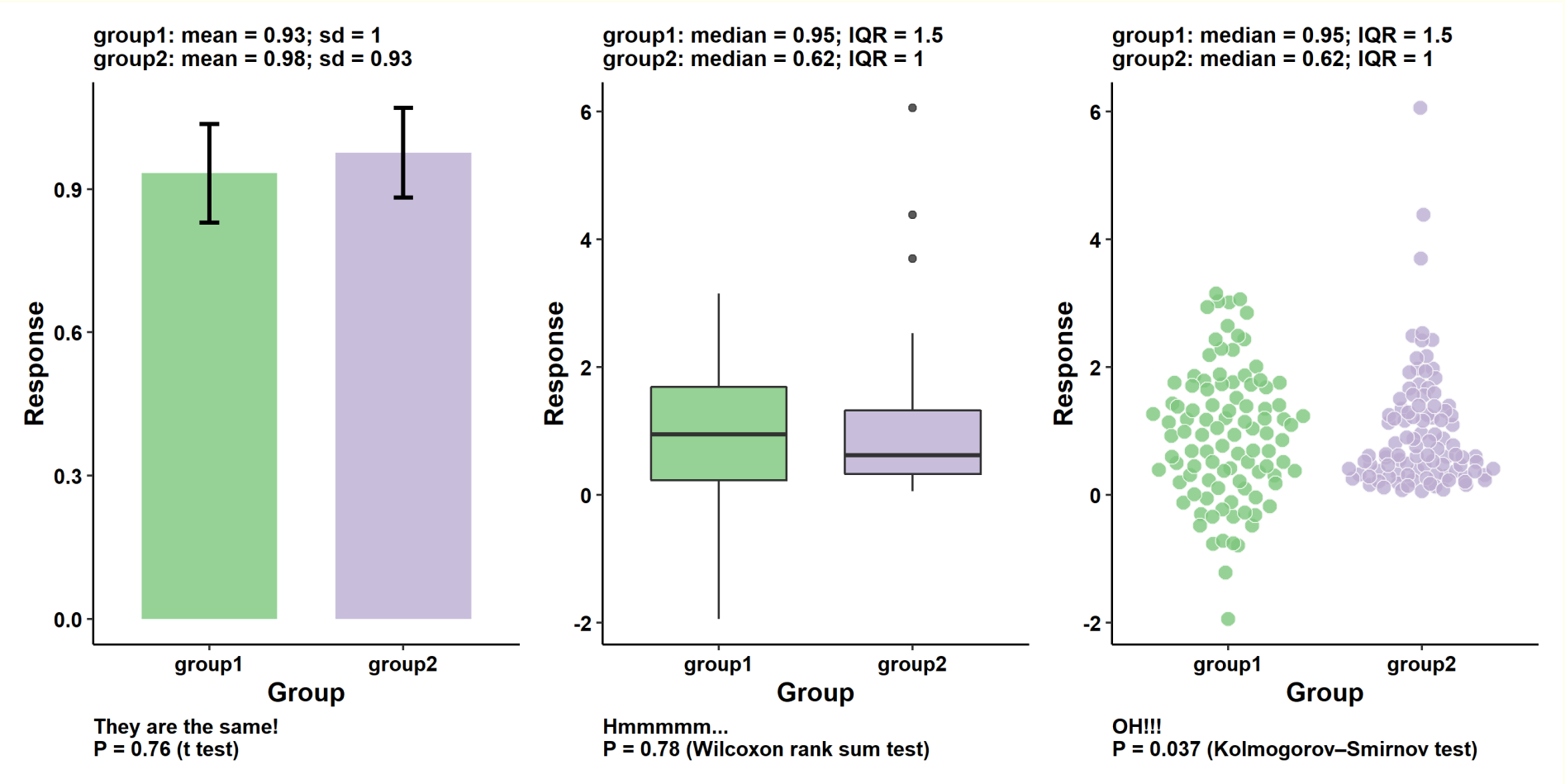
- Data often contain patterns; data viz can be tremendously helpful to identify them
- But need to look at your raw data
- That's the role of the exploratory data analysis (**EDA**)
- It may also help you formulating hypotheses (to test on other data sets)

Look at your raw data



Same relation, different patterns

Look at your raw data

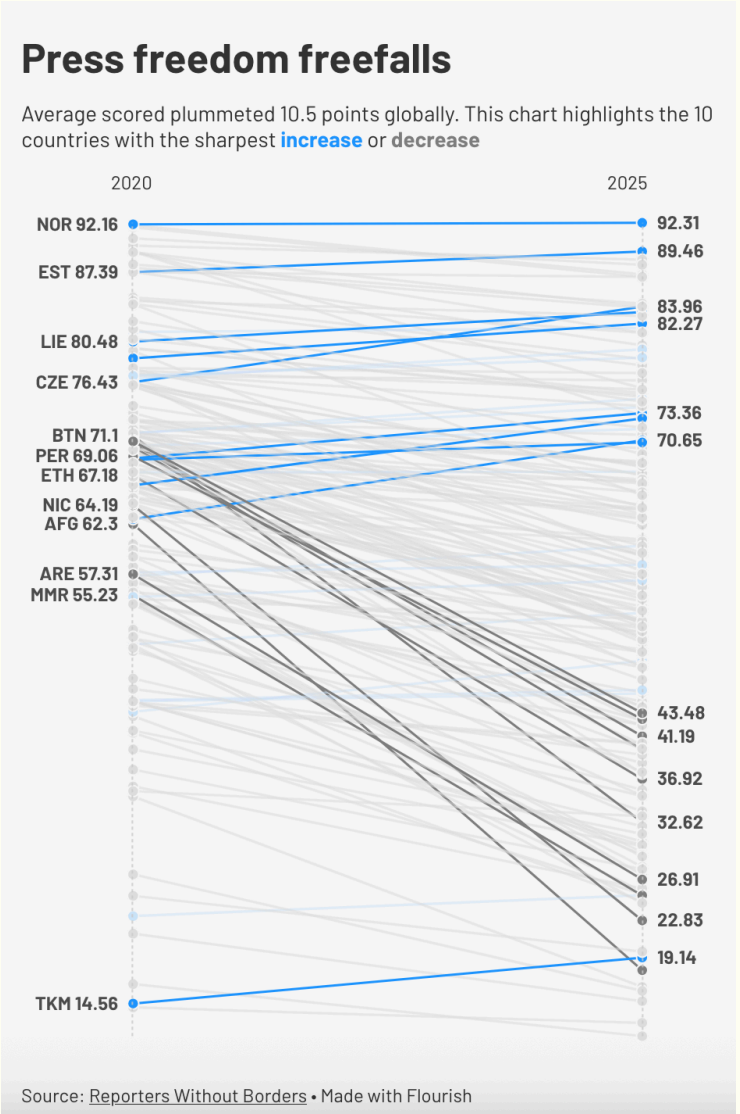


Some plots (or summary statistics) help summarize but can also **hide** information

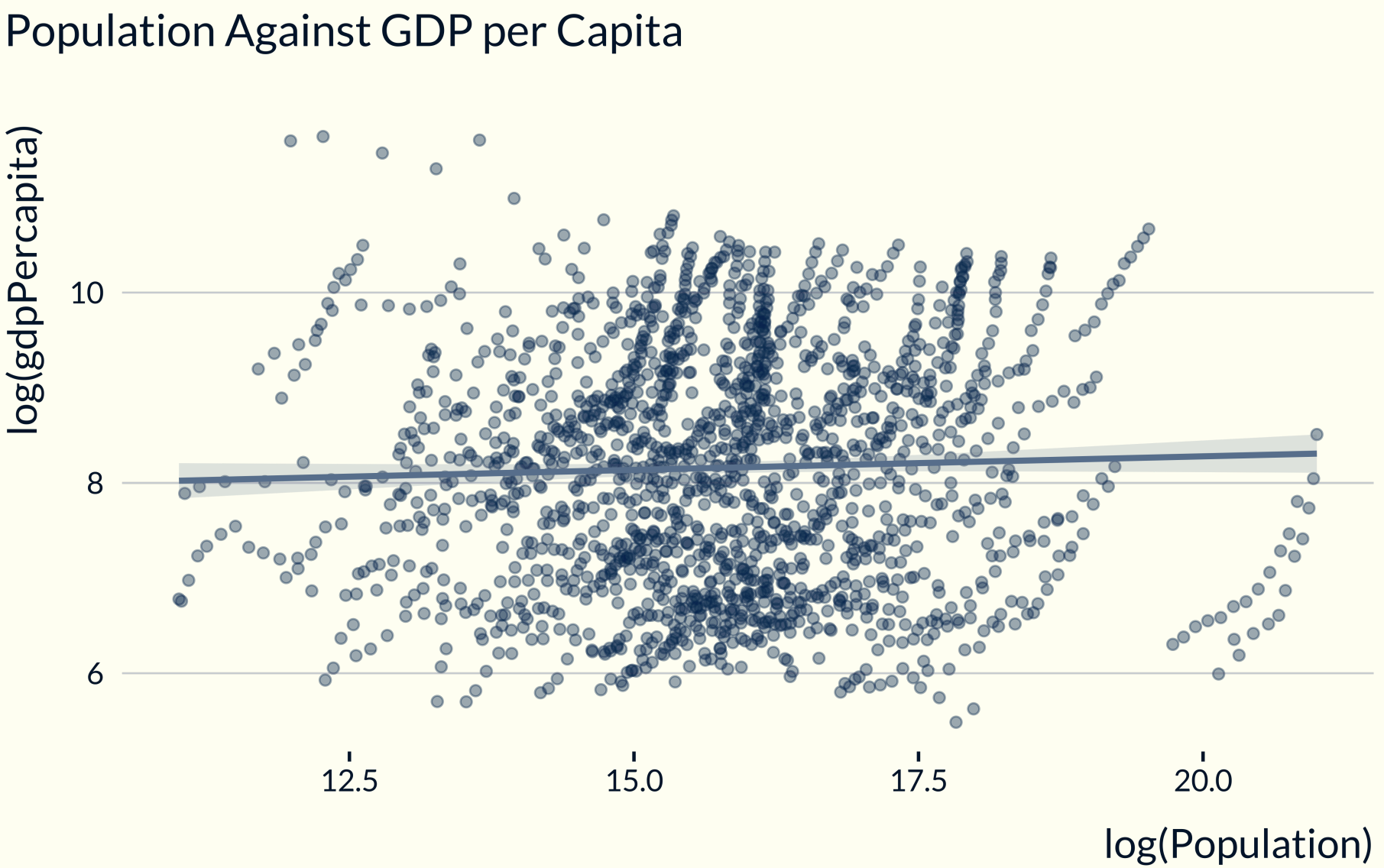
Explain: communicate your results

Country	ISO	2020	2020 Ranking	2025	2025 Ranking	Region	Pattern
Afghanistan	AFG	62.30	122	17.88	175	Asia-Pacific	Lower
Albania	ALB	69.75	84	58.18	80	EU & Balkans	Lower
Algeria	DZA	54.48	146	44.64	126	MENA	Lower
Andorra	AND	76.77	37	63.30	65	EU & Balkans	Lower
Angola	AGO	66.08	106	52.67	100	Africa	Lower
Argentina	ARG	71.22	64	56.14	87	Americas	Lower
Armenia	ARM	71.40	61	73.96	34	EECA	Higher
Australia	AUS	79.79	26	75.15	29	Asia-Pacific	Lower

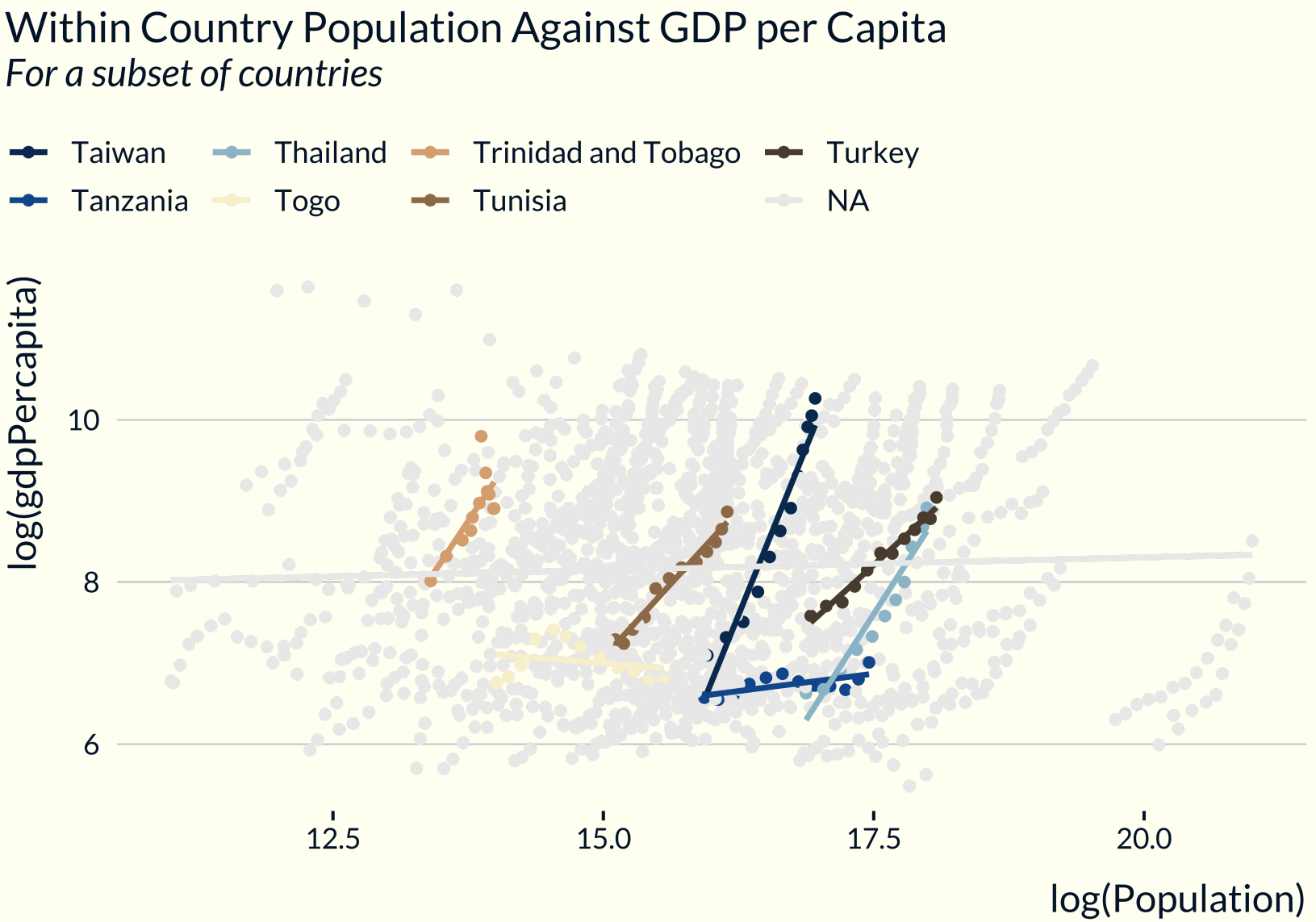
Getting your point across



Data viz as a rhetorical tool



Data viz as a rhetorical tool



The power of data viz

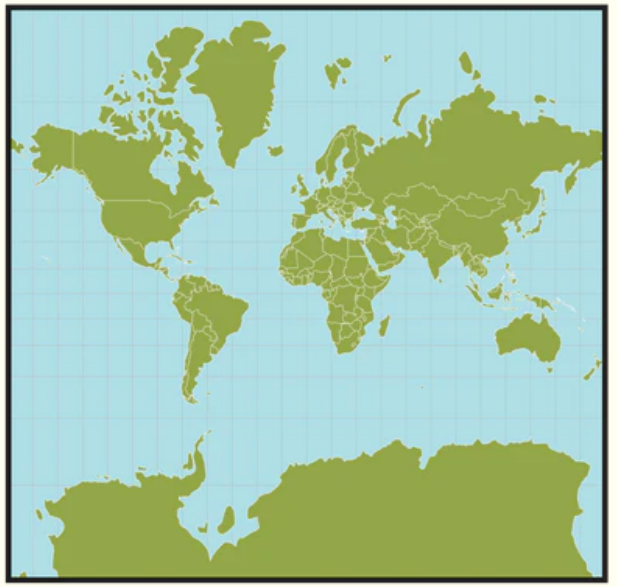
> *We can easily see patterns presented in certain ways, but if they are presented in other ways, they become invisible [..]*
>
> *Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out. If we disobey the rules, our data will be incomprehensible or misleading.*
>
> Ware, C. (2012). Information Visualization, Third Edition: Perception for Design

Data viz can be misleading

- We have briefly discussed that before
- There is a breadth of ways in which they can be misleading
- Charts can be wrong. They can also be correct BUT misleading
- See *Defense Against Dishonest Charts* on Flowing Data

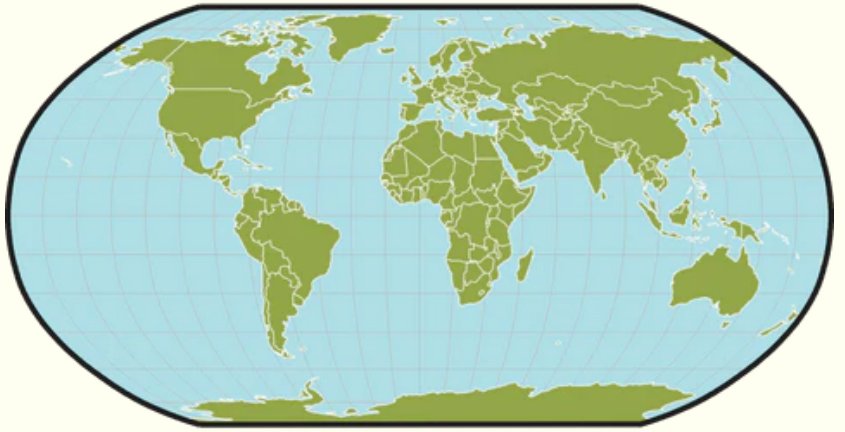
Map projections

MERCATOR



Gerardus Mercator - 1569

ROBINSON



Arthur H. Robinson - 1963

Cutting 0 on the y-axis

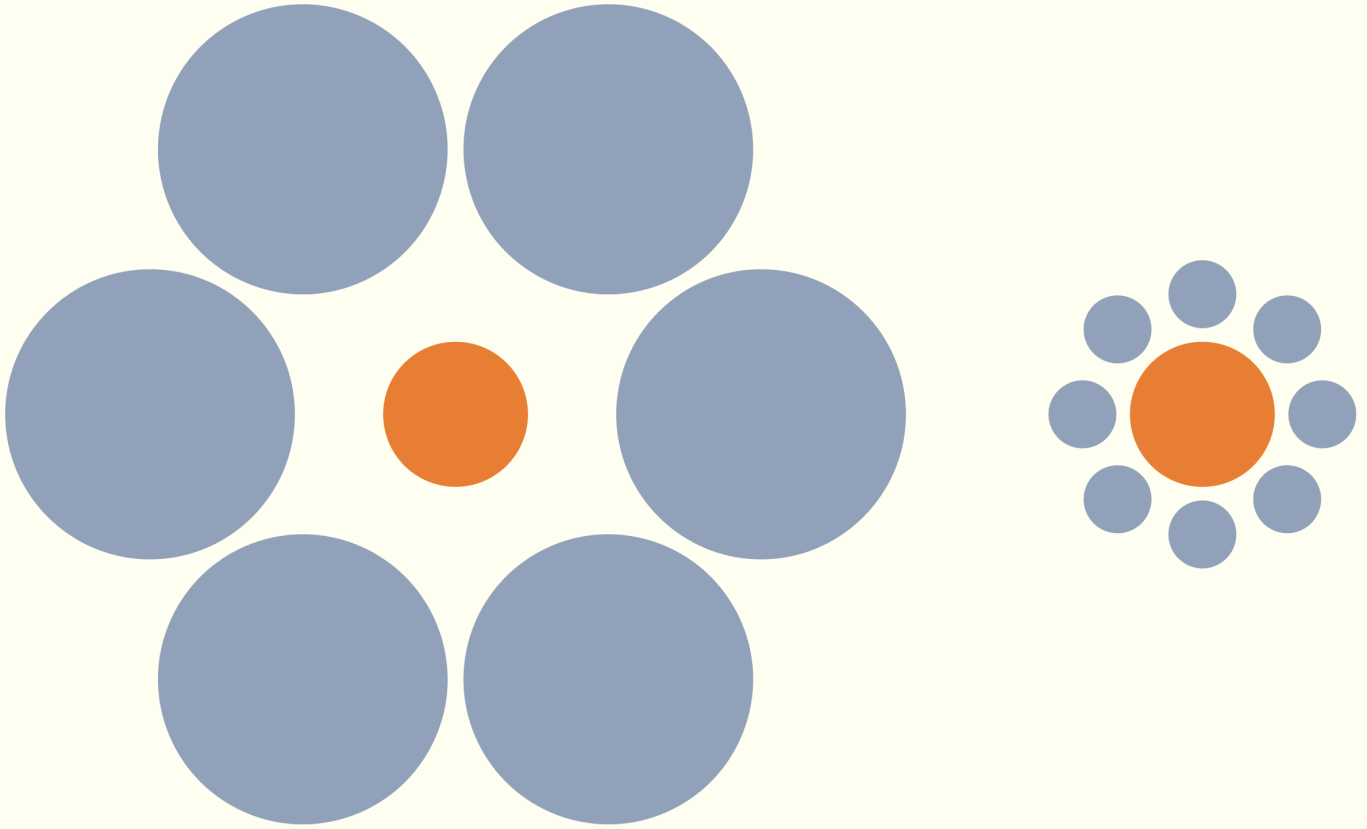
Shut up about the y-axis. It shouldn't always start at zero.

Key data viz principles

Theory

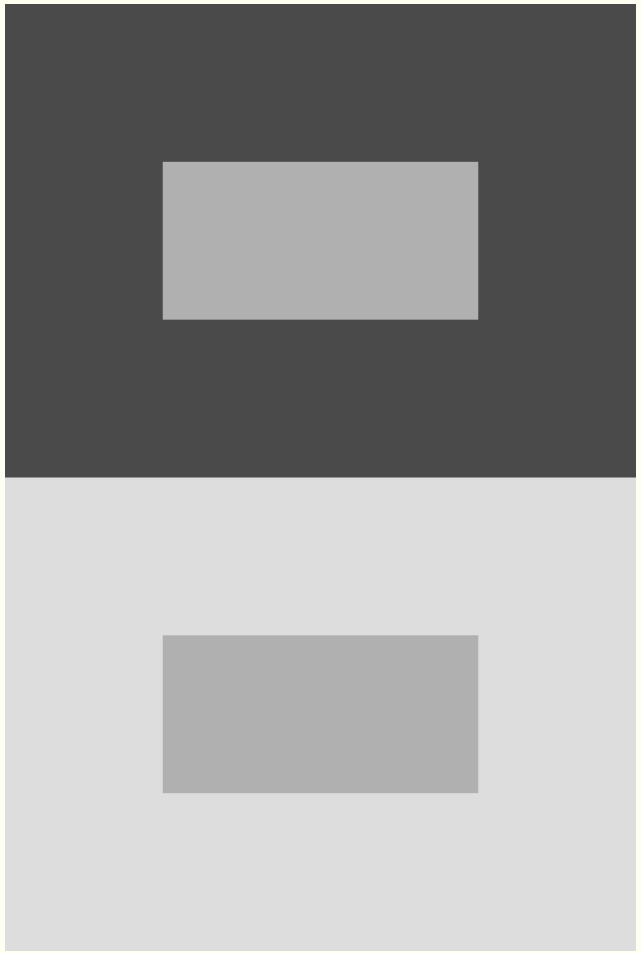
- There is actually a lot of theory behind data viz:
  - Perception,
  - Colors,
  - Design,
  - etc
- Worth learning about it and being aware of key principles
- Leverage it to make better data viz

Perception



Ebbinghaus illusion

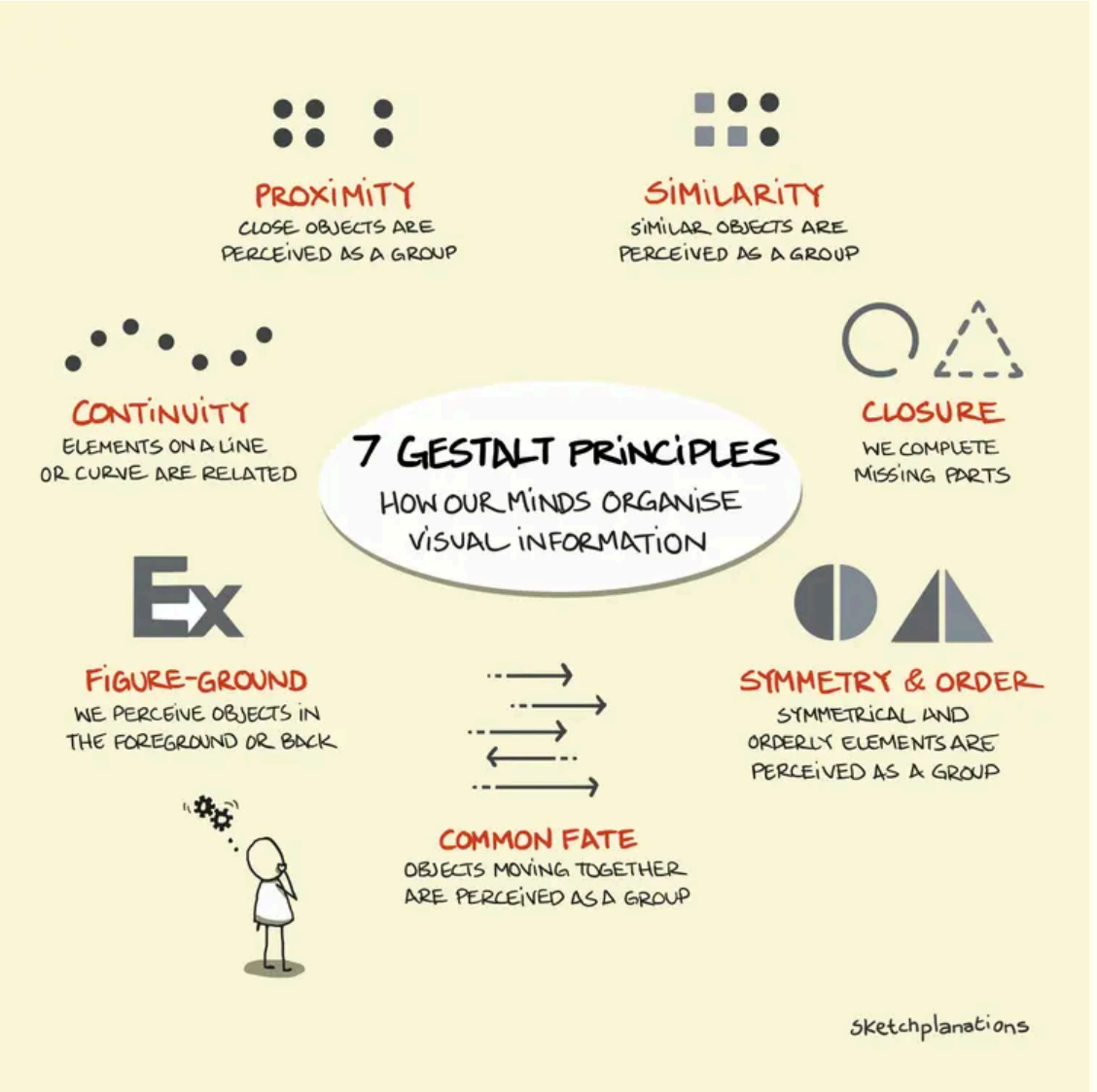
Relative differences matter



Law of simultaneous contrast

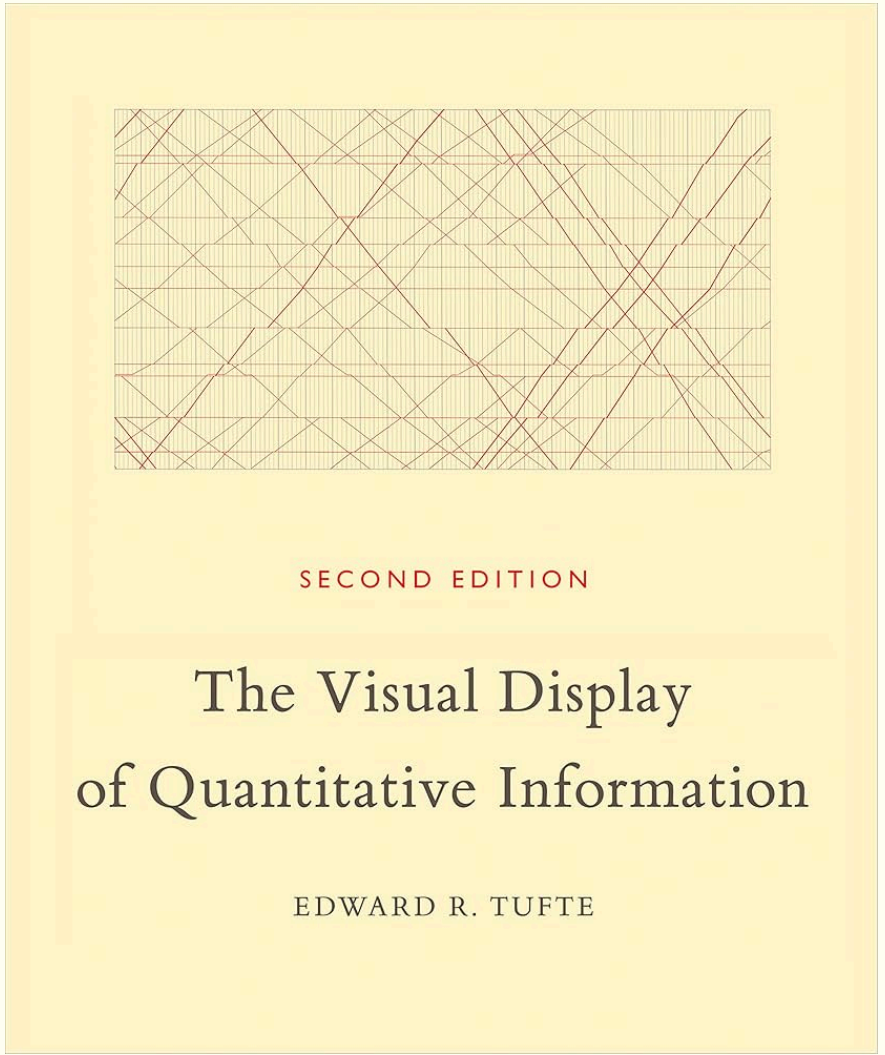
Gestalt principles

- How our brain interprets what we see
- How it organize visual information
- How we group elements together
- Use them to highlight some patterns and downplay others



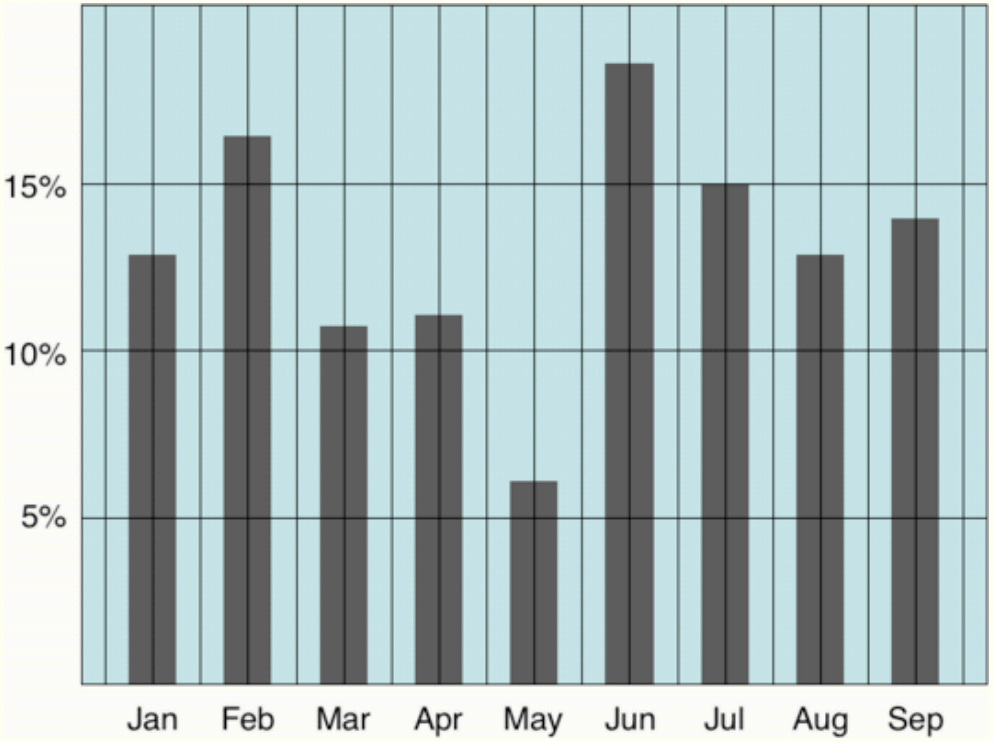
Data-to-ink ratio

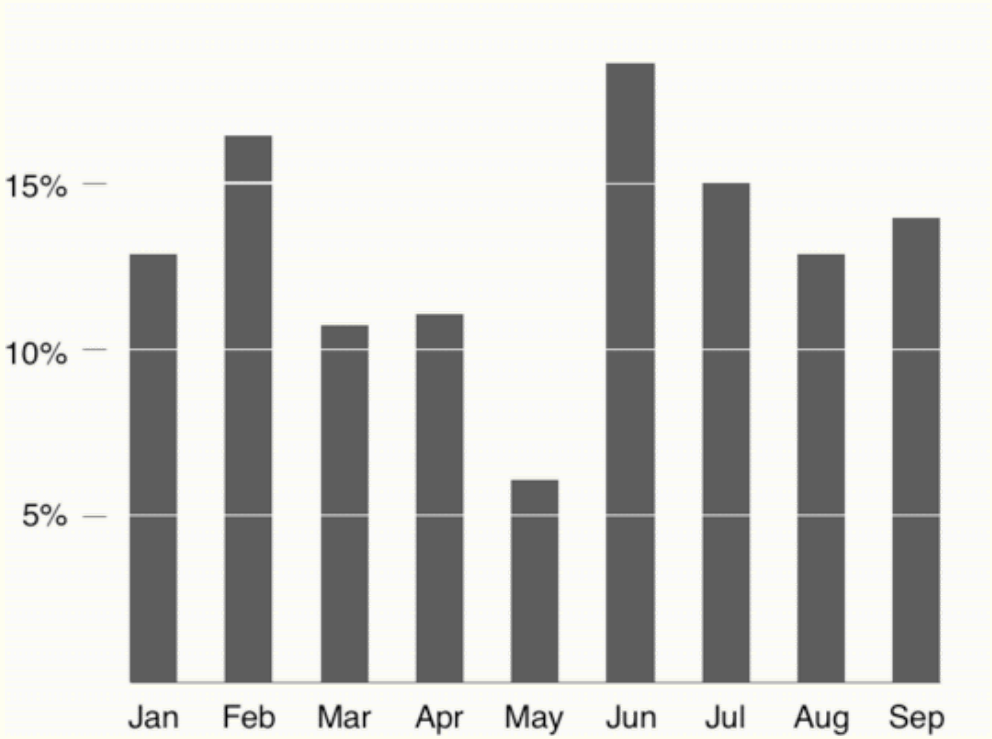
- Introduced by Edward Tufte
- In a nutshell: **avoid clutter**
- Erase non-essential and redundant information
- *“Above all else show the data”*
- Sometimes good to break these rules



Data-to-ink ratio

$$\text{data-ink ratio} = \frac{\text{data ink}}{\text{total ink on graph}}$$

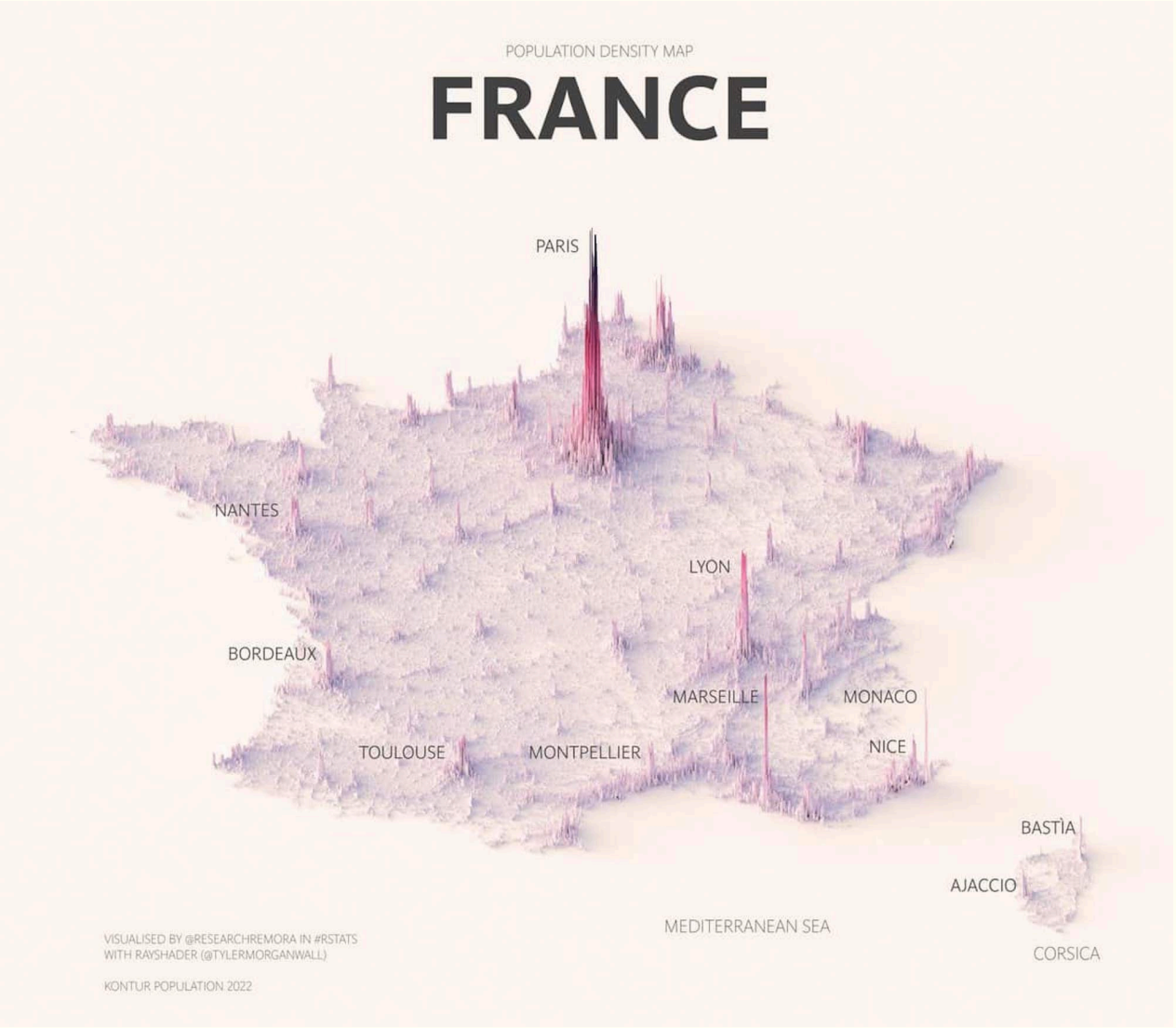




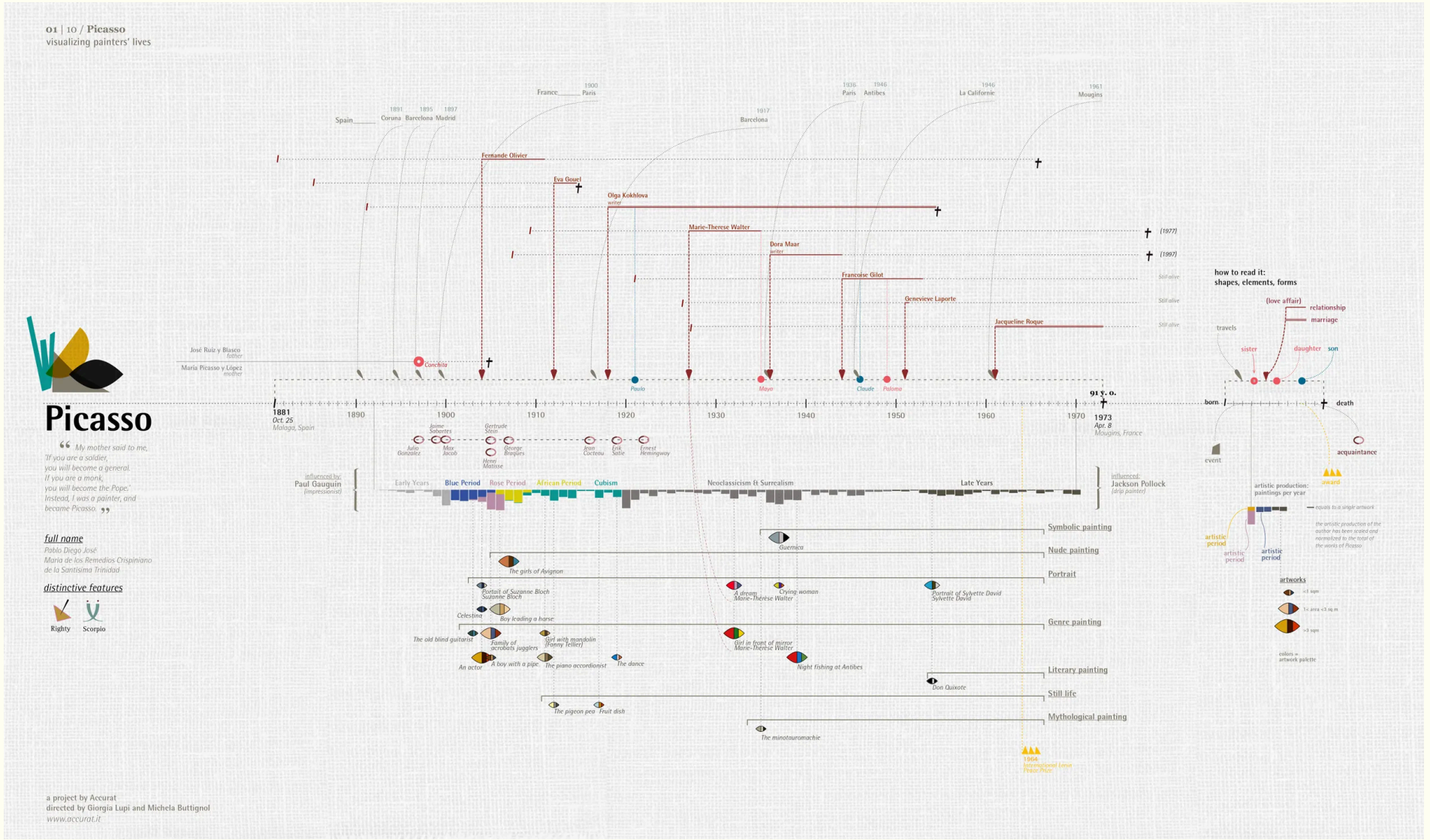
Graph aesthetics

- Why make nice looking visualizations?
- To trigger interest, to intrigue, to **catch the eye**
- That affects how people perceive information
- Nice looking visuals may be more **memorable**
- In that sense, not everything is *chartjunk*

Pretty and memorable



Engaging the audience



Graph aesthetics in academia

- Pretty graphs are useful in data journalism and so on, but what about academia?
- They are also only more pleasing to look at, they also make readers want to **engage more** with them
- Maybe better to keep the design rather minimalist
- Pretty does not always mean non-simple. Simple graphs have value.
- Opinion on **credibility** partly based on aesthetics

> "This paper did not receive the care it deserved" *comment given to a now senior researcher when they submitted a paper with a sketchy graph*

Credibility and aesthetics

VIOLET S. MANGANESE

5419 HOLLYWOOD BLVD. STE. C731, LOS ANGELES CA 90027 (323) 555-1435 VIOLET @GMAIL.COM

Education

UCLA Anderson School of Management — **Los Angeles, California**
August 2011 to June 2013
- Cumulative GPA: 3.98
- Academic interests: real-estate financing, corporations, money
- Henry Murtaugh Award

Hartford University — **Cambridge, Massachusetts**
September 2003 to June 2007
- B.A. summa cum laude, Economics
- Extensive coursework in Astrophysics, Statistics
- Van Damme Scholarship

Business experience

Boxer Bedley & Ball Capital Advisors — **New York, New York**
June 2008 to August 2011
Equity Analyst
- Performed independent research on numerous American industries, including:
- Steelmaking, croquet, semiotics, and butterscotch manufacturing
- Led company in equities analyzed in two quarters

Other work experience

Proximate Cause — **Los Angeles, California**
June 2007 to May 2008
Assistant to the Director
- Helped devise fundraising campaigns for this innovative nonprofit
- Handled lunch orders and general errands

Hot Topic — **Boston, Massachusetts**
February 2004 to March 2006
Retail sales associate
- Inventory management
- Training and recruiting

Skills and interests

- Fluent in Mandarin, Esperanto; conversational knowledge of Gaelic
- Writer of U.S. Senate-themed fan fiction
- Ocean kayaking and free diving
- Travel, cooking, hiking, playing with my dog
- Ceramics
- Backgammon
- Making paper planes

TRIXIE B. ARGON

5419 HOLLYWOOD BLVD STE C731, LOS ANGELES CA 90027
(323) 555 1435 TRIXIEARGON @ GMAIL.COM

EDUCATION

UCLA Anderson School of Management — 2011–13
- Cumulative GPA: 3.98
- Academic interests: real-estate financing, criminal procedure, corporations
- Henry Murtaugh Award

Hartford University — 2003–07
- B.A. *summa cum laude*, Economics
- Extensive coursework in Astrophysics, Statistics
- Van Damme Scholarship

BUSINESS EXPERIENCE

Boxer Bedley & Ball Capital Advisors — 2008–11
Equity analyst
- Performed independent research on numerous American industries, including:
- Steelmaking, croquet, semiotics, and butterscotch manufacturing
- Led company in equities analyzed in two quarters

OTHER WORK EXPERIENCE

Proximate Cause — 2007–08
Assistant to the director
- Helped devise fundraising campaigns for this innovative nonprofit
- Handled lunch orders and general errands

Hot Topic — 2004–06
Retail-sales associate
- Top in-store sales associate in seven out of eight quarters
- Inventory management
- Training and recruiting

“Originality” VS “familiarity”

- **Original** graphs may trigger interest
- **Familiar** graphs may convey the point more easily
- My take:
  - Use the best type of graph, regardless of its originality/familiarity
  - If it is different from what people are used to, make it easy to read

Building a graph

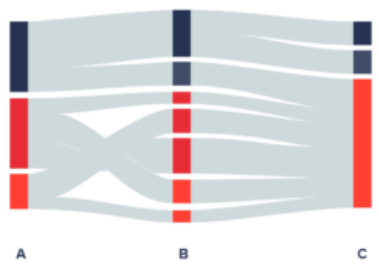
Know what chart types exist

- Know what is possible to do to find what you need
- Know **graph names** (to search the internet regarding how to code them)
- Refer to existing graph (type) galleries:
  - Data Viz Project
  - Flowing Data - Chart type
  - Dataviz Inspiration
  - R graph gallery

ALL (HTTPS://DATAVIZPROJECT.COM/) FAMILY INPUT FUNCTION SHAPE

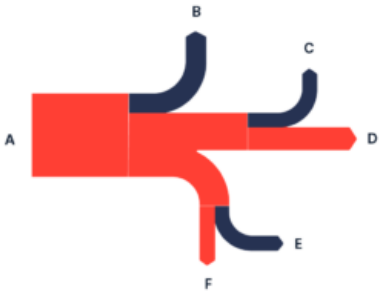
(https://datavizproject.com/)

Alluvial Diagram



(https://datavizproject.com/data-

Sankey Diagram



(https://datavizproject.com/data-

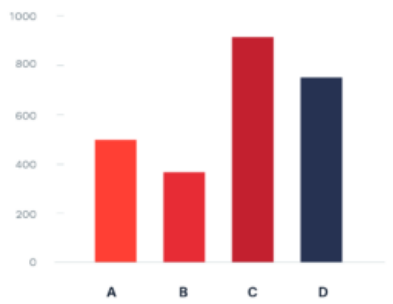
Line Graph

(https://datavizproject.com/data-

Donut Chart

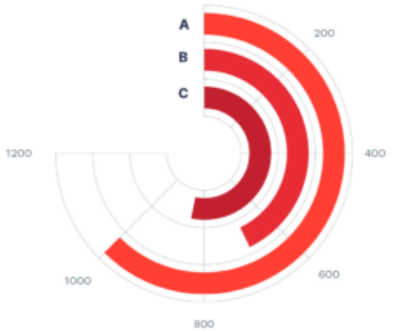
(https://datavizproject.com/data-

Bar Chart (Vertical)

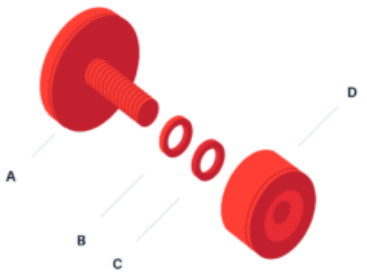


Polar Area Chart

Radial Bar Chart



Exploded View Drawing



Steps to choose a chart type

- What do you want to show in your graph?
- **What is the main message you want to convey?**
- With the same data, you can:
  - Tell a lot of different stories
  - Emphasize different points
- Making a graph = choosing a lighting for your data

1 dataset, many graphs, many stories

(https://100.datavizproject.com/)

1 dataset 100 visualizations

Can we come up with 100 visualizations from one simple dataset?

As an information design agency working with data visualization every day, we challenged ourselves to accomplish this using insightful and visually appealing visualizations.

We wanted to show the diversity and complexity of data visualization and how we can tell different stories using limited visual properties and assets.

Type of relationship to show

In your graph, you may want to show a:

- Distribution
- Evolution over time
- Magnitude
- Part of a whole
- Ranking
- Geographical patterns
- Flow
- Correlations
- Deviation

Graph type decision tree

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine

Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Distribution

Show values in a dataset and how often they occur. The shape (or skew) of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram

The standard way to show a statistical distribution – keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot

A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Cumulative curve

A good way of showing how unequal a distribution is; y axis is always cumulative frequency, x axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

Columns work well for showing change over time – but usually best with only one series of data at a time.

Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)

Use to show the uncertainty in future projections – usually this grows the further forward to projection.

Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.

Priestley timeline

Great when date and duration are key elements of the story in the data.

Circle timeline

Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline

Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram

Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph

A type of area chart, use when seeing changes in proportions over time is more important than individual values

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Bar

See above. Good when the data are not time series and labels have long category names.

Paired column

As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar

See above.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)

Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop

Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar

A space-efficient way of showing value of multiple variables– but make sure they are organised in a way that makes sense to reader.

Parallel coordinates

An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Bullet

Good for showing a measurement against the context of a target or performance range.

Grouped symbol

An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi

A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc

A hemicycle, often used for visualising parliamentary composition by number of seats.

Gridplot

Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Venn

Generally only used for schematic representation.

Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information, relationship graphs.

Sankey

Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network

Used for showing the strength and inter-connectedness of relationships of varying types.

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

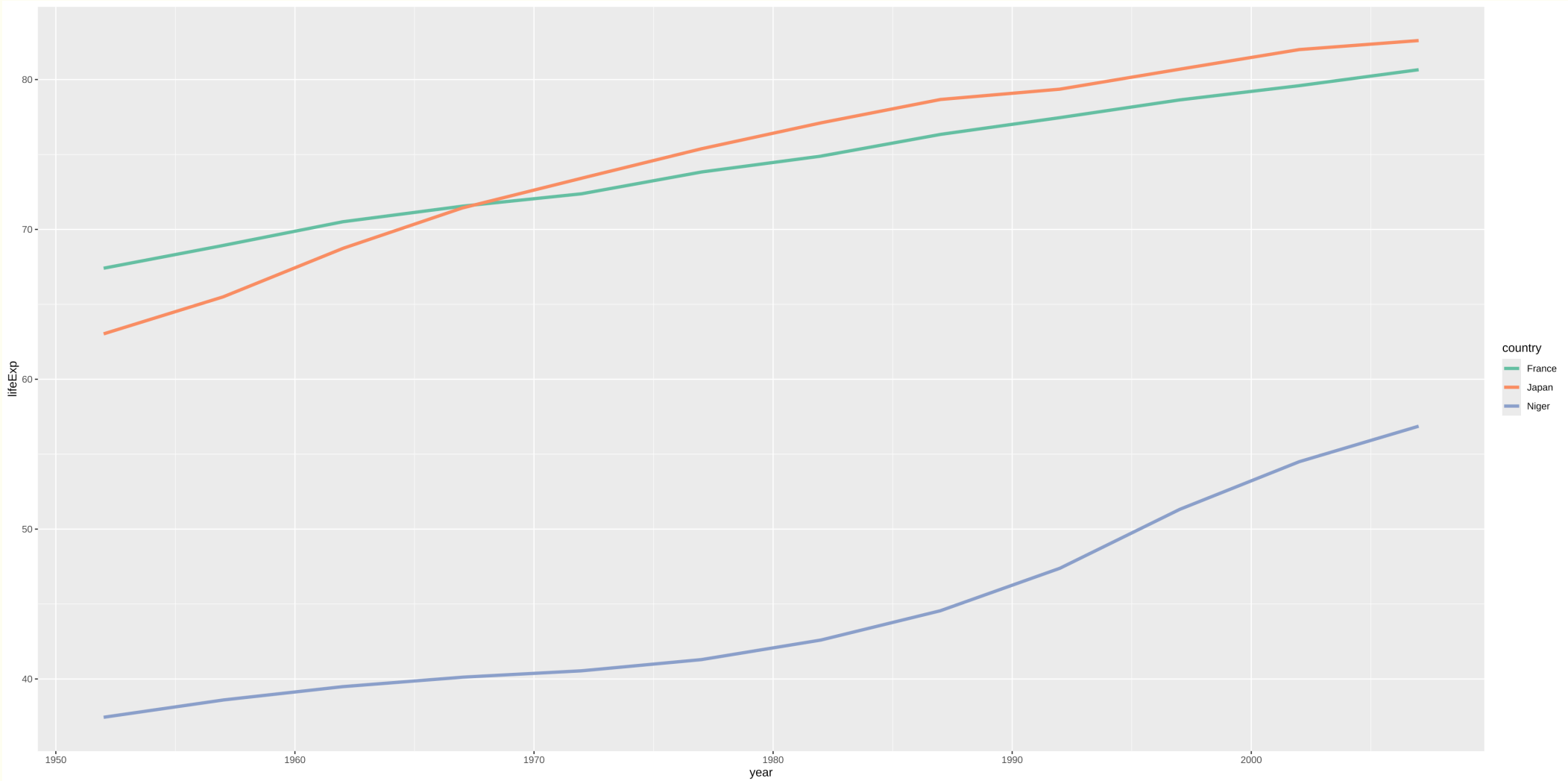
FT graphic: Alan Smith, Chris Campbell, Ian Bott, Liz Faunce, Graham Parrish, Billy Ehrenberg-Shannon, Paul McCallum, Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

ft.com/vocabulary

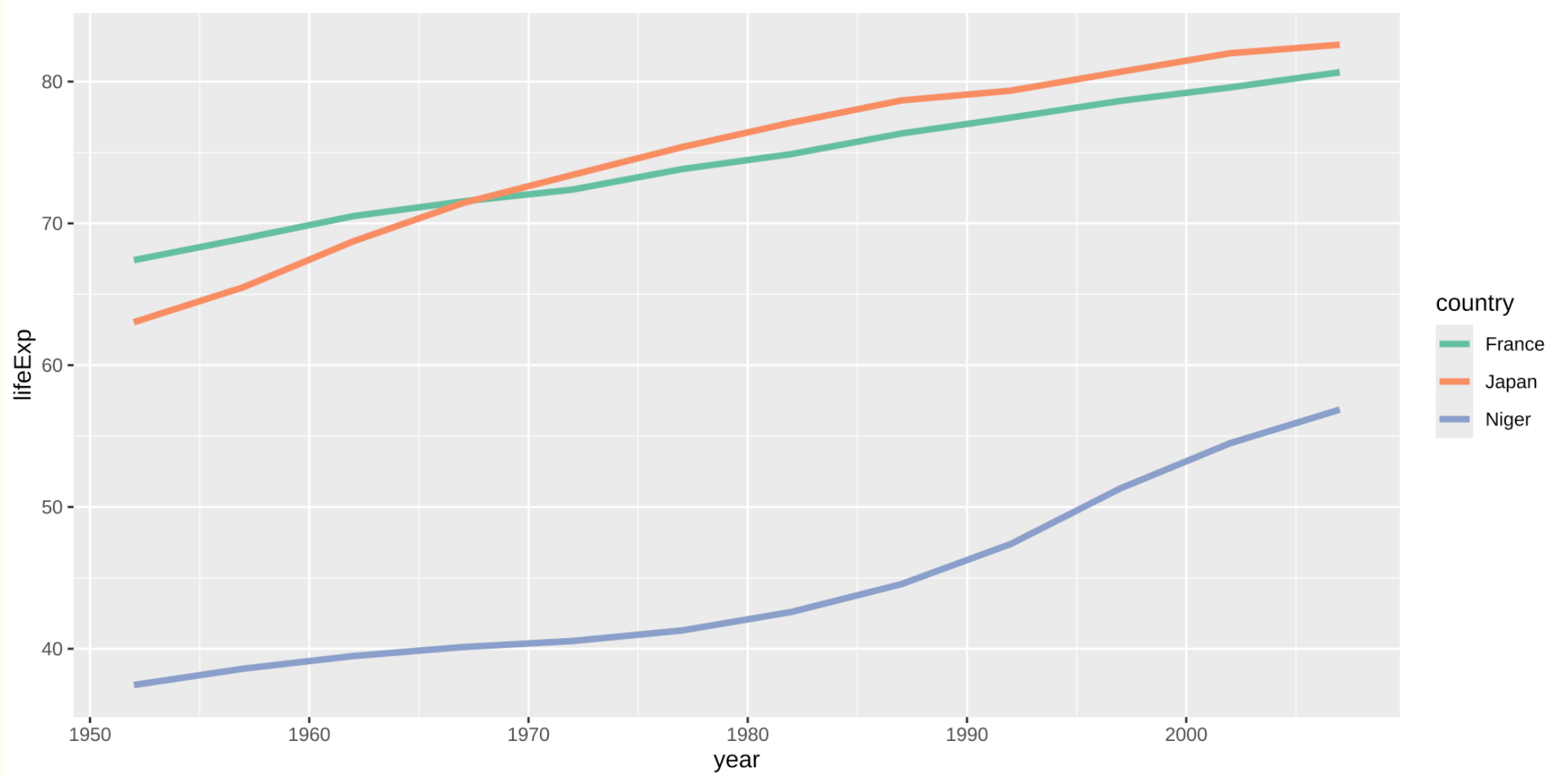
Example data

country	year	lifeExp
France	1952	67.410
France	1957	68.930
France	1962	70.510
France	1967	71.550
France	1972	72.380
France	1977	73.830
France	1982	74.890
France	1987	76.340
France	1992	77.460
France	1997	78.640
France	2002	79.590
France	2007	80.657
Japan	1952	63.030

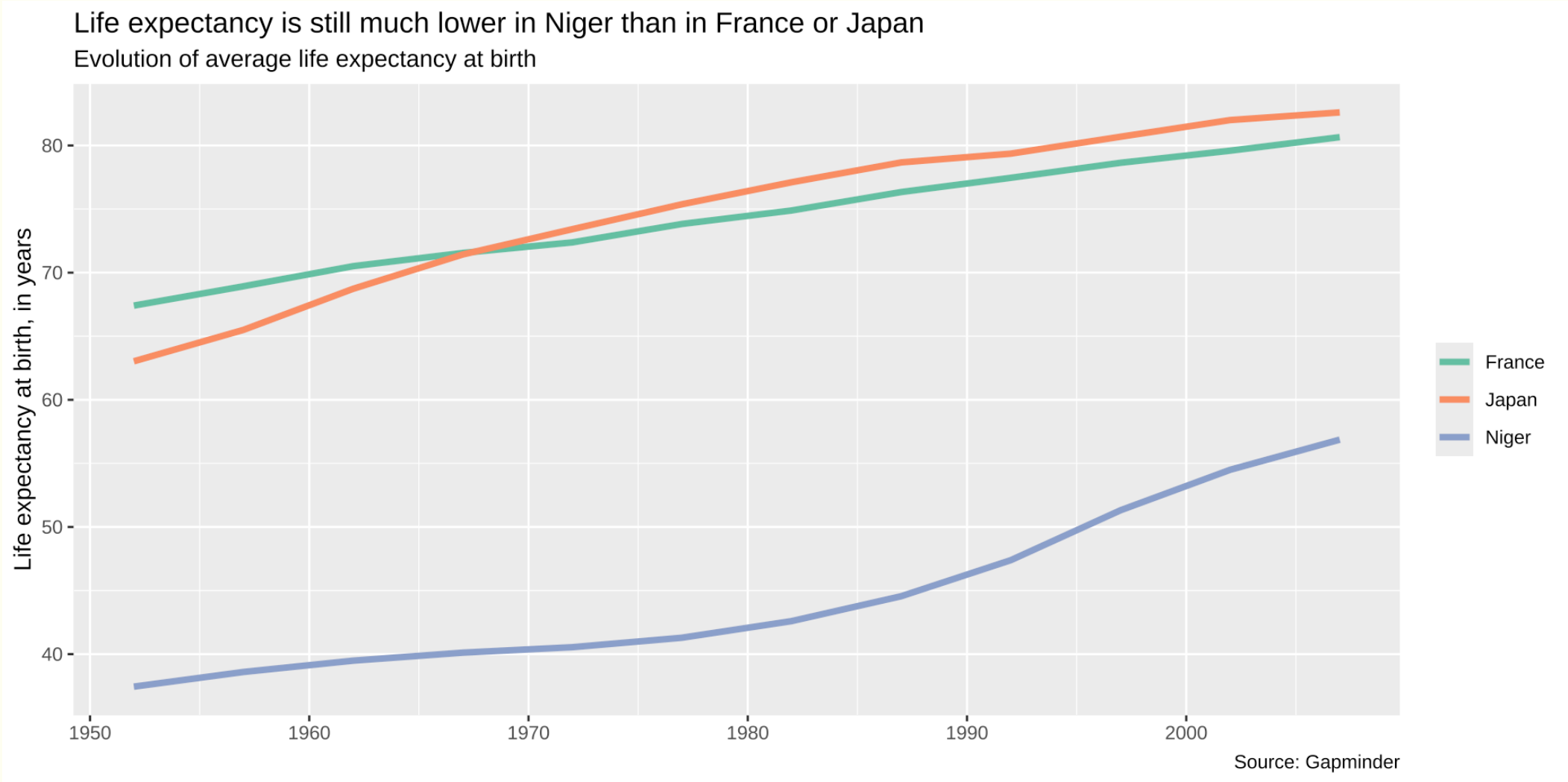
A concrete example



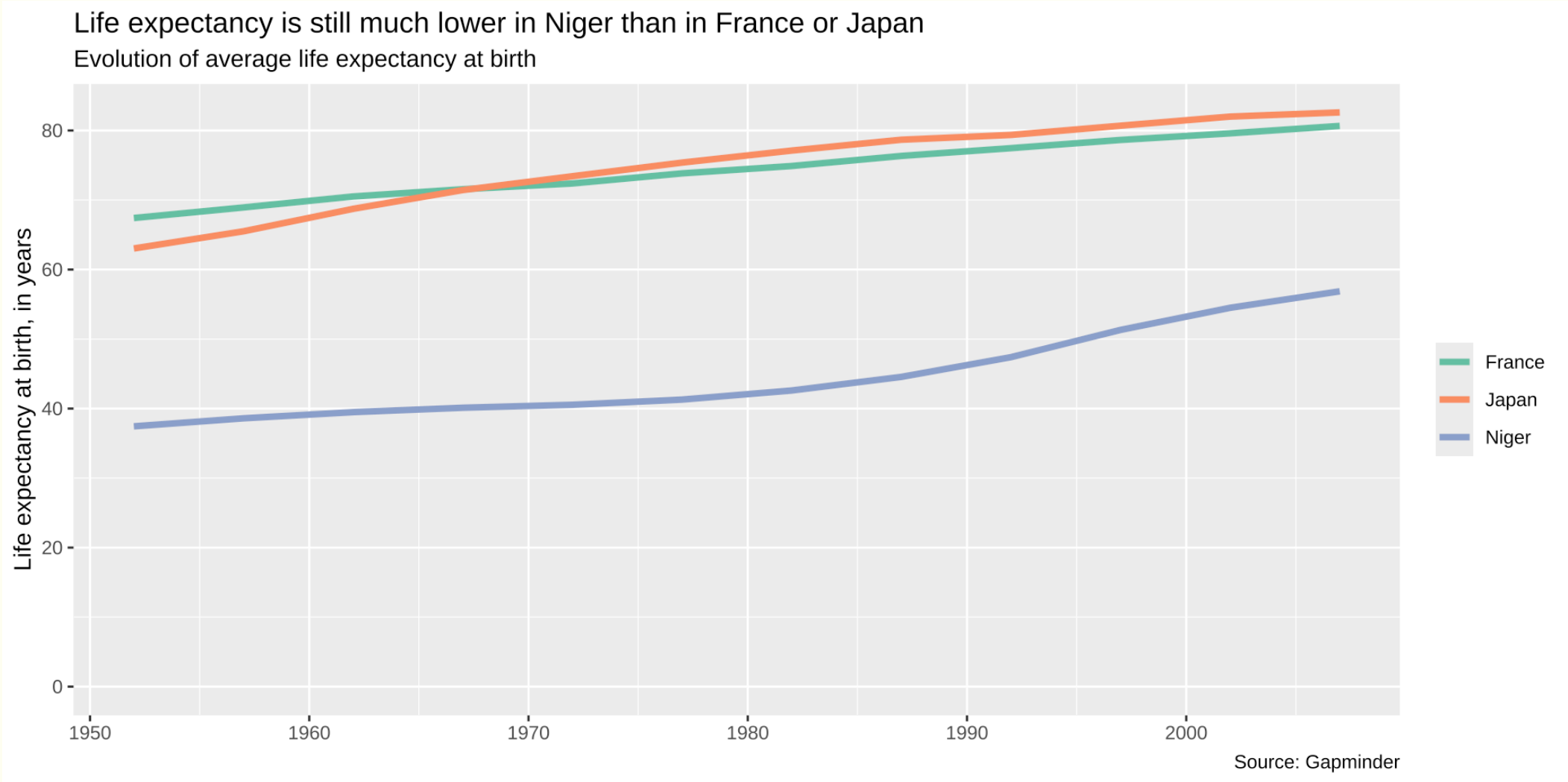
Legible text



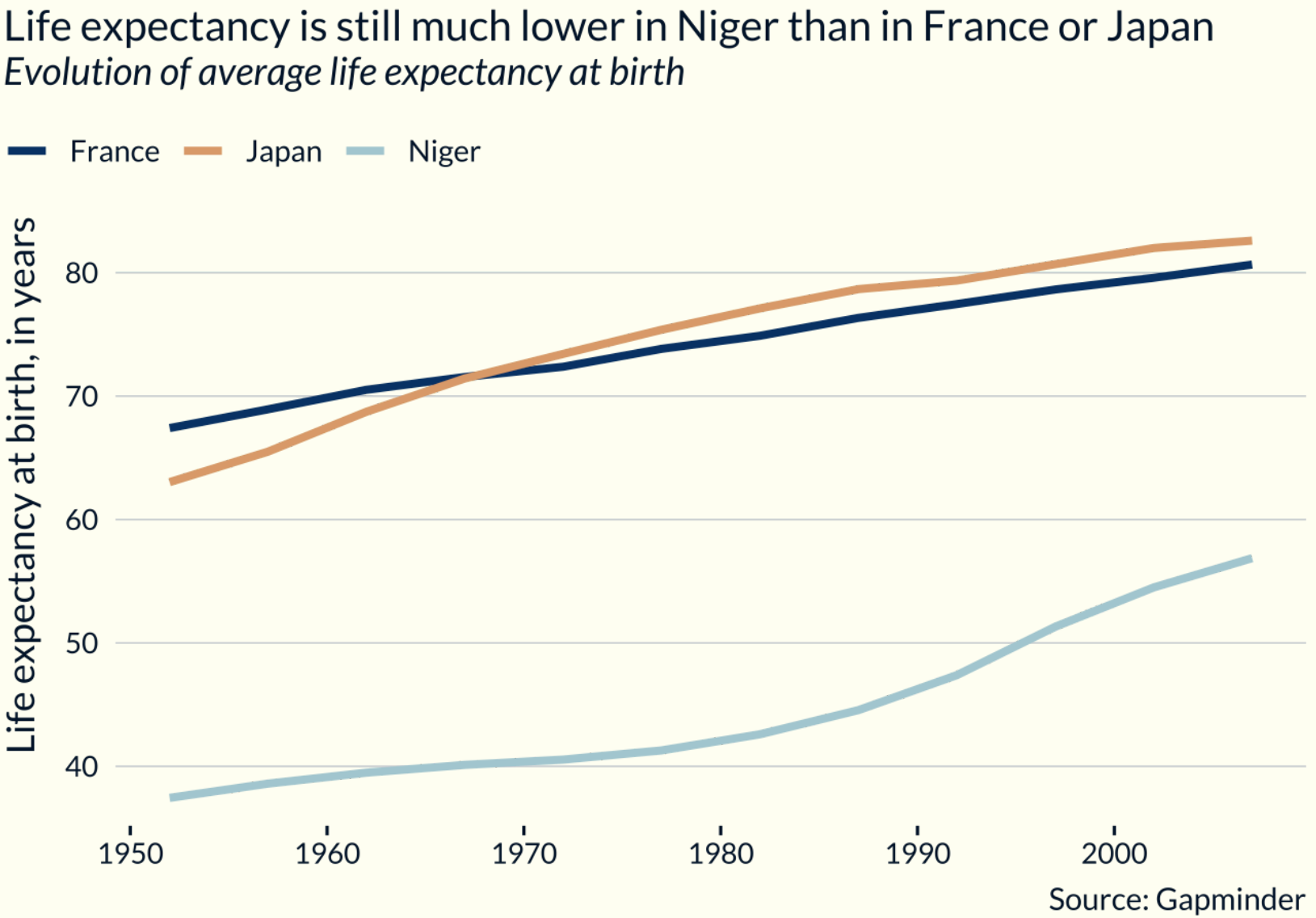
Title, clear axis labels and source



Scale of the y-axis



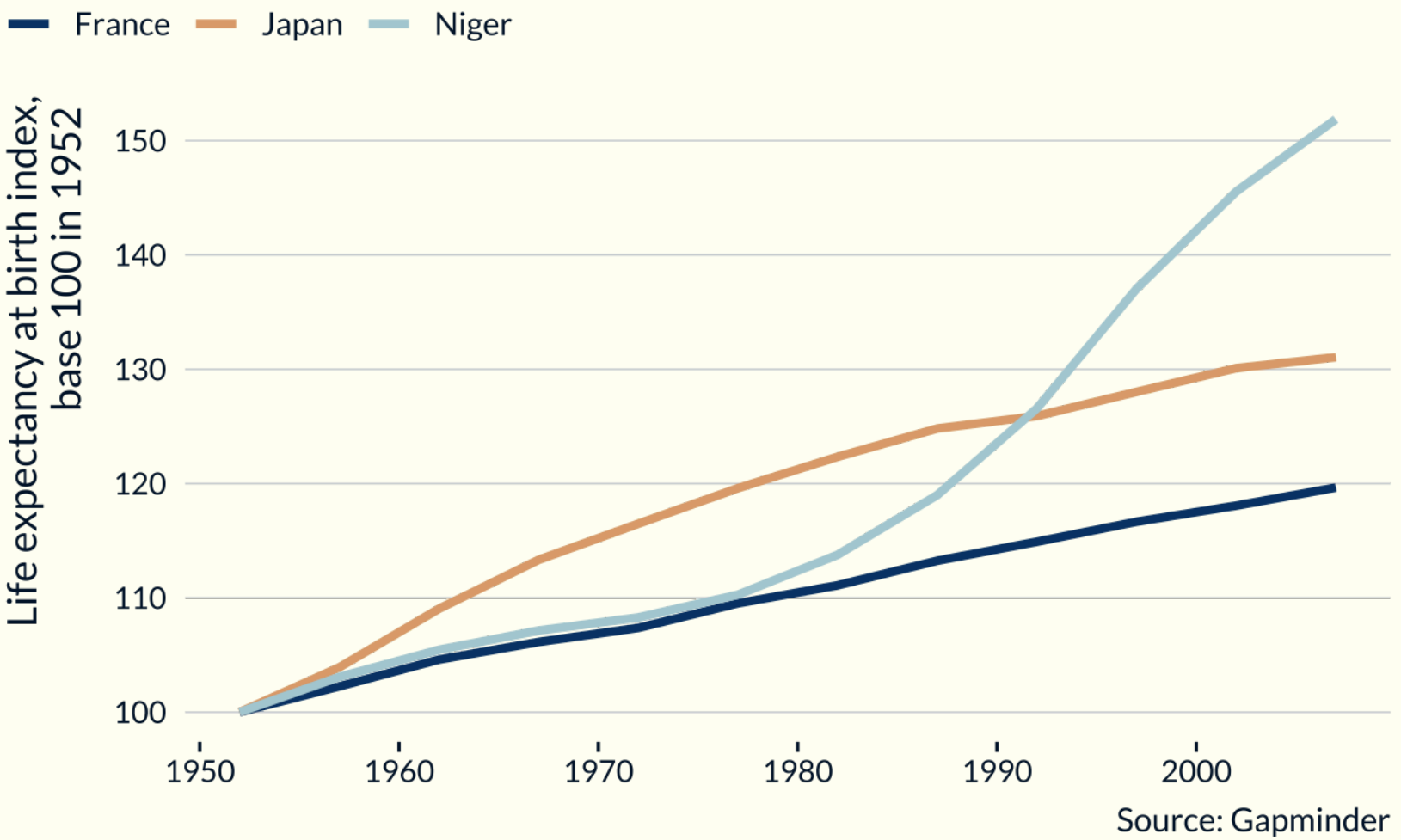
Stylize


Life expectancy is still much lower in Niger than in France or Japan
Evolution of average life expectancy at birth
France
Japan
Niger
Life expectancy at birth, in years
80
70
60
50
40
1950
1960
1970
1980
1990
2000
Source: Gapminder

An alternative story

Life expectancy at birth increased sharply in Niger since 1980

Evolution of an average life expectancy at birth index



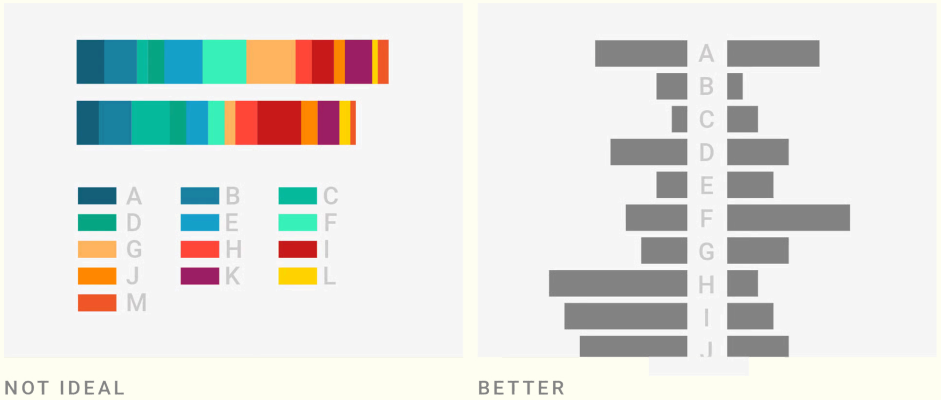
Graphs in an oral presentation

- **Explain orally what your graph represents!**
  - What is on the x-axis?
  - What is on the y-axis?
  - What is the message you want to convey?

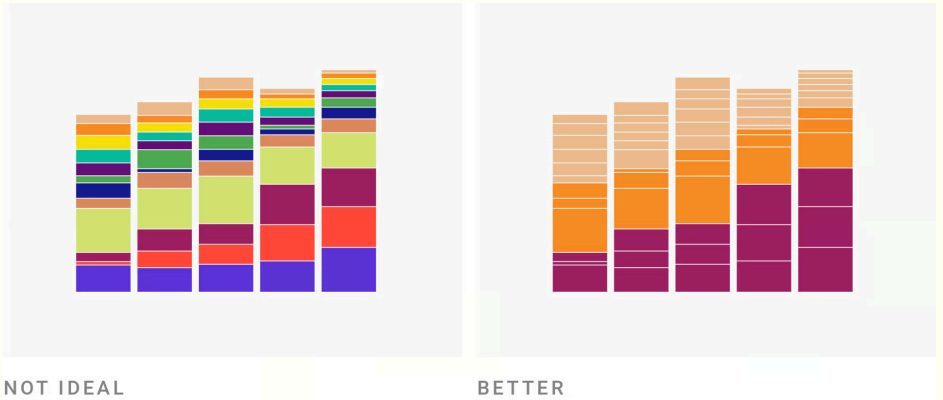
Avoid using many different colors

If need more than 7 colors or so:

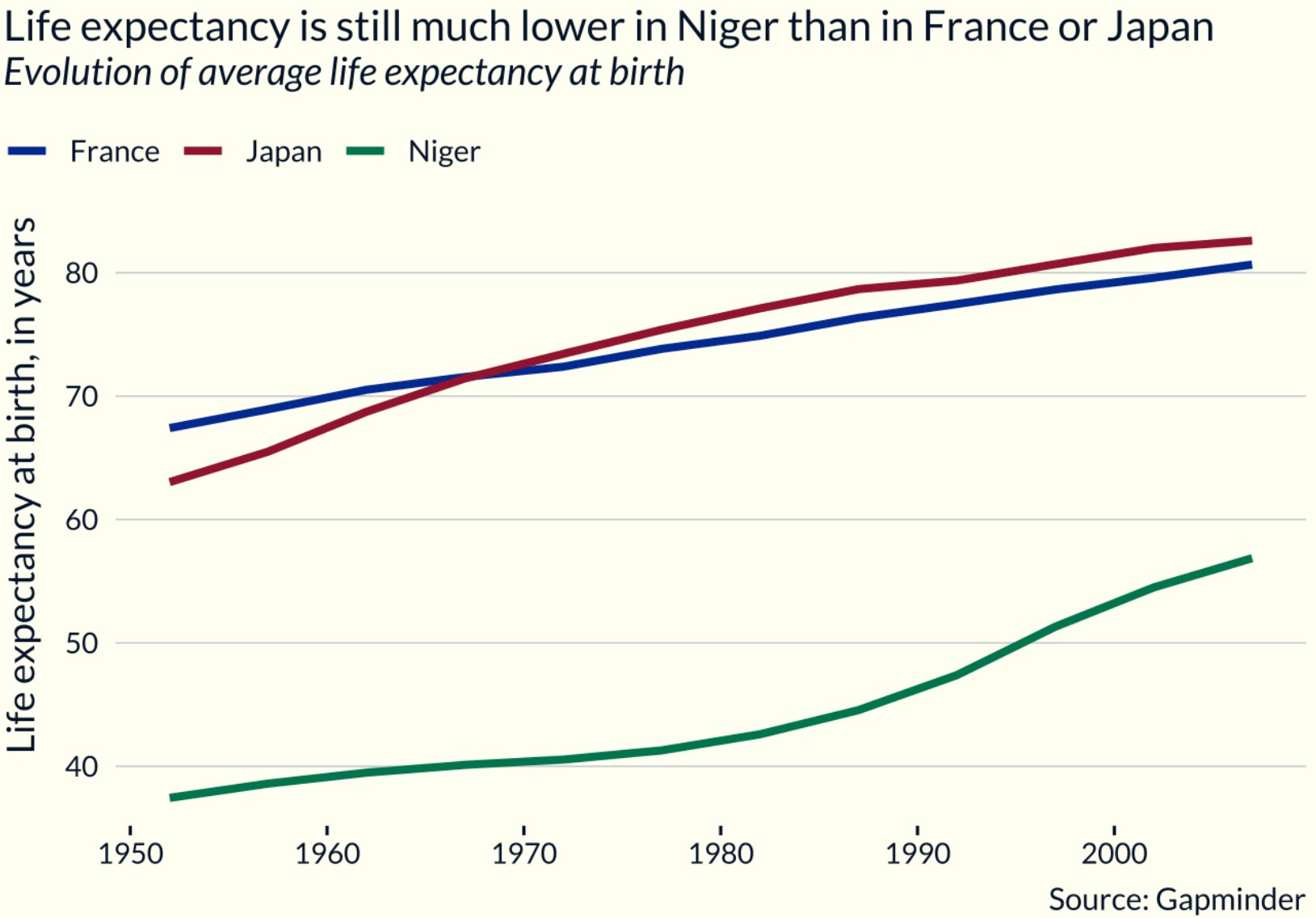
- Use another graph



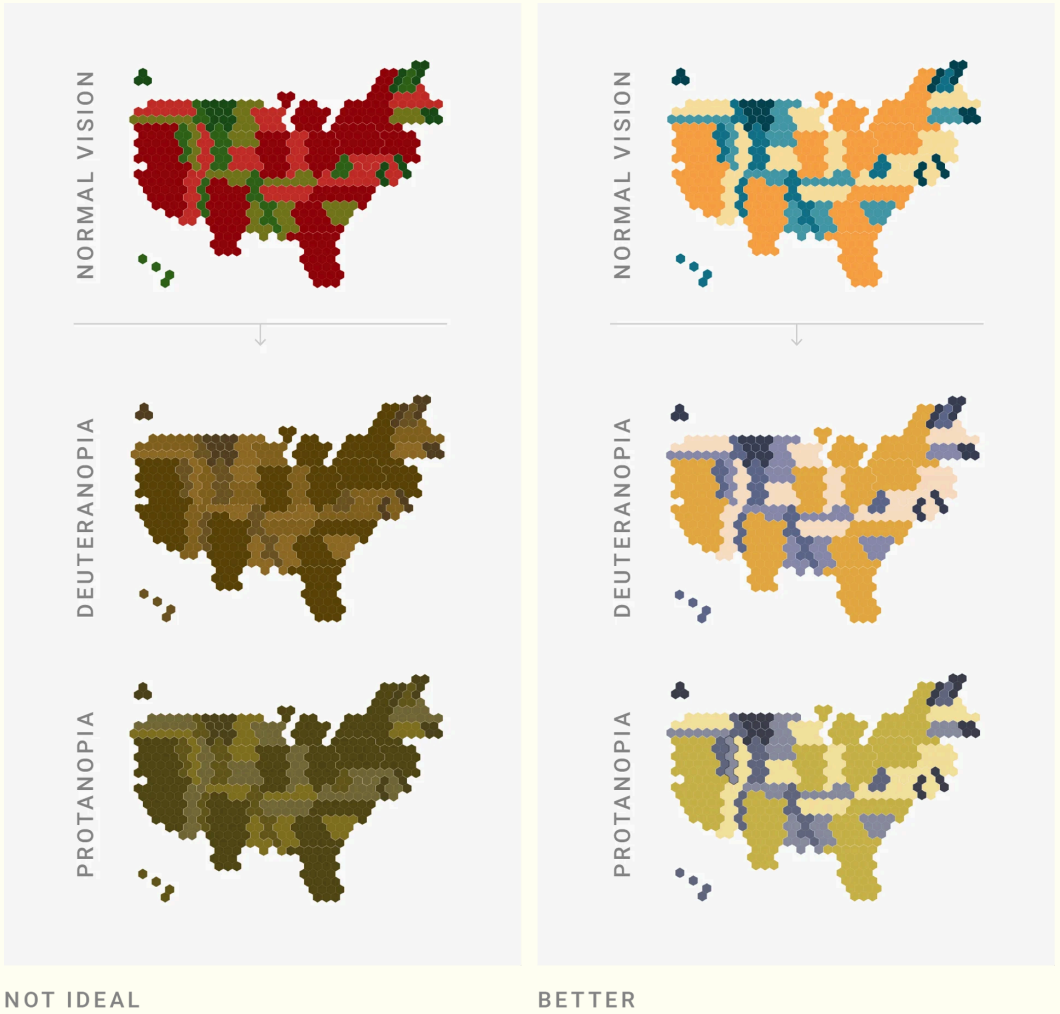
- Group categories together



Use intuitive colors



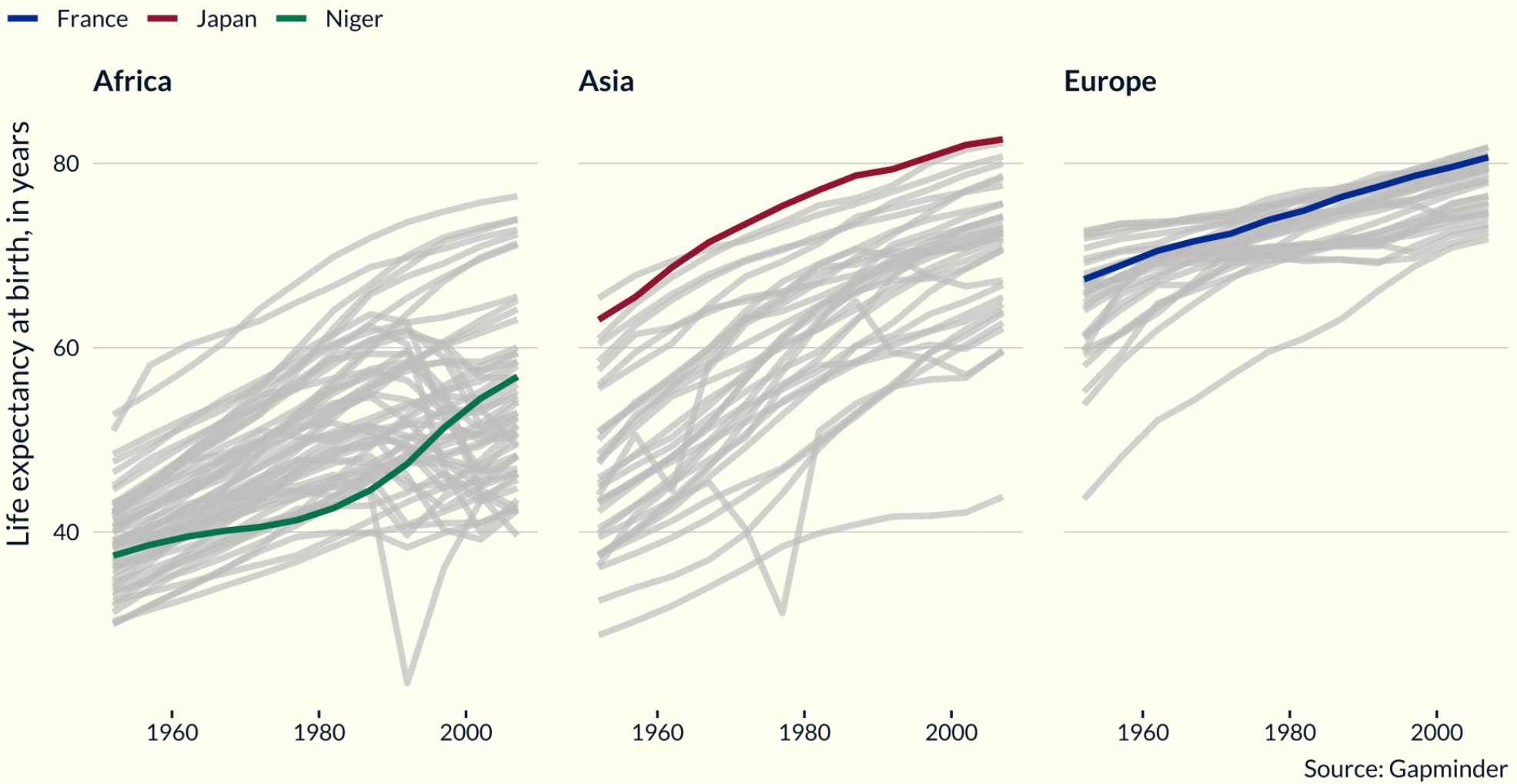
Colorblind-friendly visualizations



Use gray. Emphasize.

Life expectancy in France, Japan and Niger as compared to their own continent

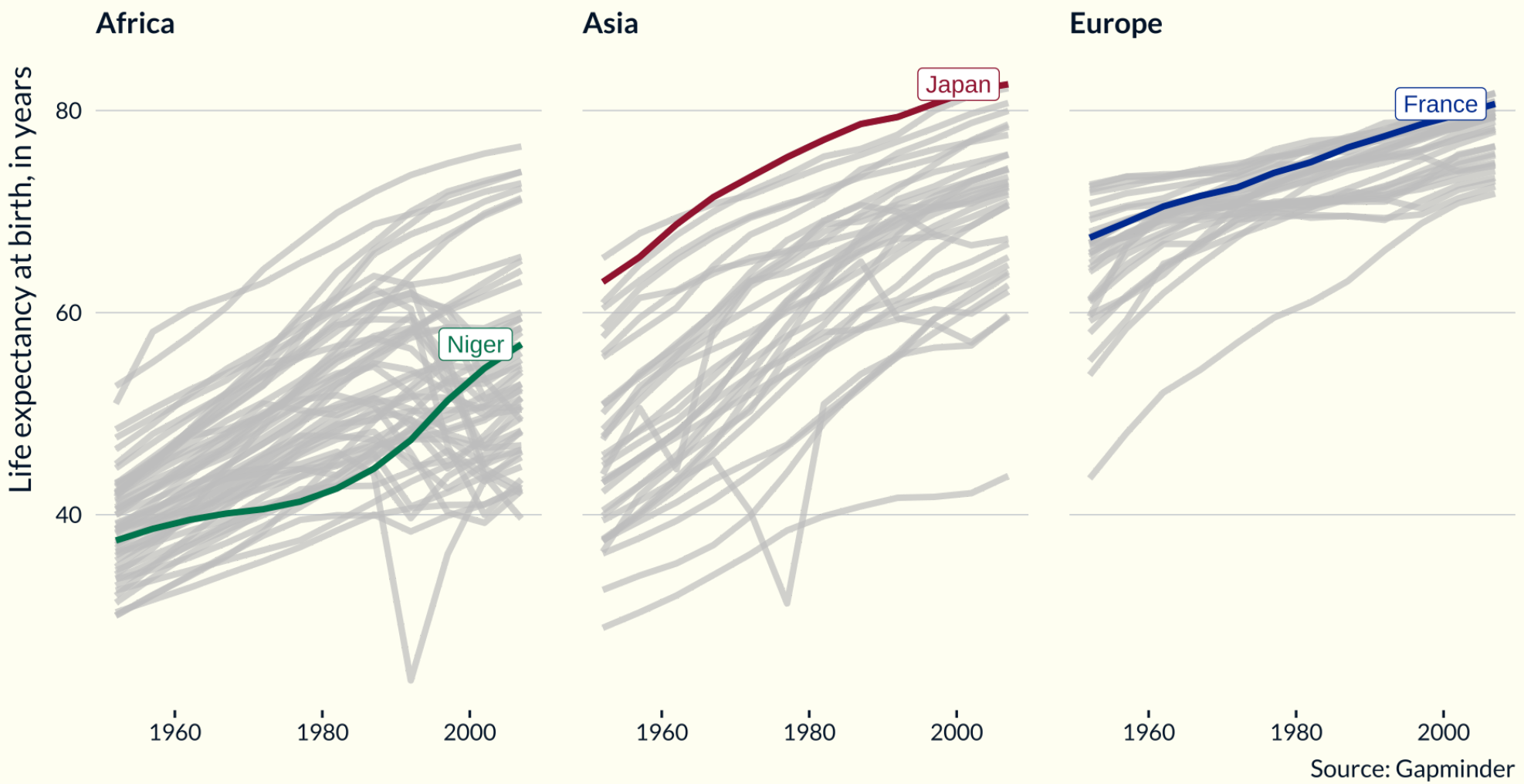
Evolution of average life expectancy at birth, by continent



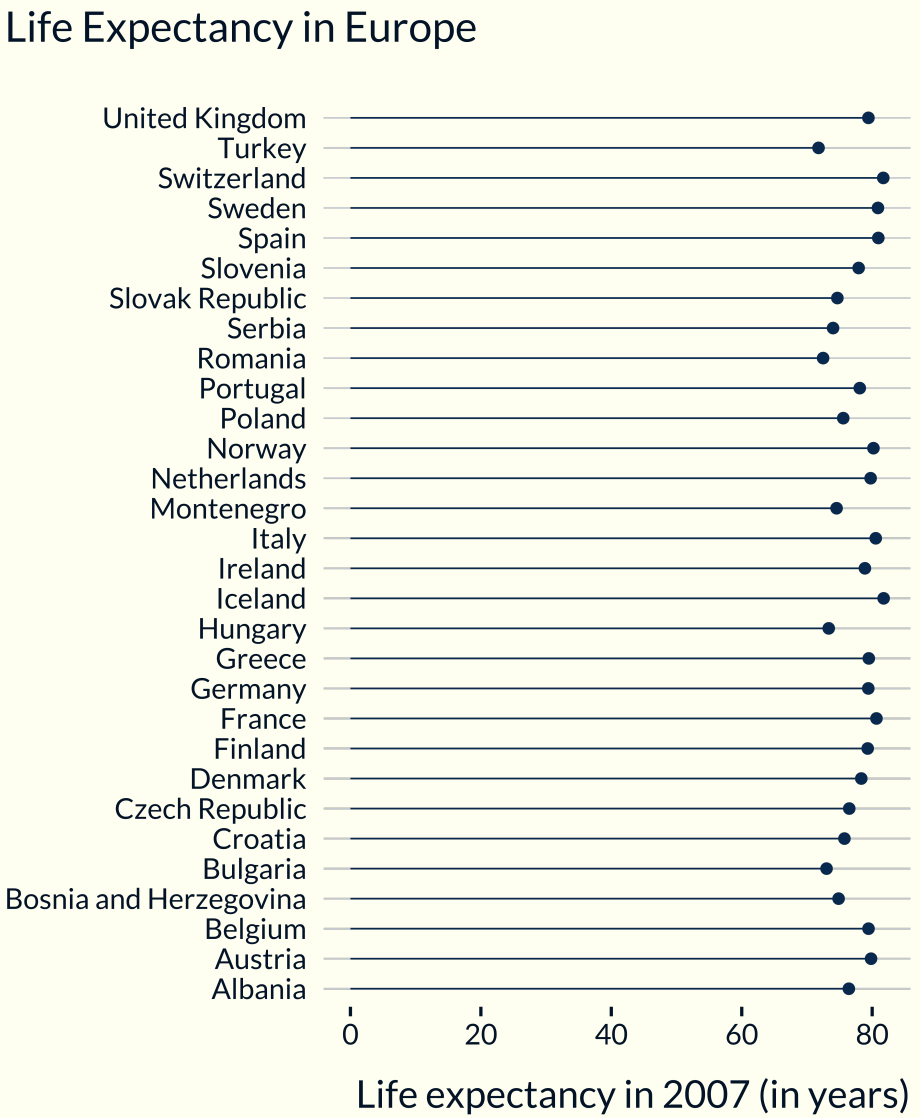
Label directly

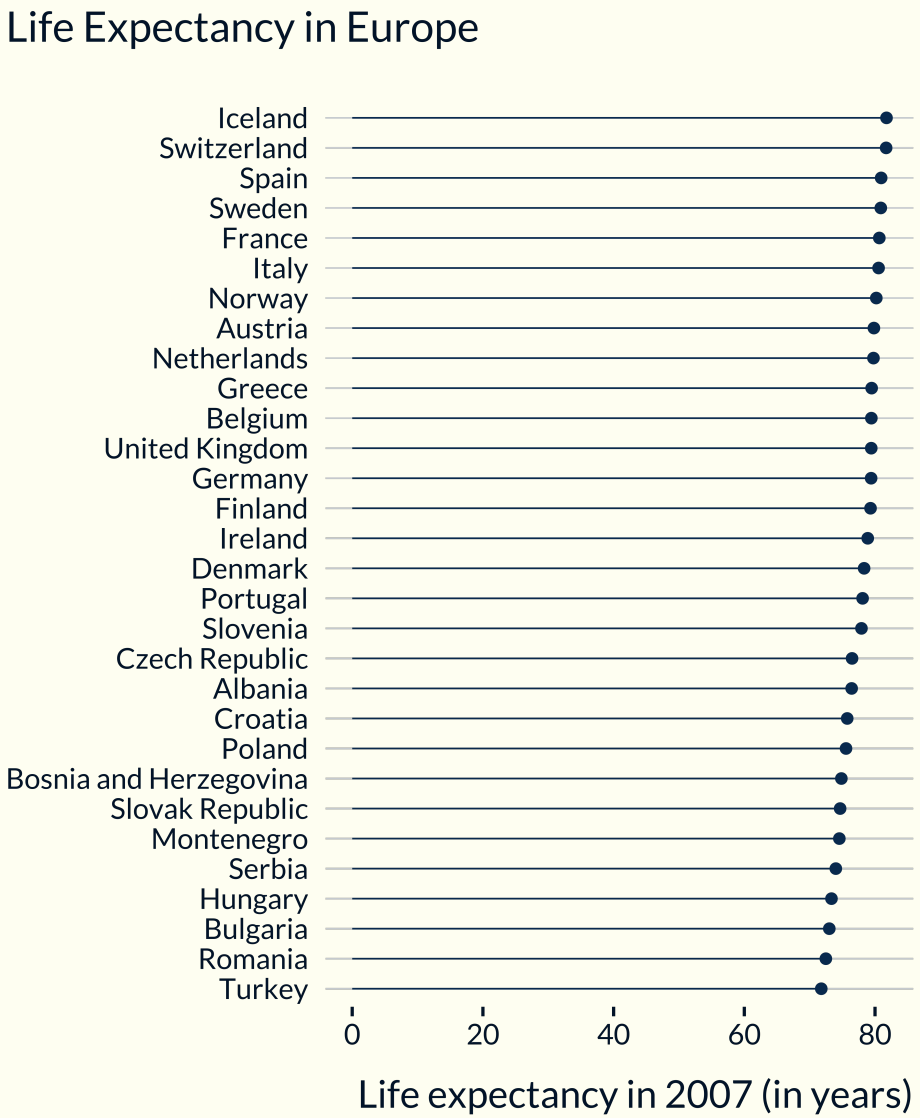
Life expectancy in France, Japan and Niger as compared to their own continent

Evolution of average life expectancy at birth, by continent



Order your data



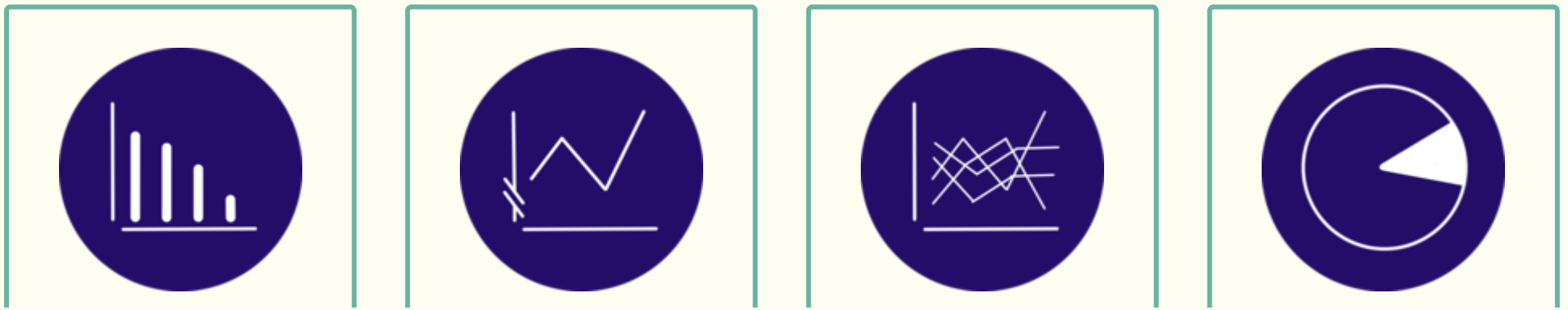


Data viz caveats

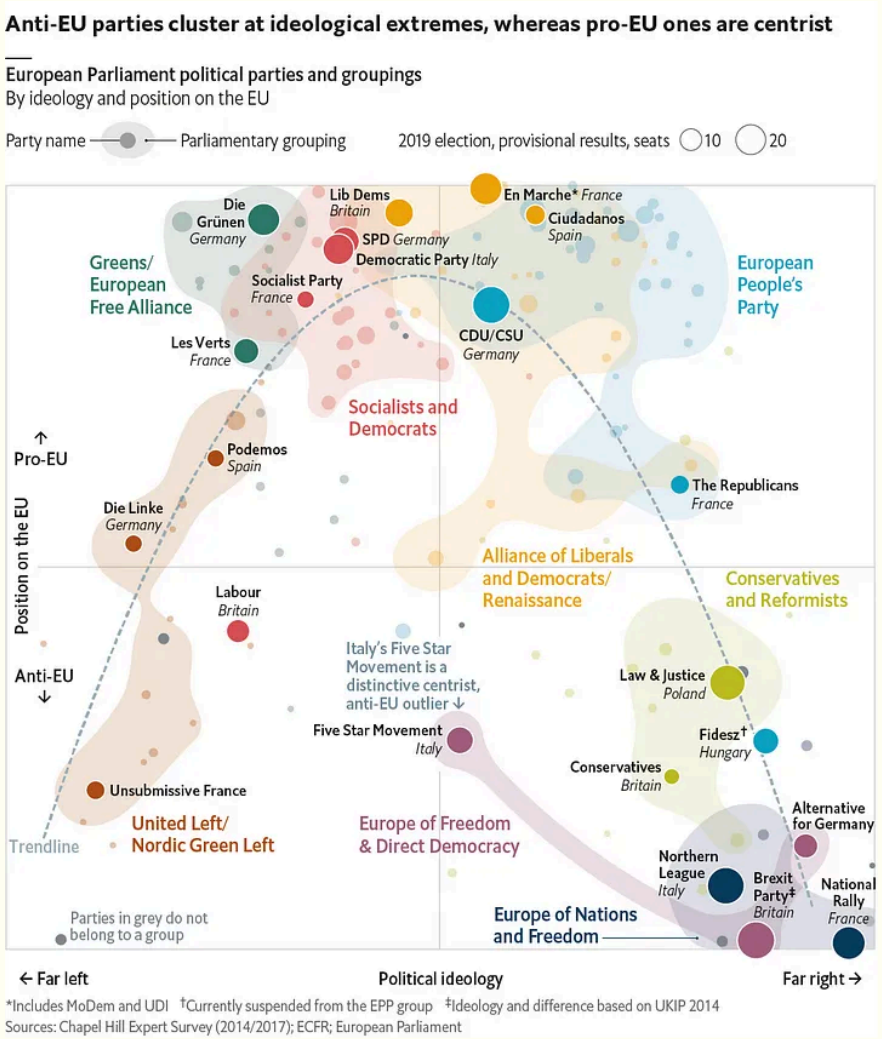
CAVEATS

A collection of dataviz caveats by data-to-viz.com

Show all | Top 10 | Improvement | Misleading | Map | Bar



Sometimes you may break the rules

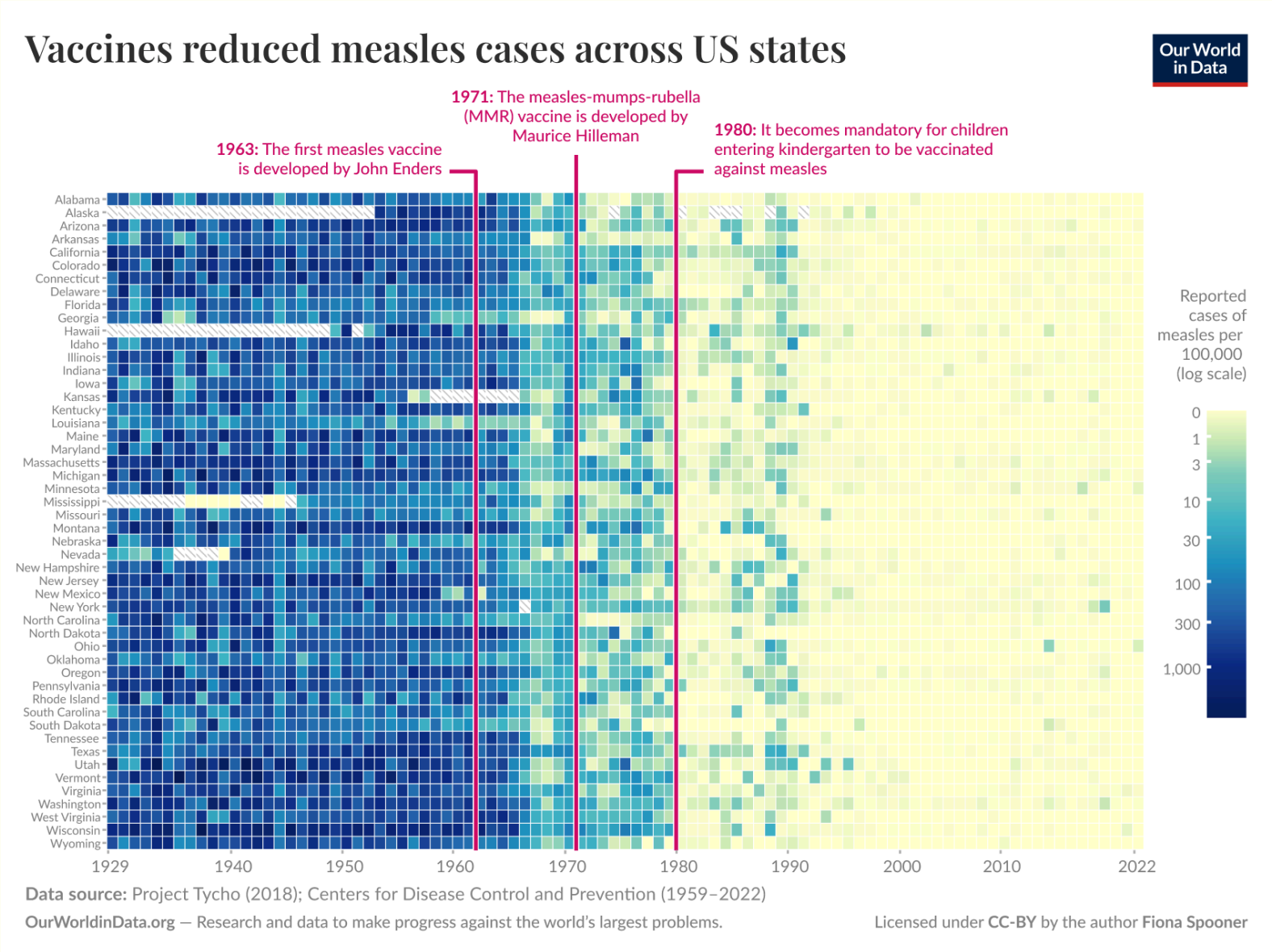


9 colors

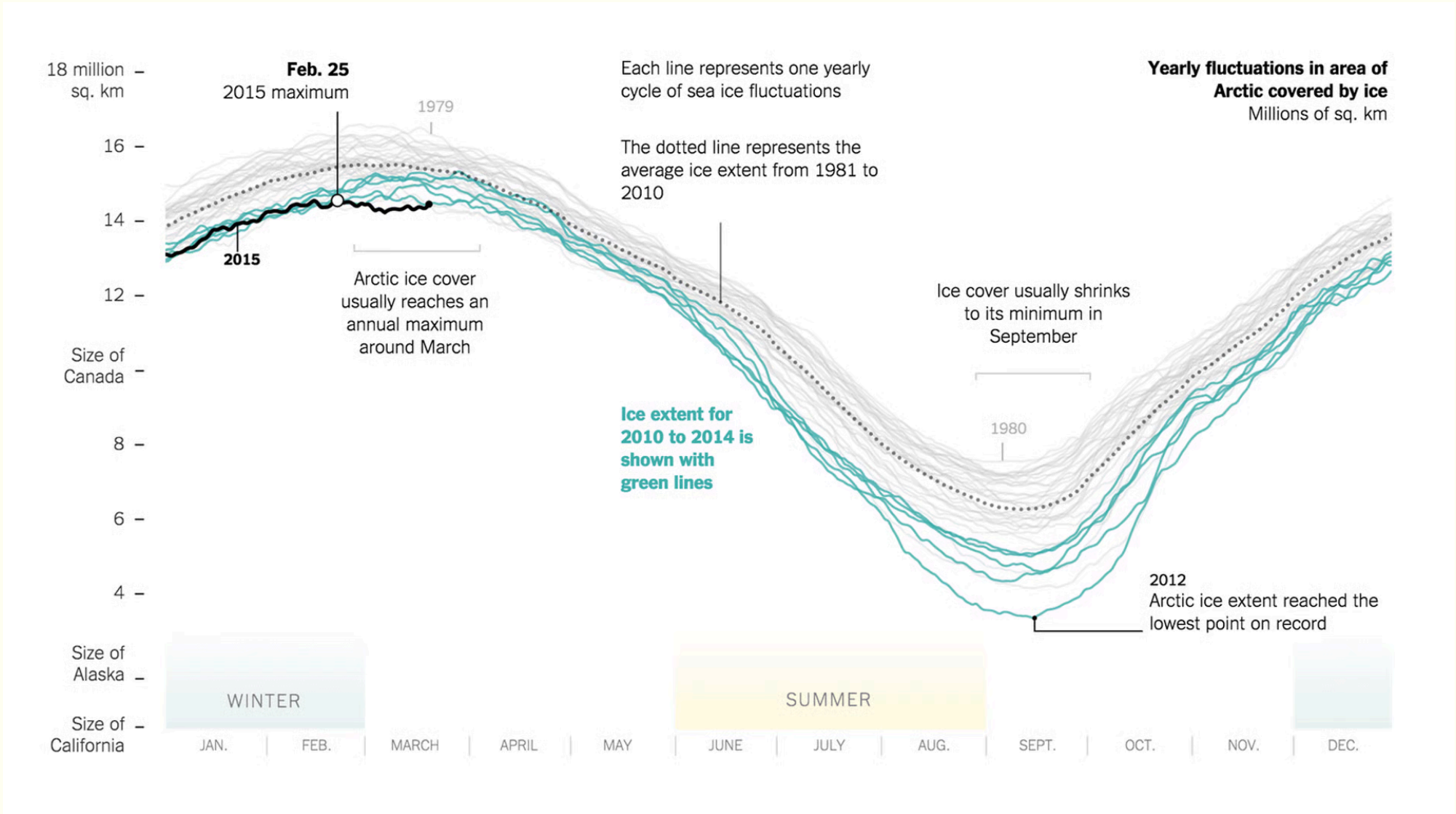
BUT

- Labeled directly
- Shade reinforces grouping

Nice data viz



Nice data viz



Pay attention to data viz

- If you start paying attention to data viz when you see them, you will see what works, what does not
- It is a process that takes place in the “background”: you will learn quickly and not realize it
- You will see nice looking graphs (and hopefully enjoy it)
- You will build better and more impactful graphs i

Summary of concrete recommendations

Take-away messages

- Build legible, understandable and nice looking graphs.
- Have a title and explicit axes; present them.
- Limit the number of colors you use. Use gray.
- Label your graphs directly, add annotations.
- Think twice before cutting the y-axis
- Overall, facilitate the retrieval of information.

Data viz in economics

When do we use graphs in econ?

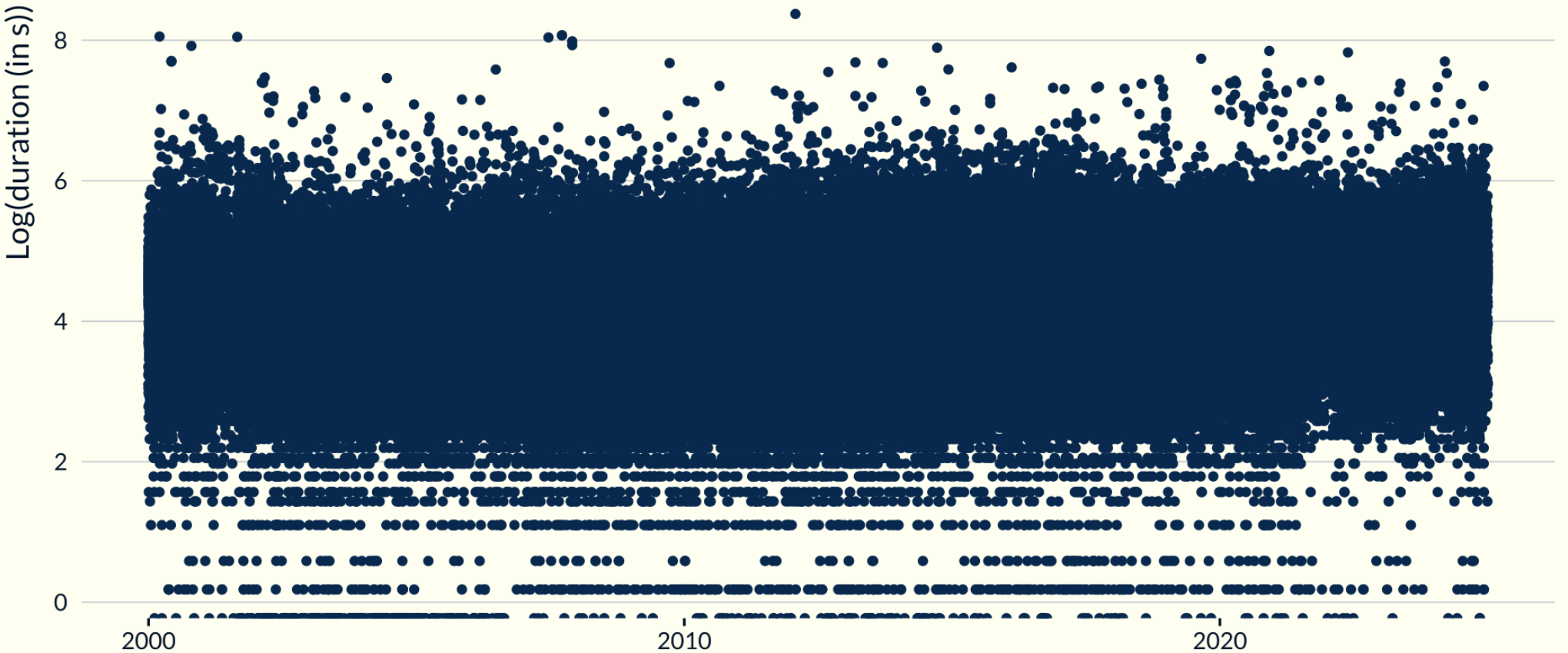
- As **rhetorical** visualization tools for models
- To **explore** our data
- To check the validity of our **models**
- As **diagnostics**
- To **communicate** results

Specificities of viz in economics

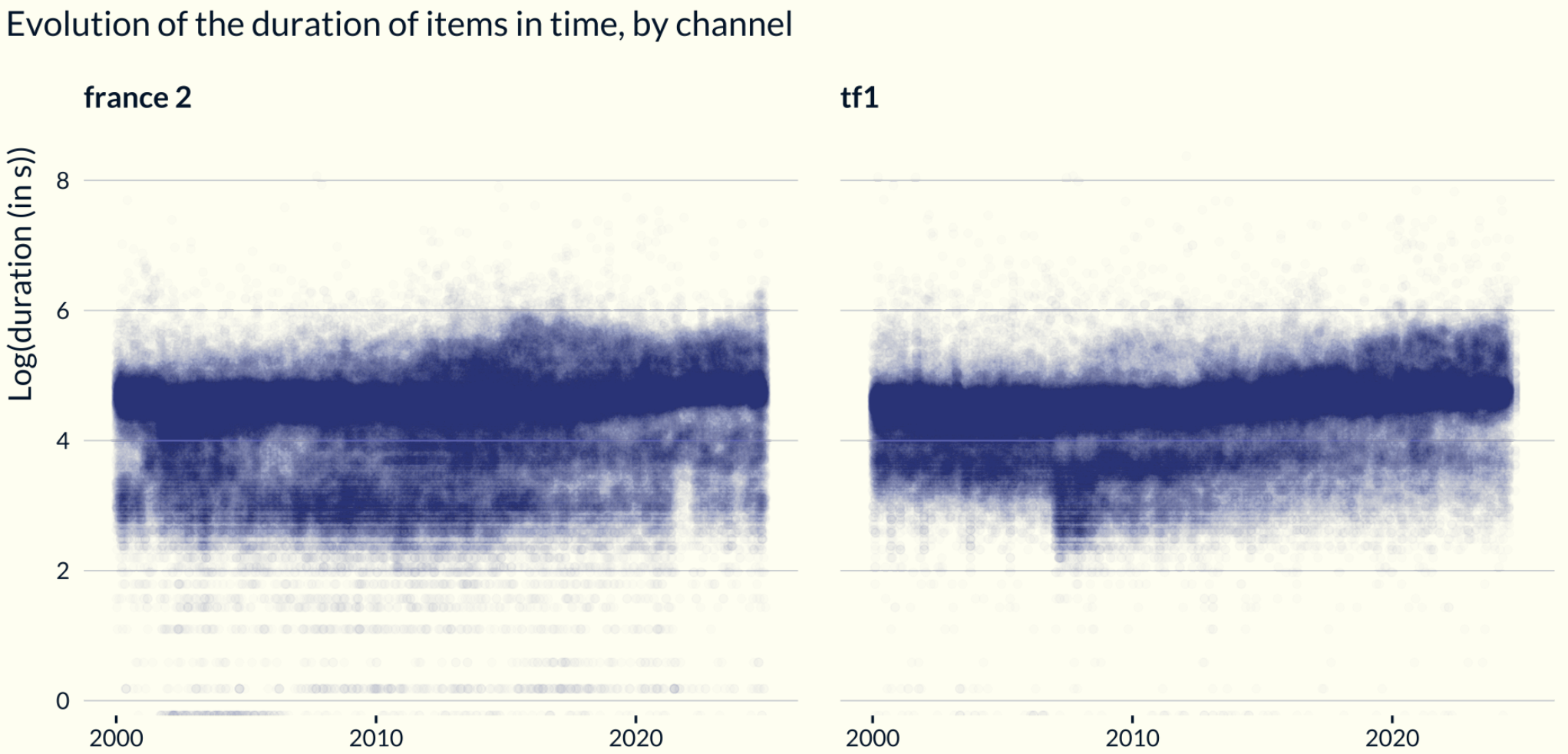
- We often have to deal with a **massive number of observations**
- We often want to display a specific type of graph: **estimation output**
- Our analysis are often based on **identifying assumptions** that we can sometimes check through graphs
- We often use very complex models. Visualization can help **understand what we are actually estimating**

Large numbers of observations

Evolution of the duration of items in time



Opacity and subsamples



Use heatmaps instead of scatter plots

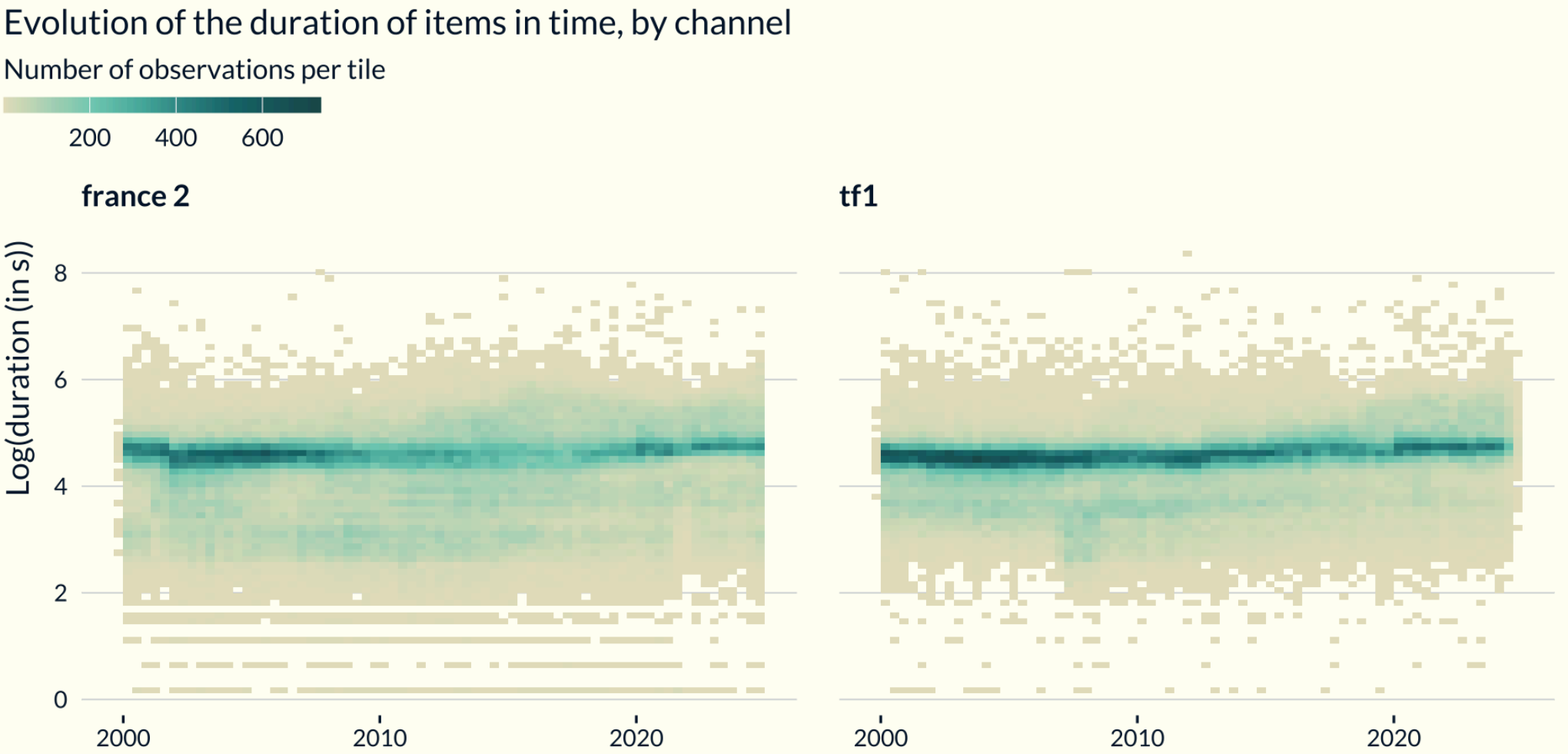
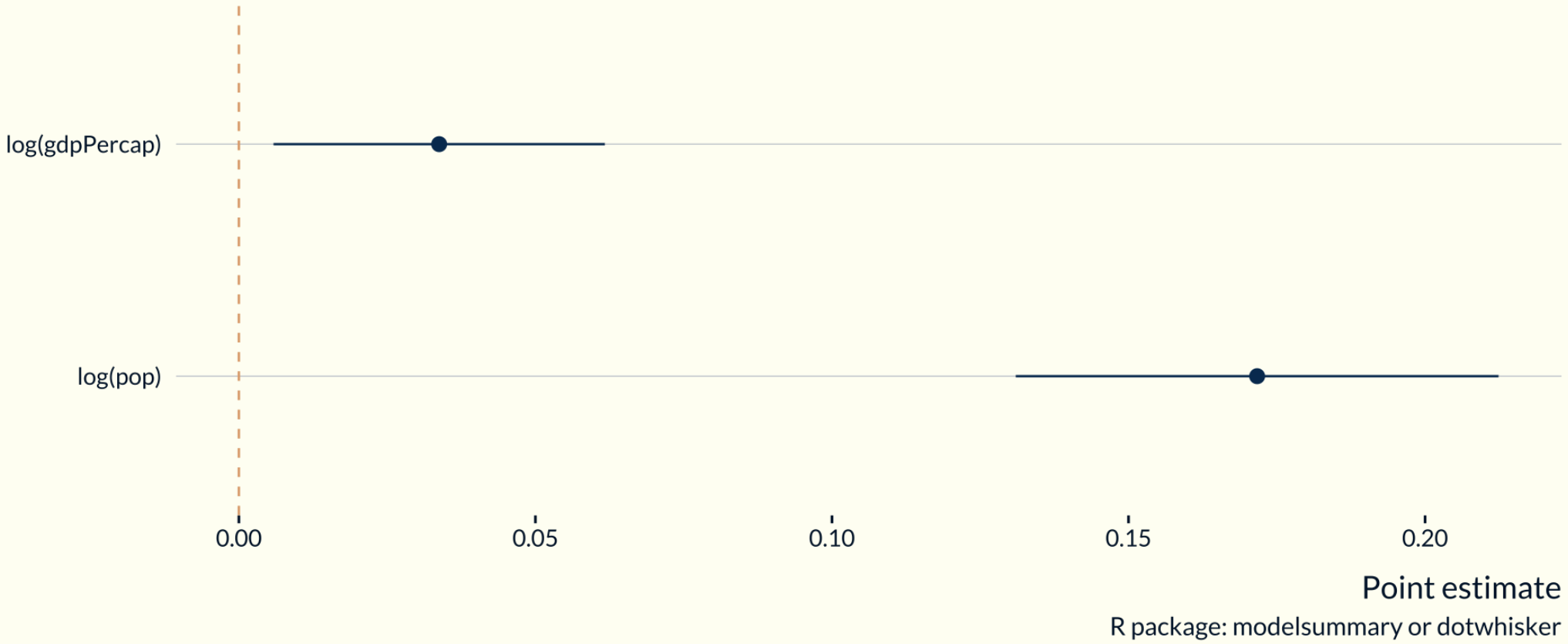


Table of estimation outputs

	(1)
log(pop)	0.172
	(0.021)
log(gdpPercap)	0.034
	(0.014)
Num.Obs.	1704
R2	0.935
R2 Adj.	0.929
RMSE	0.06
Std.Errors	by: country
FE: country	X
FE: year	X

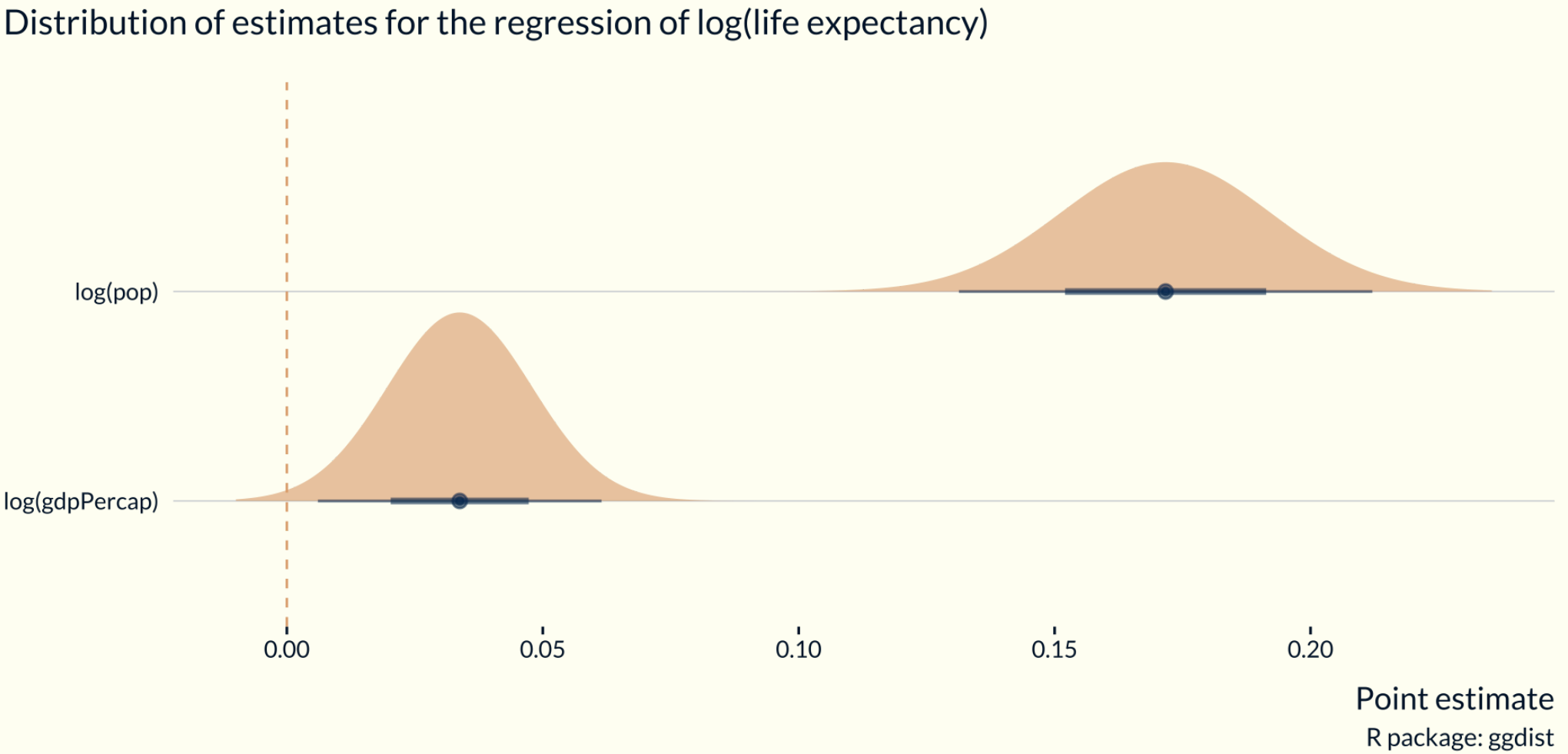
Whisker plot of estimation outputs

Point estimates for the regression of log(life expectancy)
With 95% confidence intervals



R package: modelsummary or dotwhisker

Whisker and distribution



Graphs hierarchy

- We make graphs for different audiences:
  - For **you** (and your future self): can be quite rough on the edges, but you will want to be able to understand it in the future
  - For **presentations**: you have some leeway for explaining orally your graphs
  - For the **paper**: there is only a couple of graphs in a paper; make them perfect

Main take-away points

Data viz at large

- Data viz is **powerful**, harness its power
- It can be super **insightful** or equally **deceptive**
- It can make your point **memorable**
- It can also be truly **beautiful**
- Leverage perception and data viz **principles**

How to build a graph?

Take-away messages

- There are **many rules** in data viz
- The main goal is to **facilitate the transmission of your message**
- What is the main point you want to convey?
- Choose (one of) the right graph types
- Explain your graphs, orally and by facilitating reading, on your graph
- BUT avoid clutter

Data viz for economics

- Most data viz principles and ideas also apply to economics and academia in general
- There are however some specificities
- In particular, some graphs and **types of analyses** are specific to academic research
- Data viz can be extremely useful for research and communication

Thanks