

Lecture 5 - Communicating

Topics in Econometrics - M2 ENS Lyon

Vincent Bagilet

2026-09-30

Graphs and tables, not so simple

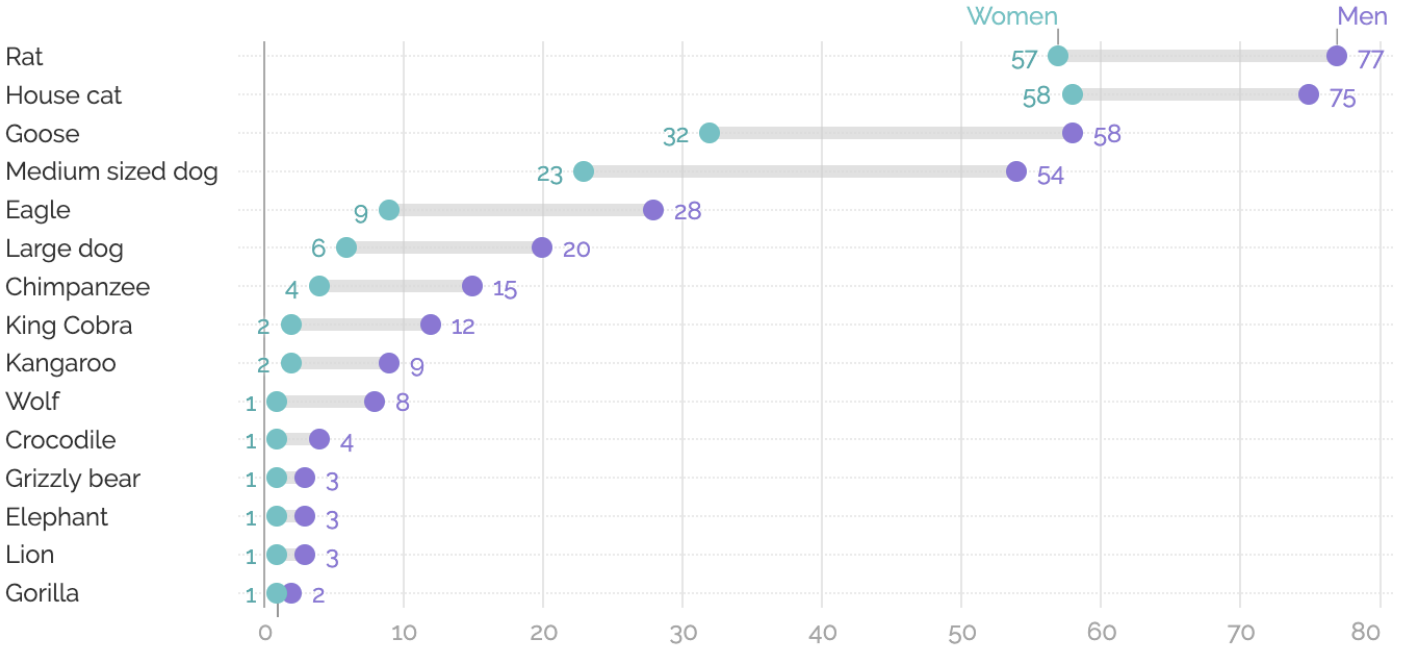
- Key to produce credible estimates but only worth it if we **communicate** them properly
- **Graphs** and **tables** are **everywhere** in economics
- They are often the result of analyses
- Readers often start looking at graphs and tables
- **Seems pretty simple**, we all know how to make graphs and tables
- Sure BUT handled carelessly a graph can be **inefficient** and even **misleading**
- Once you pay attention to data viz, it is fun, instructive and satisfying!

Many graph types

- Type 1
- Type 2
- Type 3
- Type 4

What animal could Britons beat in a fight? Compared to women, men feel most able to take on medium-sized dogs and geese

Which of the following animals, if any, do you think you could beat in a fight if you were unarmed?
% of 2,082 GB adults



Trying to go beyond the obvious

- I will probably state **obvious** stuff BUT
 - Maybe still good to hear it sometimes
 - I will provide **references** to explore all this in more details
 - I will try to structure and **conceptualize**
- Objectives of the class:
 - Provide **tools** to nail your data viz and your tables
 - Communicate that **how you present results matters** and provide a conceptual lens

Outline

- **Why** good data viz matters?
- **What** are the characteristics of good data viz?
- **How** to make an ok data viz?
- **Communicating** in economics
- Main **take-away** points

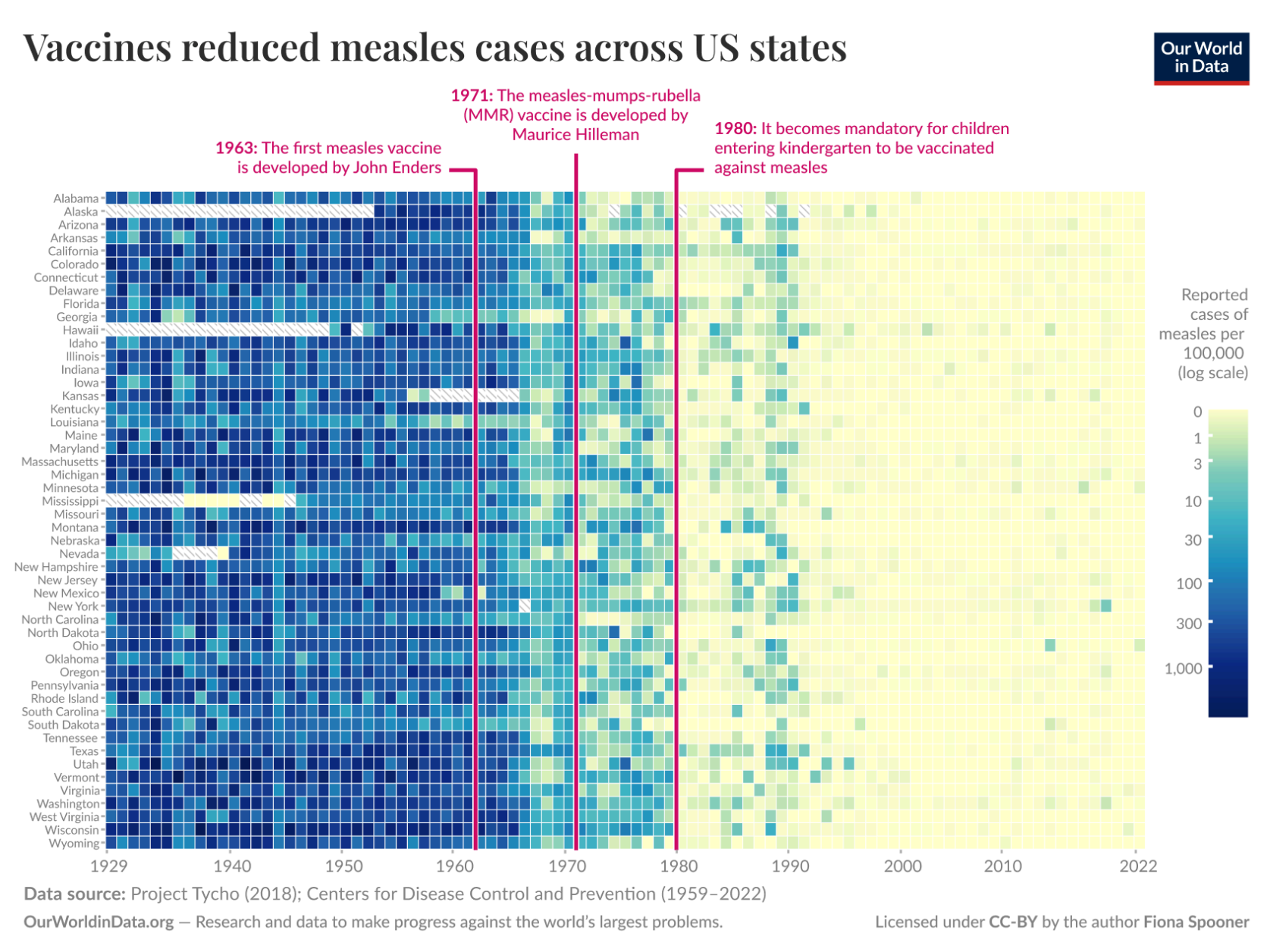
Why *good* data viz matters?

What do we use data viz for?

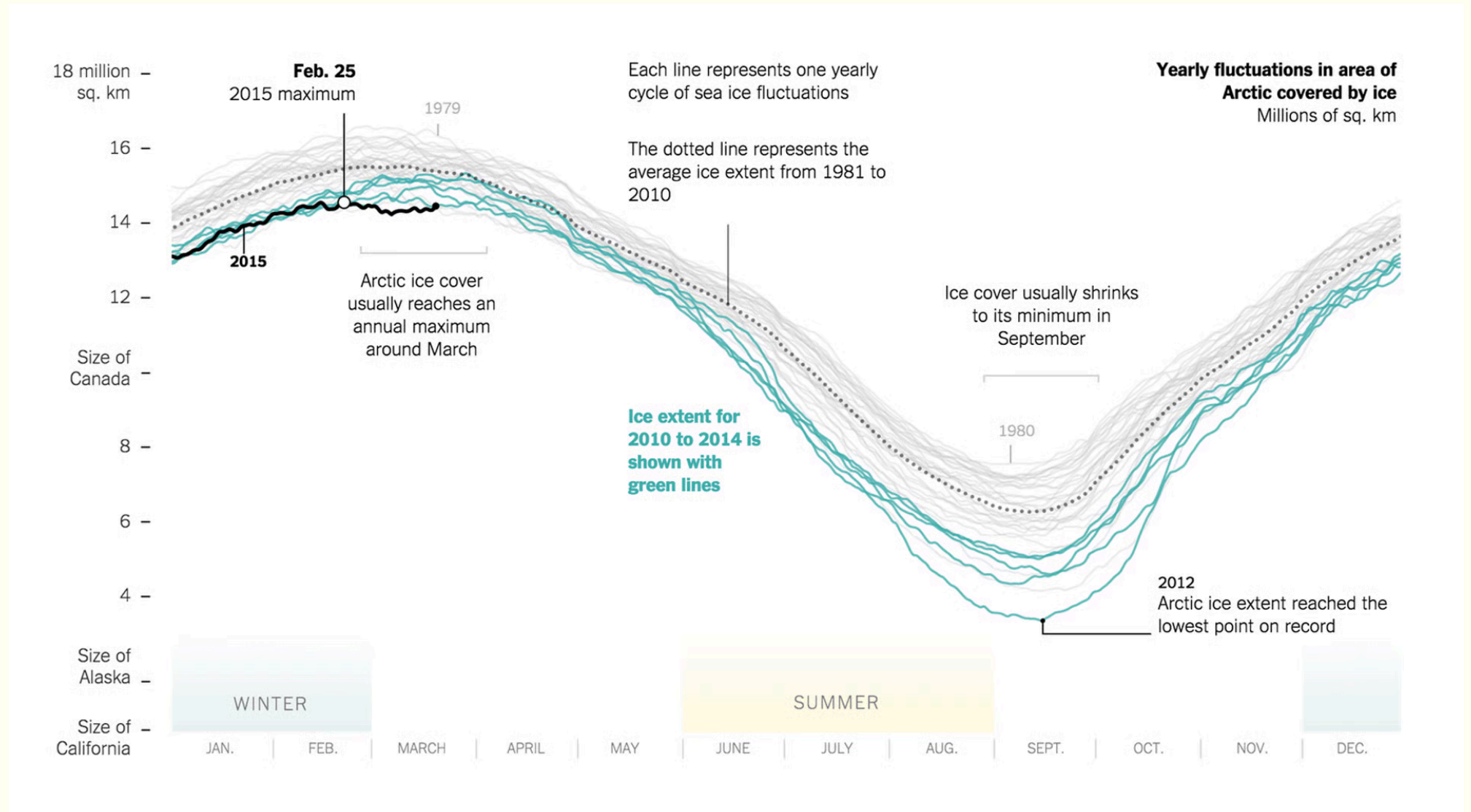
Usefulness of data viz, overall

- Allows to **synthesize** large amounts of information into a limited space (that also applies to tables)
- Ease **interpretation** (reduce the cognitive load)
- For your audience to not only receive info but also **understand** and **remember** it

Synthetic and memorable



Yearly fluctuations in area of Arctic covered by ice



Several uses of data viz

Graphs to explore

- Analyze
- Confirm

Graphs to explain

- Inform
- Convince

Graphs for inference

- Econ specific

They have different goals and audiences

Adapting graphs to the audience

- Which potential audiences?
 - **Peers, economists**: can use complex but familiar types of graphs
 - **Non-economists**: think about how they are
 - **You** today: super quick but with basic labels
 - **Future you**: still need to be able to remember what the graph is about 6 months later

Different levels of polishing

- Graphs 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!

Explore: make sense of your data

- Data viz helps a lot to **identify patterns** in data
- That's the role of the exploratory data analysis (**EDA**)
- It may help to:
 - Understand what is your data and in your data
 - Spot issues in your data
 - Formulate hypotheses (to test on other data sets)
 - Assess the relevance of some key assumptions



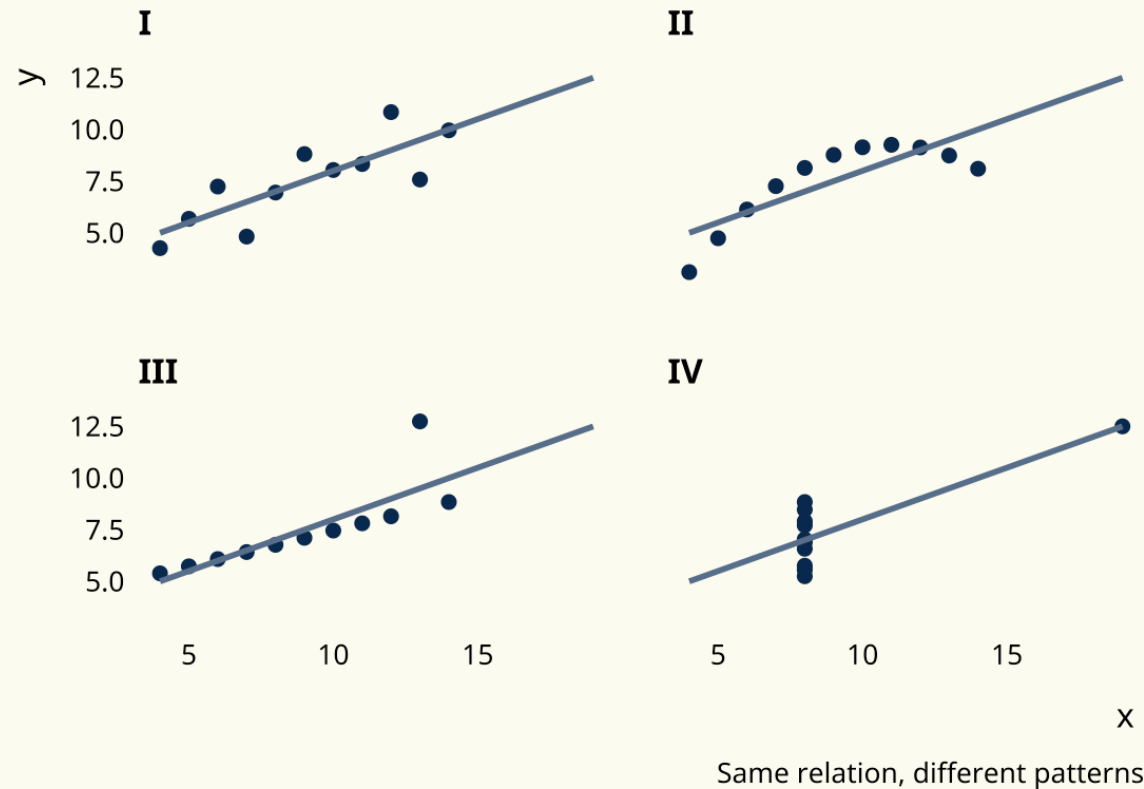
Rule

Always look at your raw data

Look at your raw data

Same regression line, different data

Anscombe's quartet

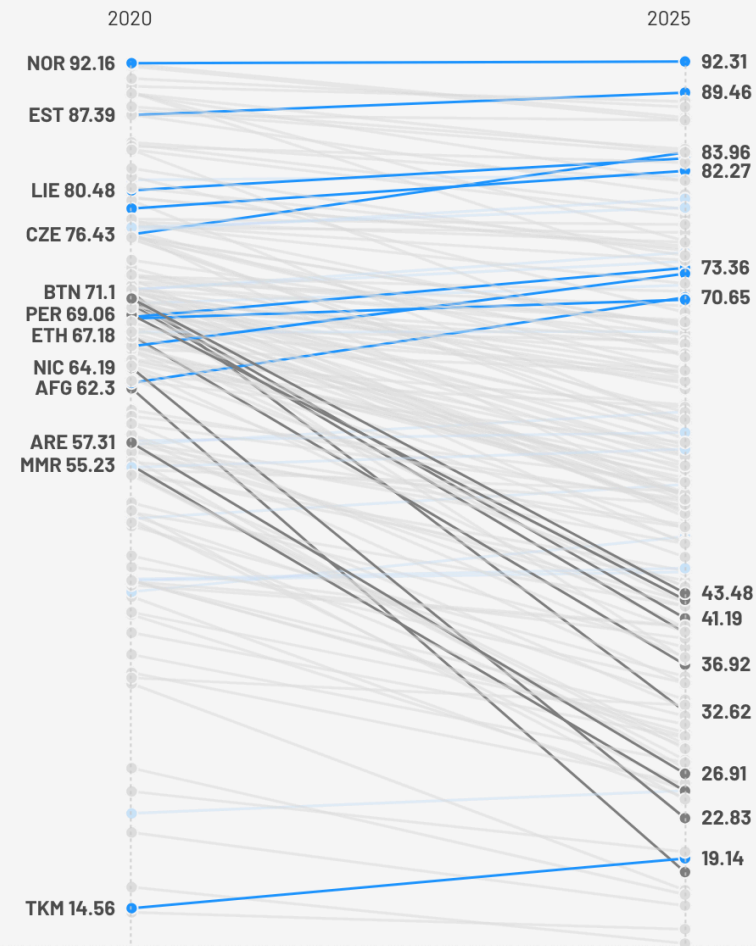


- eg might be helpful to **evaluate your modeling assumptions**
- Spot outliers driving your results

Explain: get your point across

Press freedom freefalls

Average scored plummeted 10.5 points globally. This chart highlights the 10 countries with the sharpest **increase** or **decrease**

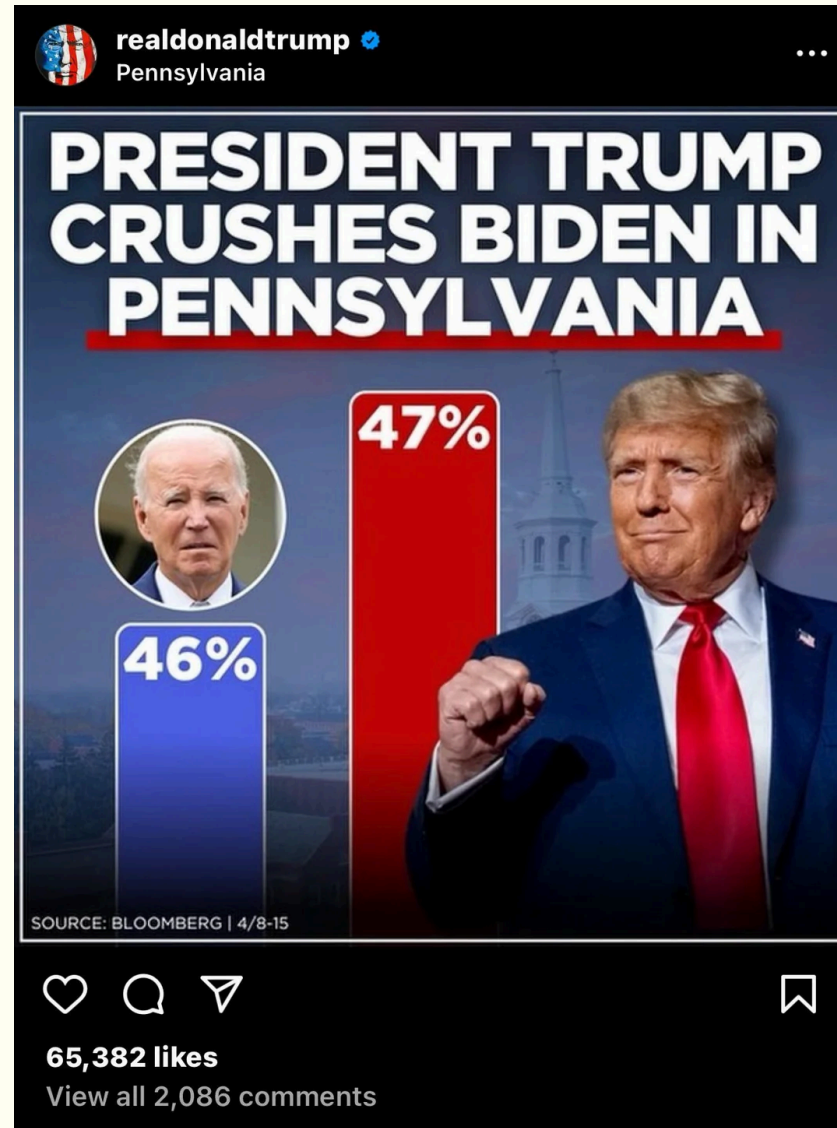


Source: [Reporters Without Borders](#) • Made with Flourish

- A data viz can be much clearer than a table (not always)
- Can help focus on one specific point
- Convey a **story**

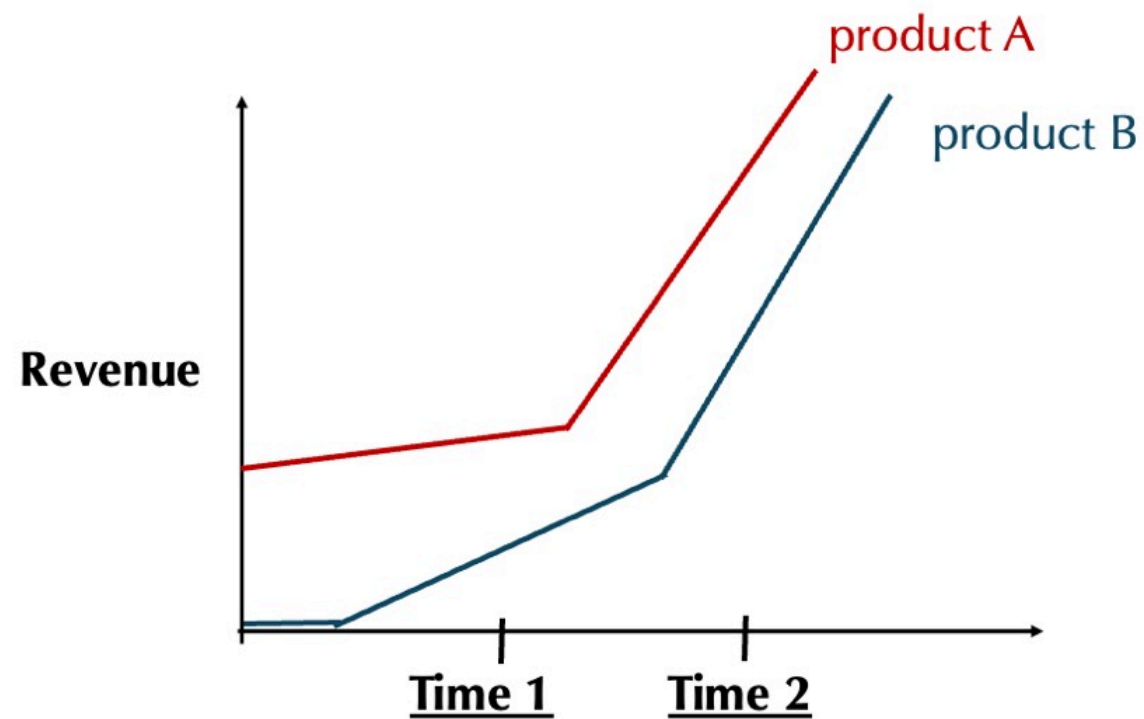
Why make *good* data viz?

Data viz can be obviously deceptive



...or difficult to interpret

On which day was there a bigger difference in revenue between A and B?



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

- Charts can be **wrong**. They can also be correct BUT **misleading**¹
- Be mindful of this power when make and read graphs

Cutting 0 on the y-axis

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

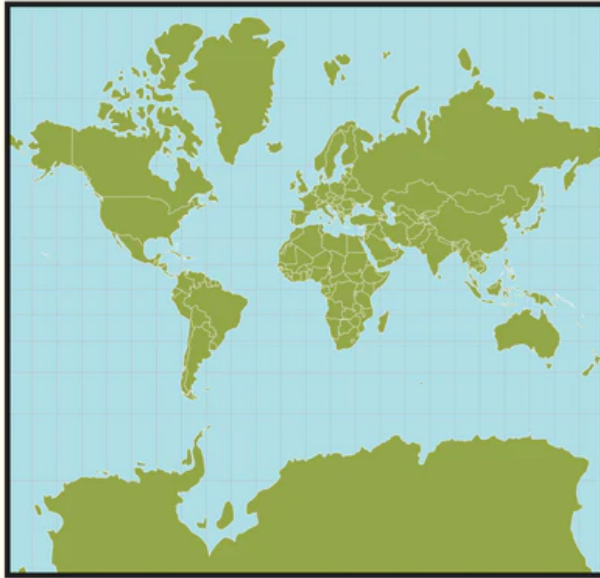
Vox



Watch on

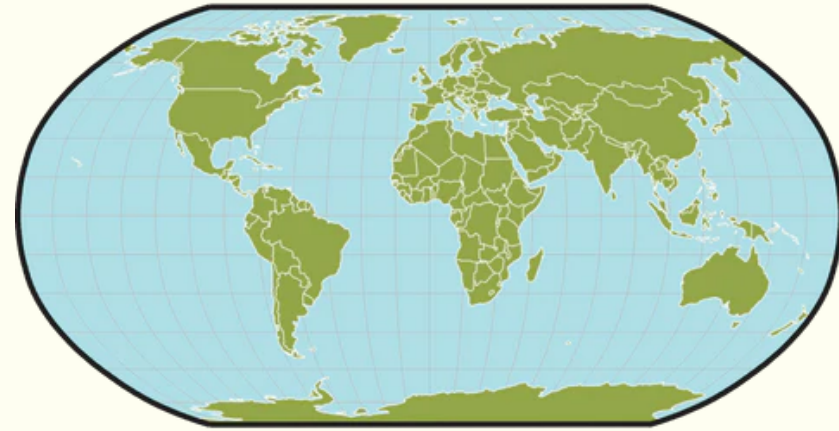
Map projections distort the reality

MERCATOR



 Gerardus Mercator - 1569

ROBINSON



 Arthur H. Robinson - 1963

Data viz and credibility

“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

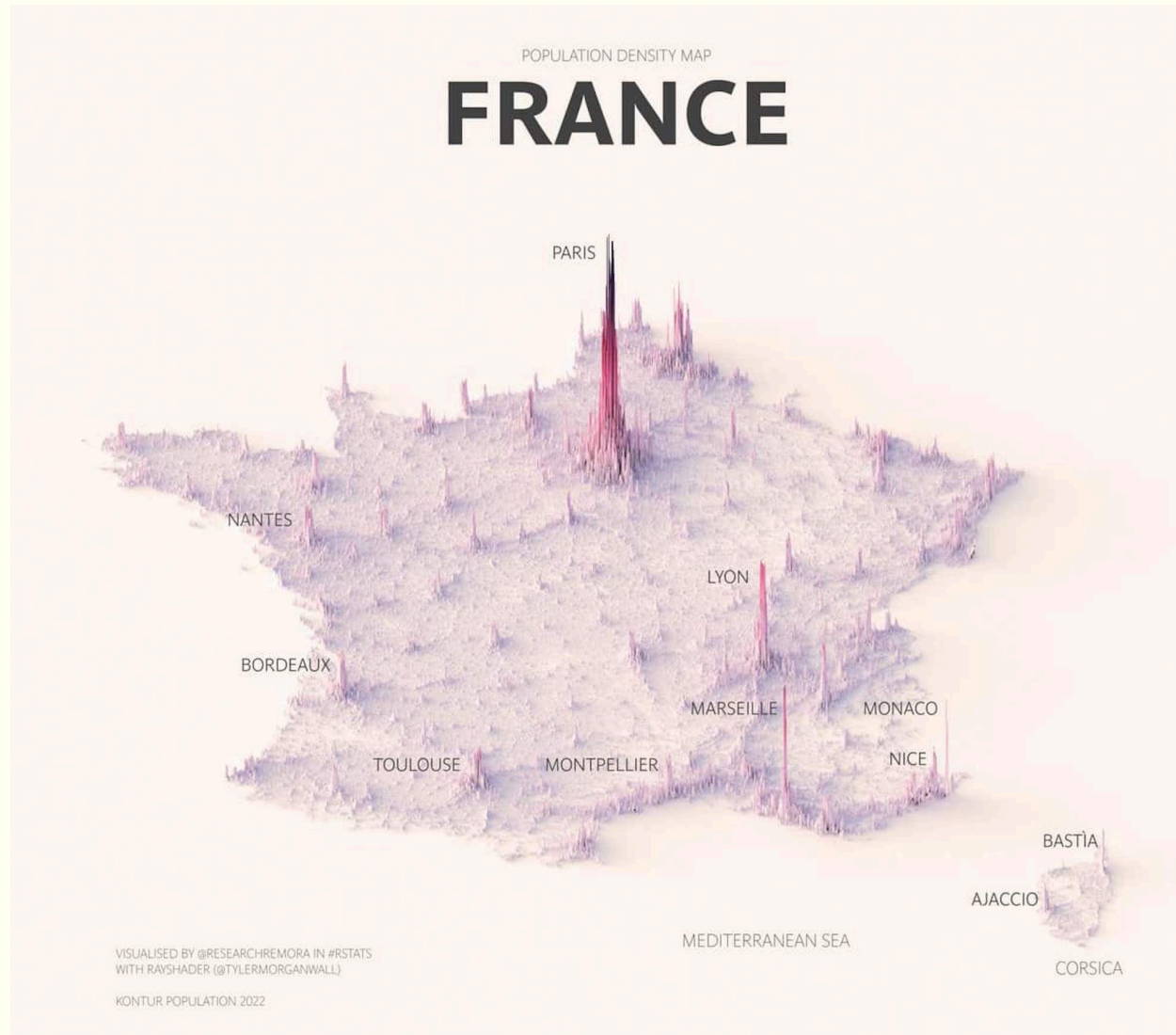
Proximate Cause 2007–08
Assistant to the director
• Helped devise fundraising campaigns for this innovative nonprofit
• Handled lunch orders and general errands

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

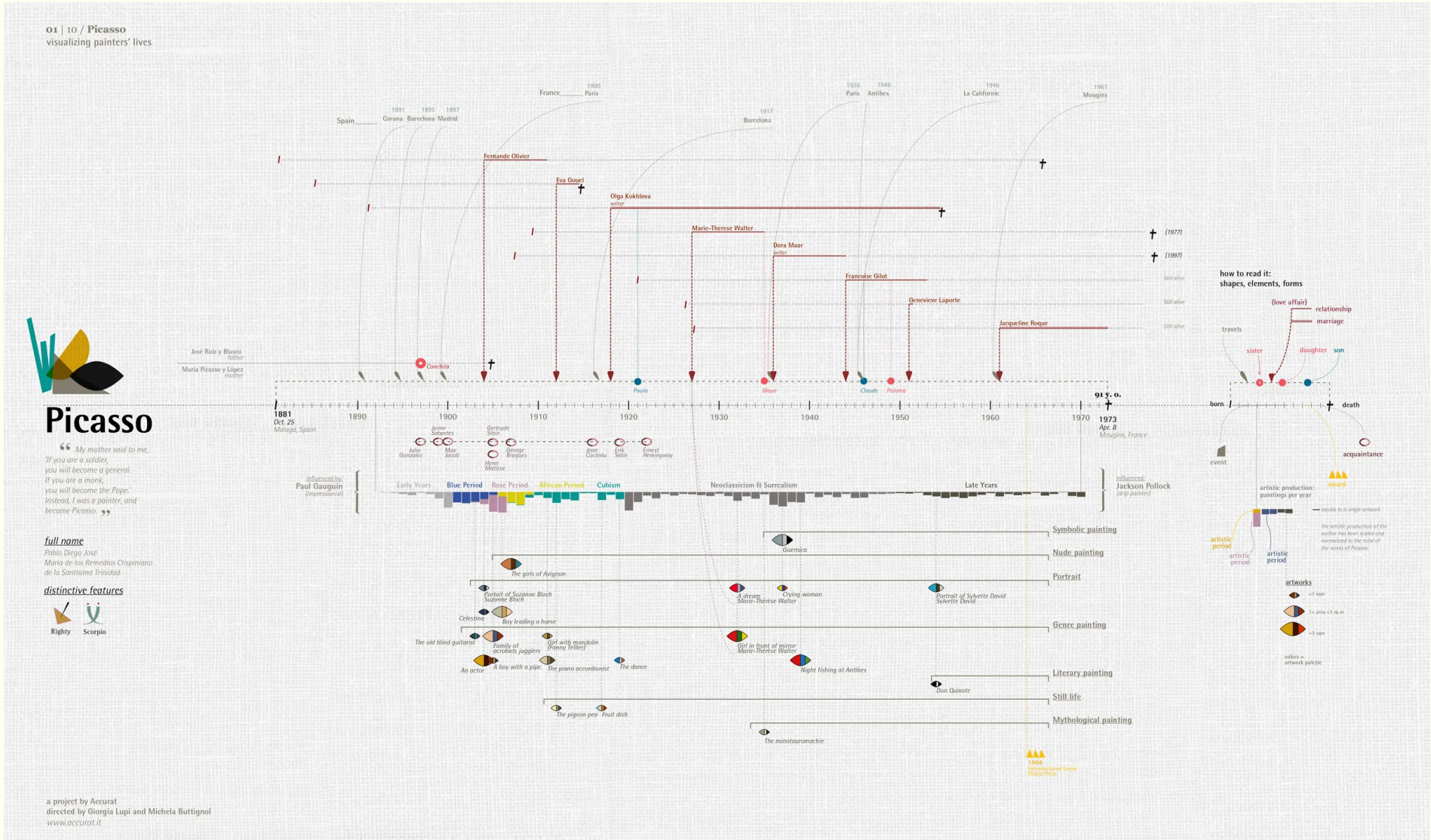
Graph aesthetics

- Beyond credibility, 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**
- They can help engage the audience
- Pretty does not always mean non-simple. Simple graphs have value.
- In academia, maybe better to keep the design rather minimalist

Pretty and memorable



Engaging the audience



What are the characteristics of good data viz?

Tell a story

Which story?

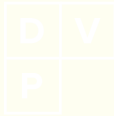
- Stories resonate with us, humans
- What do you want to show in your graph?
- What is the **message** you want to convey?
- Again, you do not want your audience to receive info, you want them to **understand** and **remember** your message

! Important

Clearly stating the message you want to convey is the first and most important step of a successful visualization

1 dataset, many stories, many graphs

- With the same data, you can:
 - Tell a lot of different stories
 - Emphasize different points
- Making a graph = **choosing a lighting** for your data
- The **type of graph** you choose matters and depends on your story



(<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.

Number of World Heritage Sites			
	Norway	Denmark	Sweden
2004	5	4	13

Cognitive load

A guiding principle

A rough definition

How easy or difficult it is to parse new information

- Often want to **decrease the cognitive load** (particularly for academia)
- Can often use this **a guide** for making our data visualization
- As all rules, it is meant to be broken

Different types of cognitive loads

- **Intrinsic** load: inherent complexity of the information
 - *eg*, $3/7$ more difficult to compute than $3 + 7$
- **Germane** load: how familiar is the audience with the new information
 - *eg*, box plots, coefficient plots might be difficult to understand without prior experience
- **Extraneous** load: how the new information is presented
 - *eg*, horizontal text, highlight important data points, annotate, etc

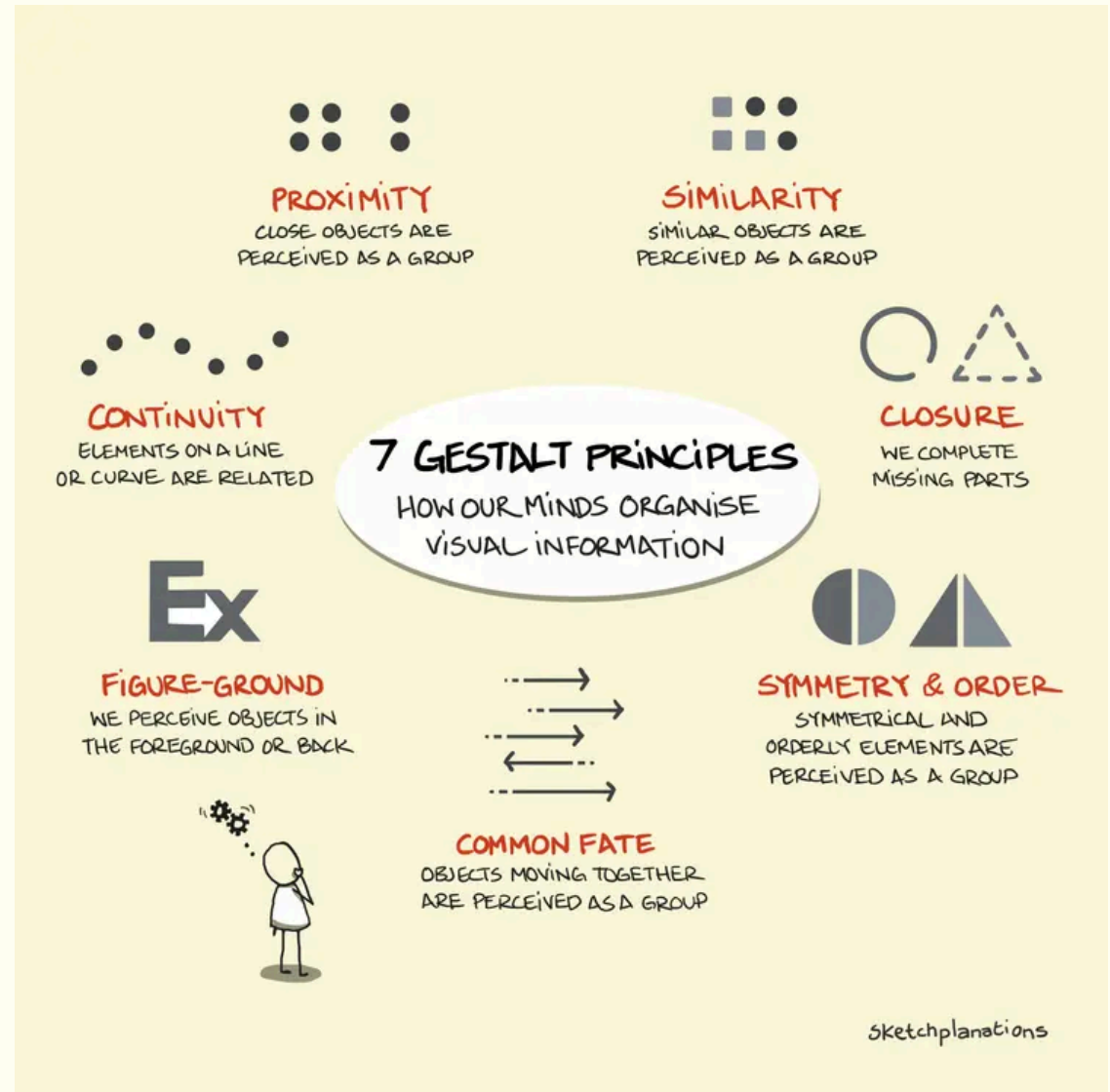
Data viz theory

Importance of theory

- There is actually a lot of theory behind data viz:
 - Perception,
 - Colors,
 - Design,
 - etc
- Can relate to cognitive load
- But can also **leverage it to make better data viz**
- Worth learning about it and being aware of key principles

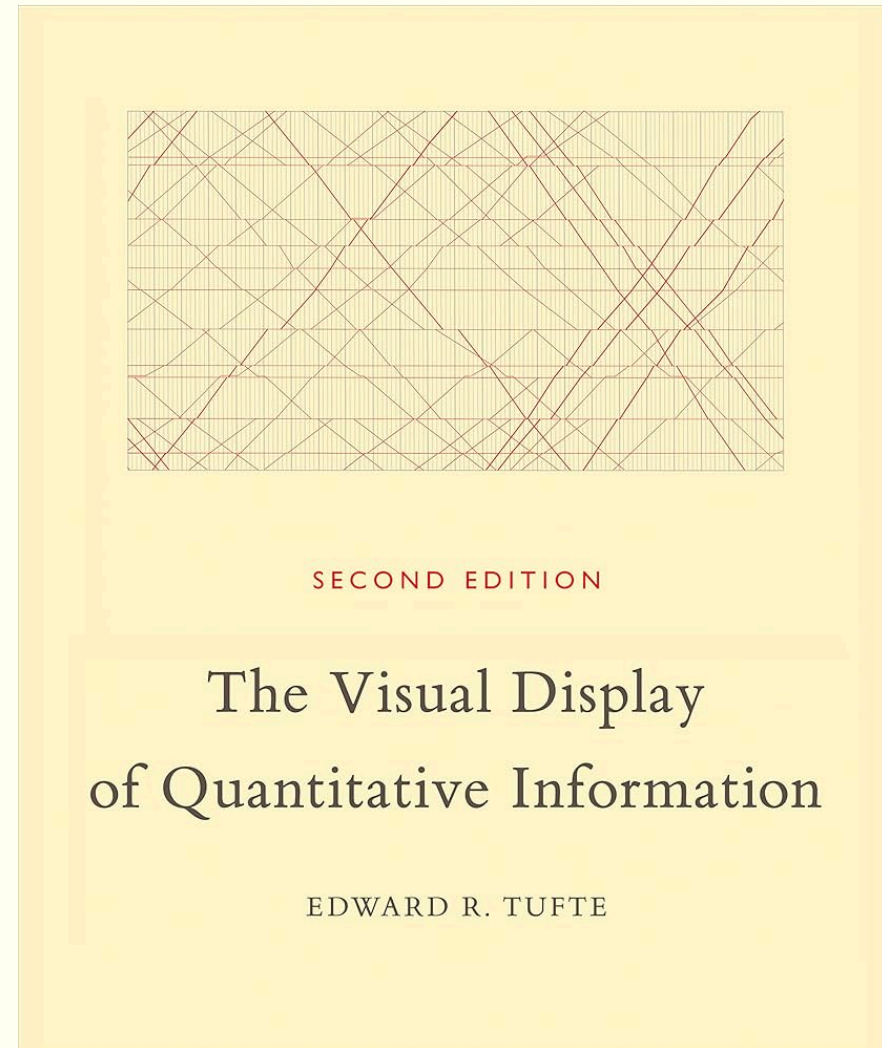
Gestalt principles

- 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



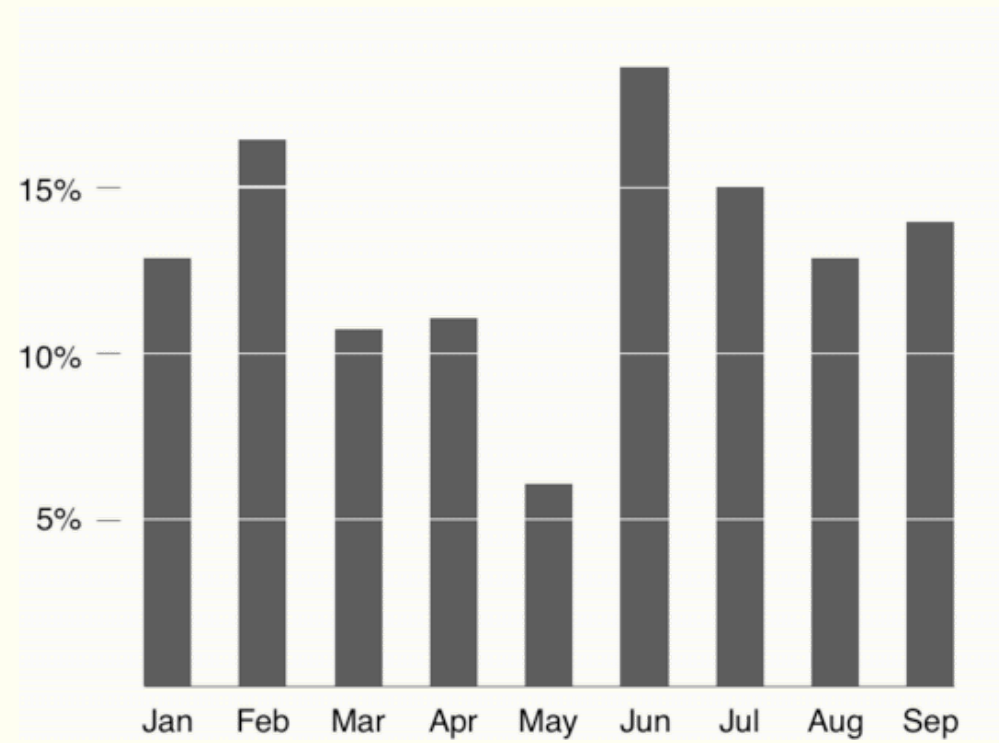
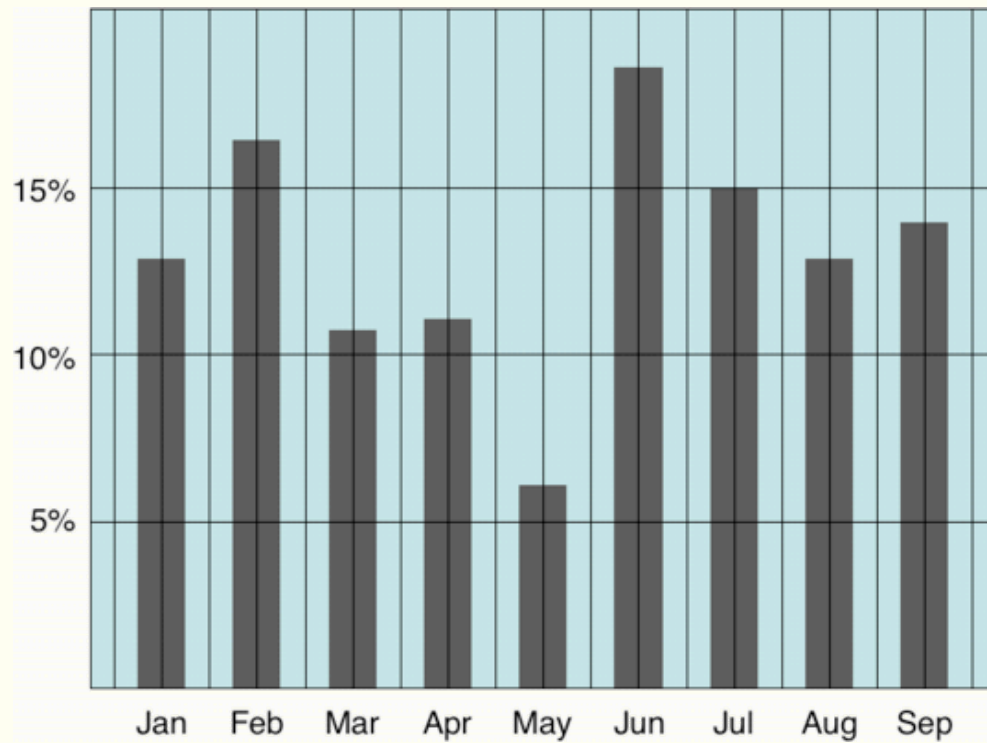
Data-to-ink ratio

- Introduced by Edward Tufte
- In a nutshell: **avoid clutter**
- Erase non-essential and redundant information
- *“Above all else show the data”*
- Somehow “reduce the cognitive load”



Data-to-ink ratio

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



How to make an ok data viz?

Non-negotiable rules

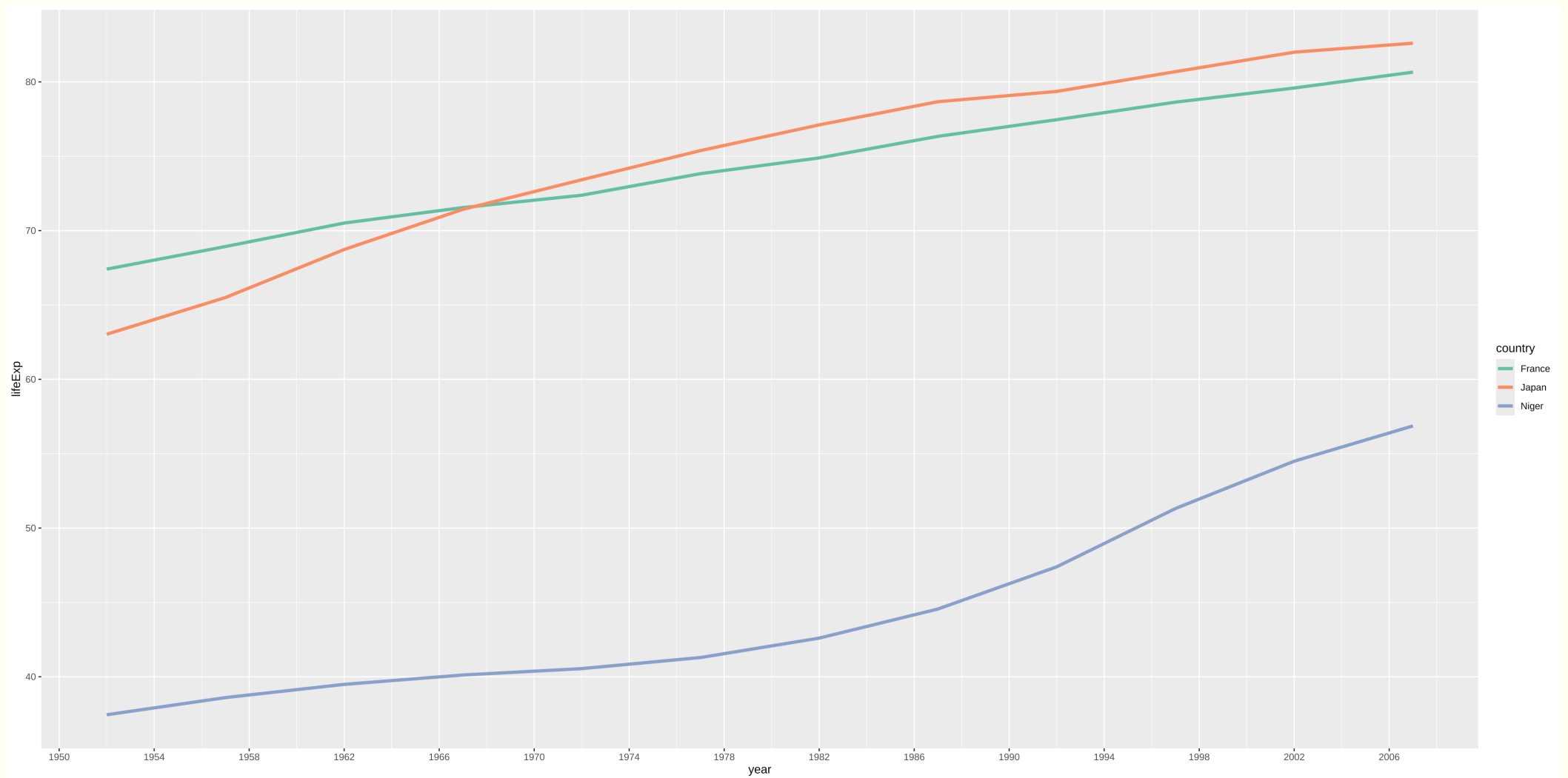
Example data

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
Japan	1957	65.500

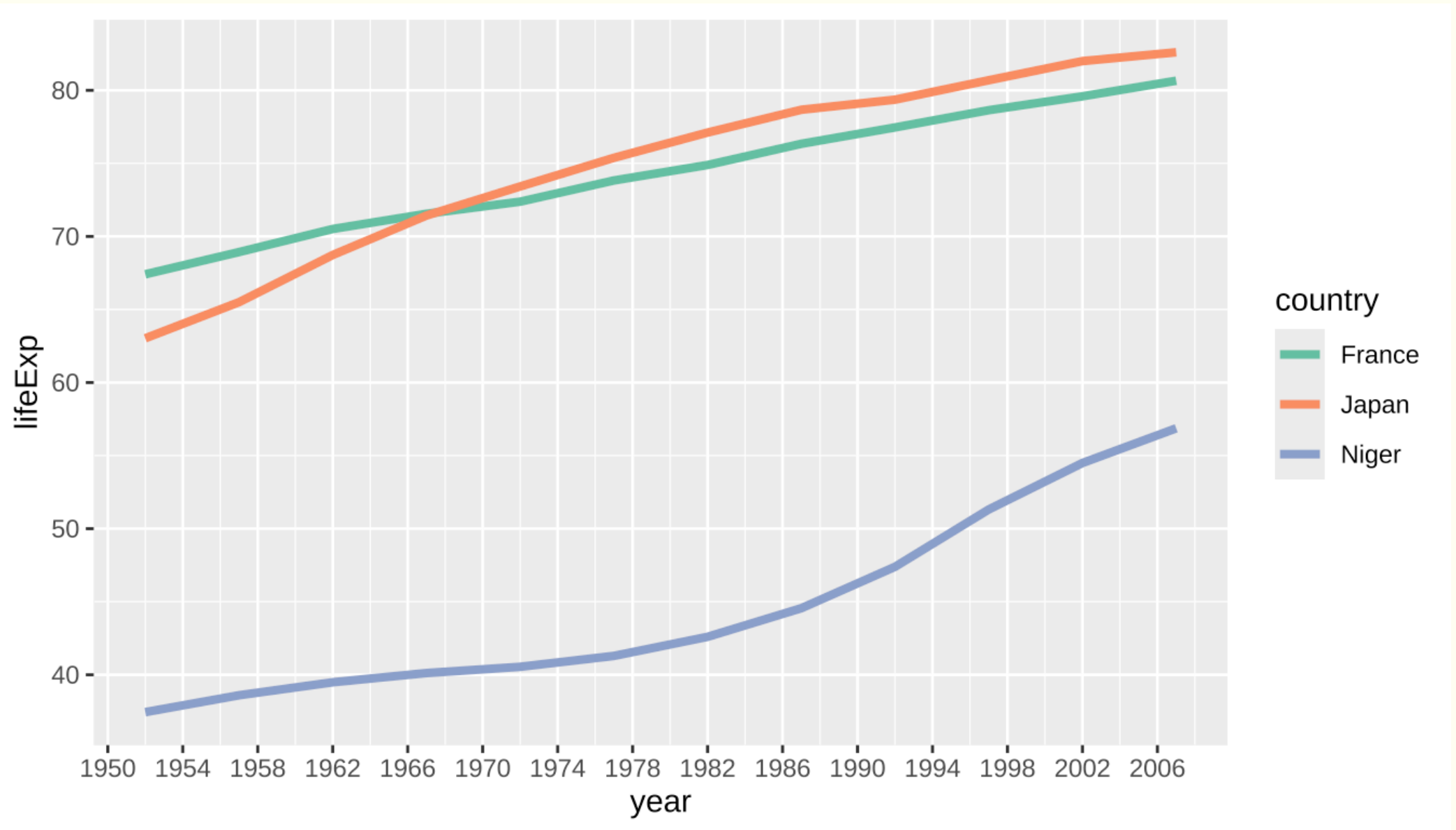
country	year	lifeExp
Japan	1962	68.730
Japan	1967	71.430
Japan	1972	73.420
Japan	1977	75.380
Japan	1982	77.110
Japan	1987	78.670
Japan	1992	79.360
Japan	1997	80.690
Japan	2002	82.000
Japan	2007	82.603
Niger	1952	37.444
Niger	1957	38.598
Niger	1962	39.487
Niger	1967	40.118
Niger	1972	40.546
Niger	1977	41.291

country	year	lifeExp
Niger	1982	42.598
Niger	1987	44.555
Niger	1992	47.391
Niger	1997	51.313
Niger	2002	54.496
Niger	2007	56.867

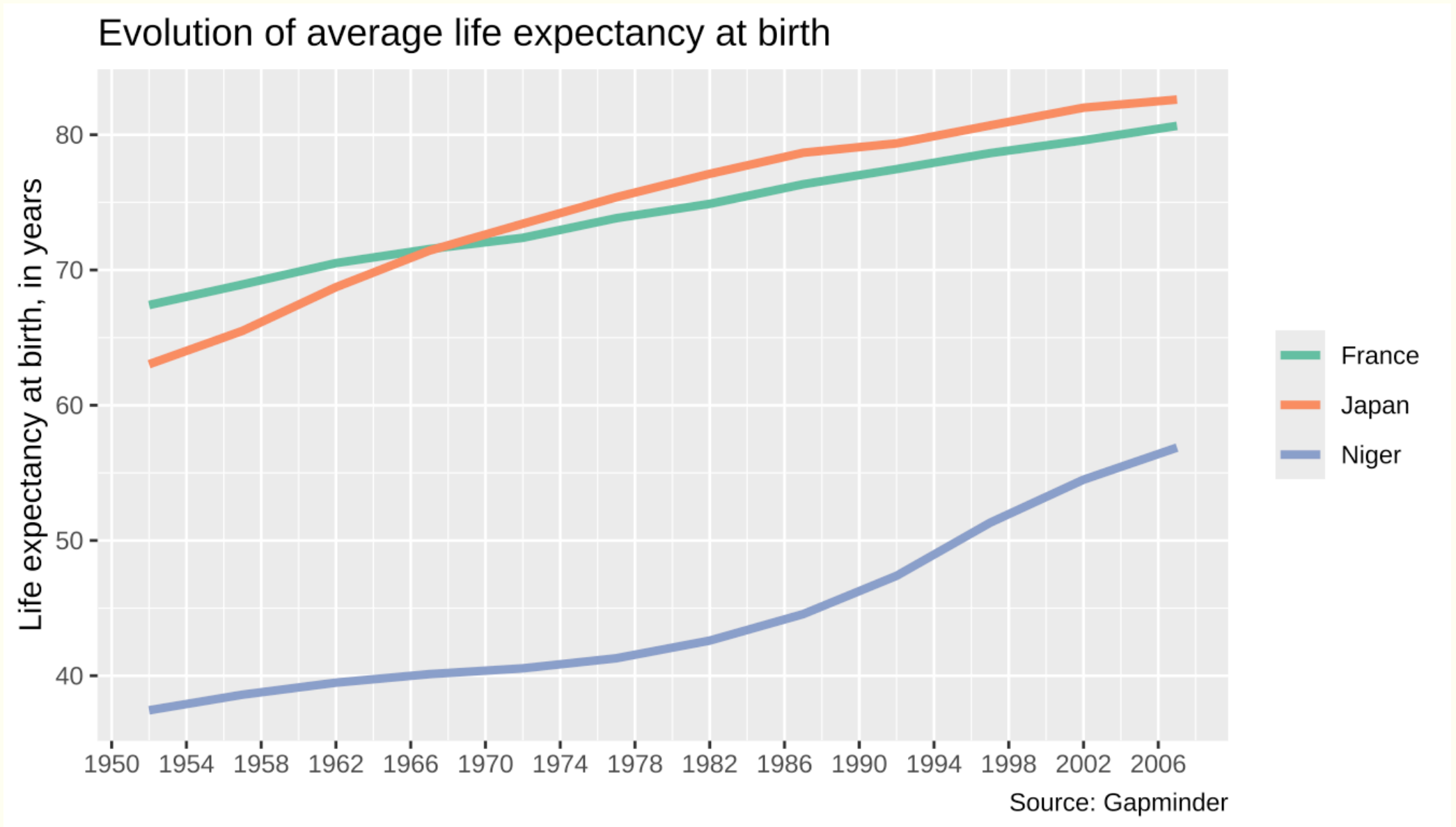
A concrete example



Legible text



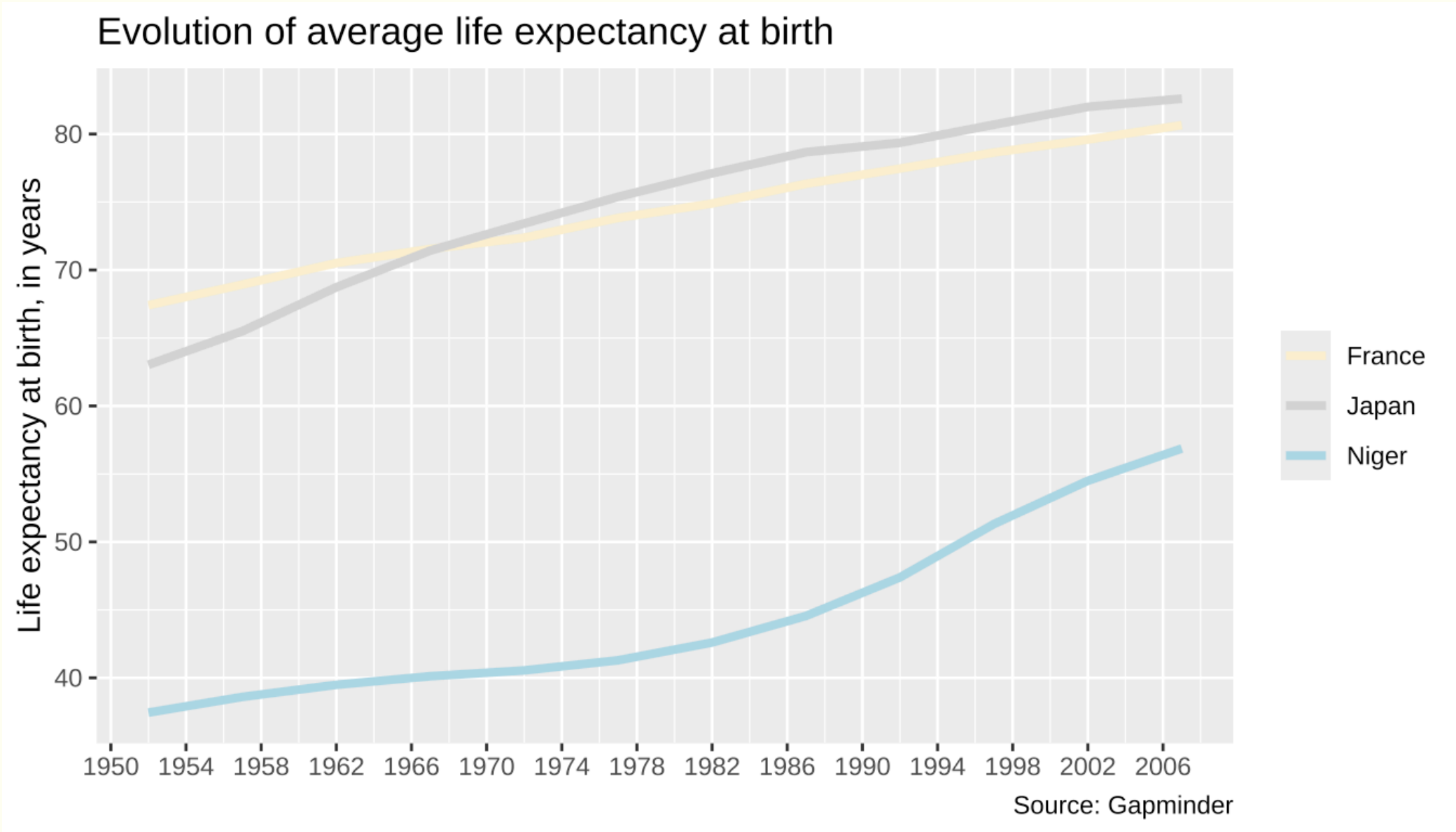
Title, clear axis labels and source



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?

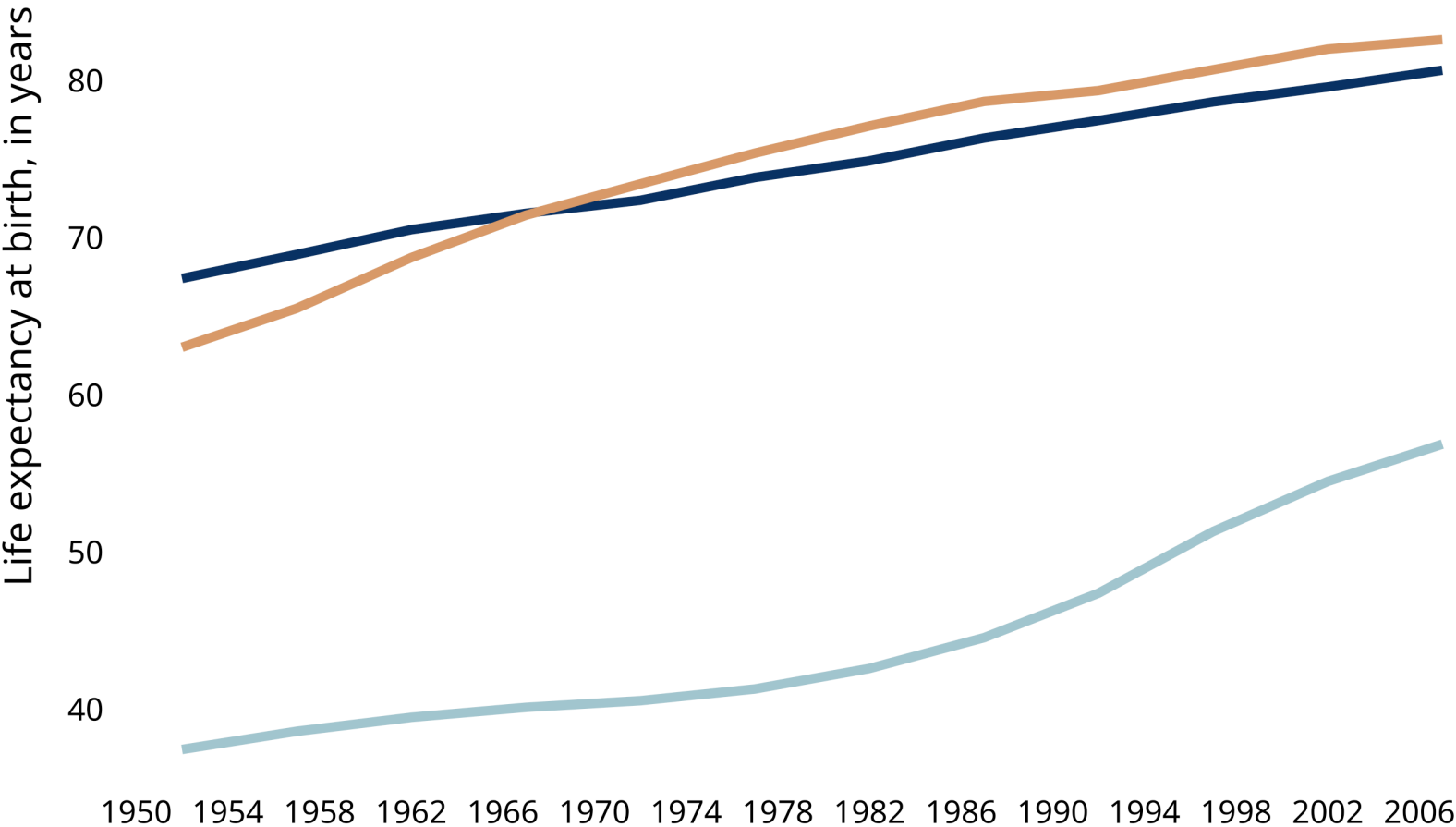
Avoid low contrast



Good practice

Evolution of average life expectancy at birth

■ France ■ Japan ■ Niger

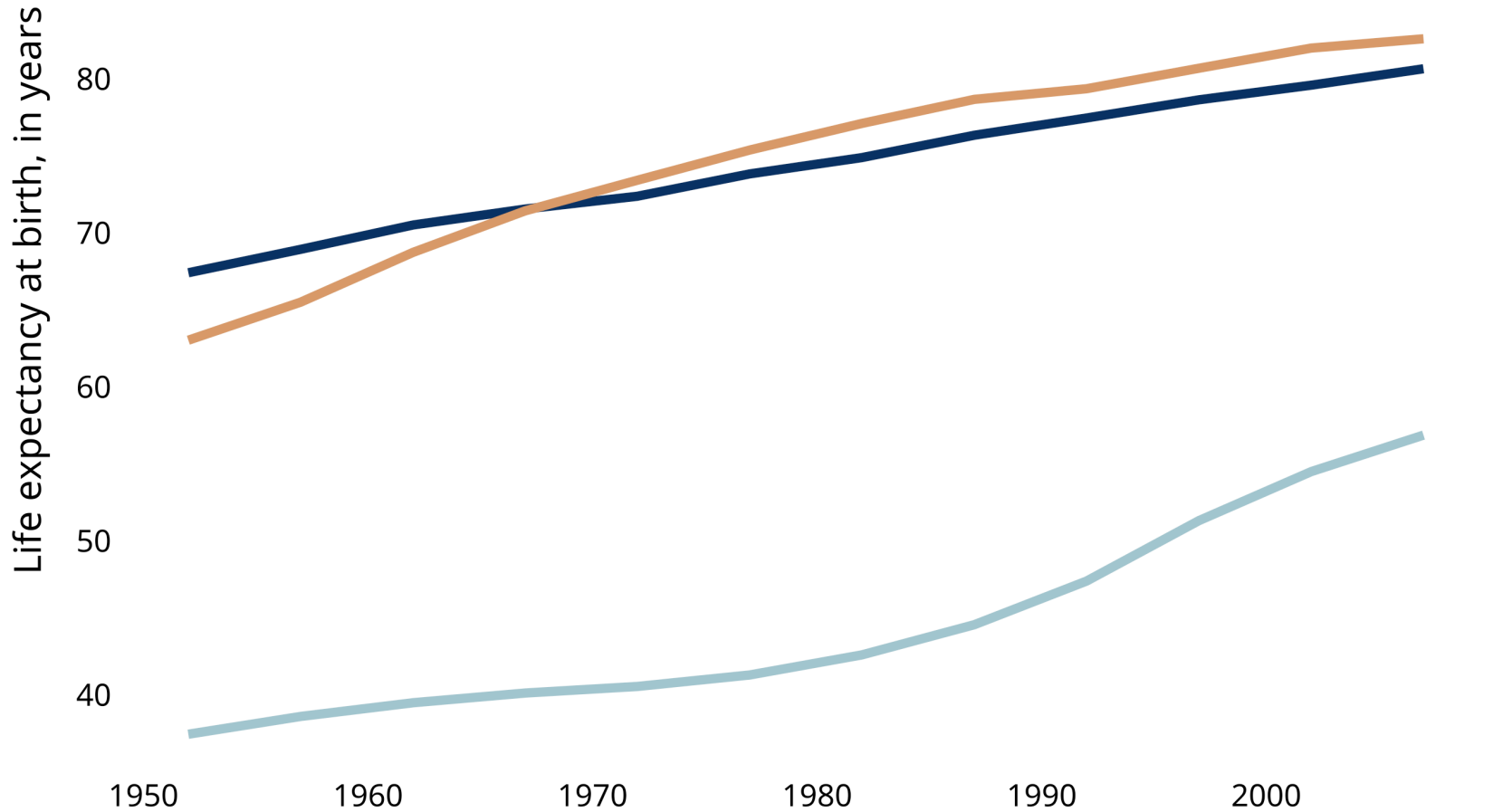


Source: Gapminder

Limited number of guides break

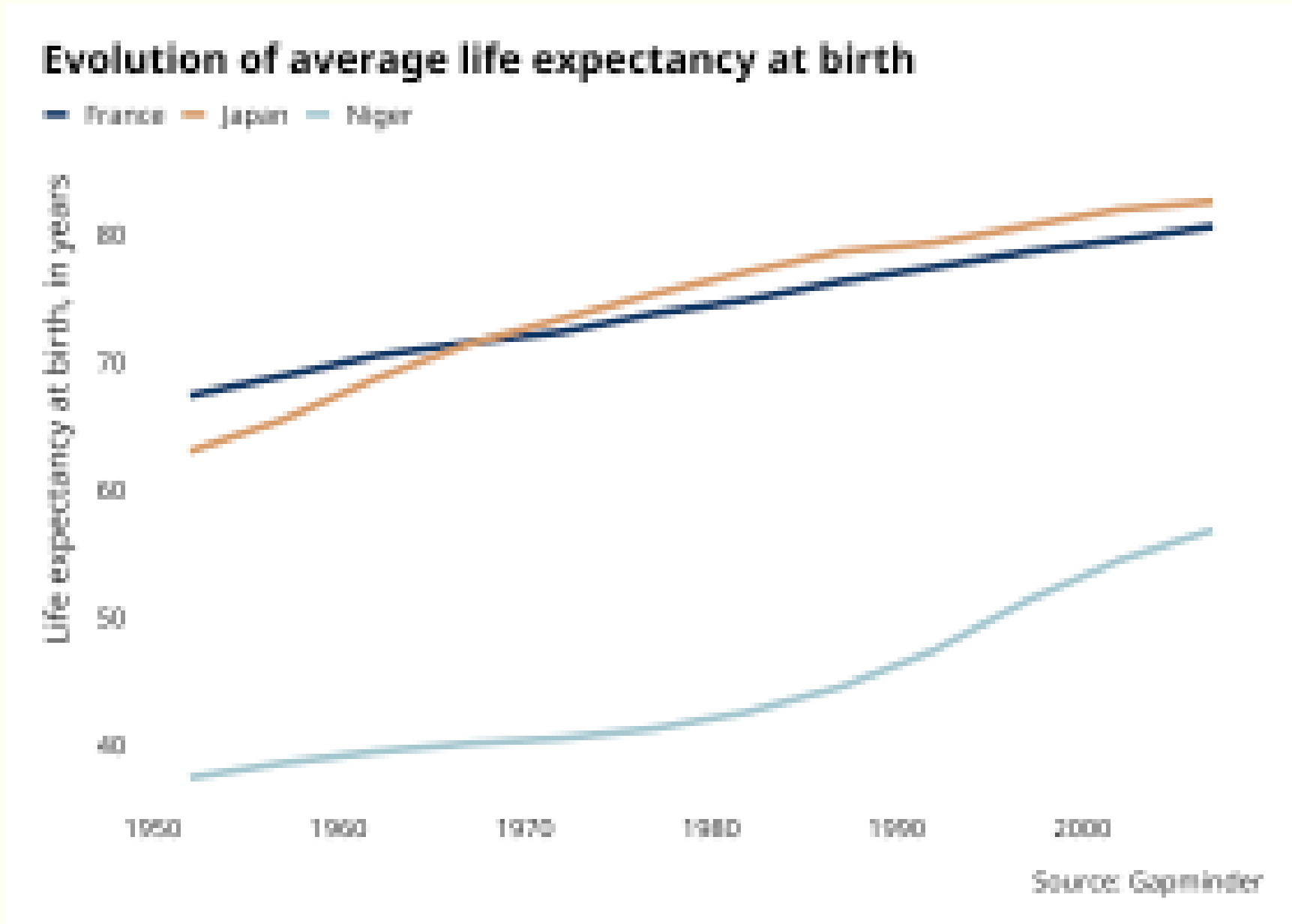
Evolution of average life expectancy at birth

■ France ■ Japan ■ Niger



Source: Gapminder

Export graphs in PDF (vectorized)

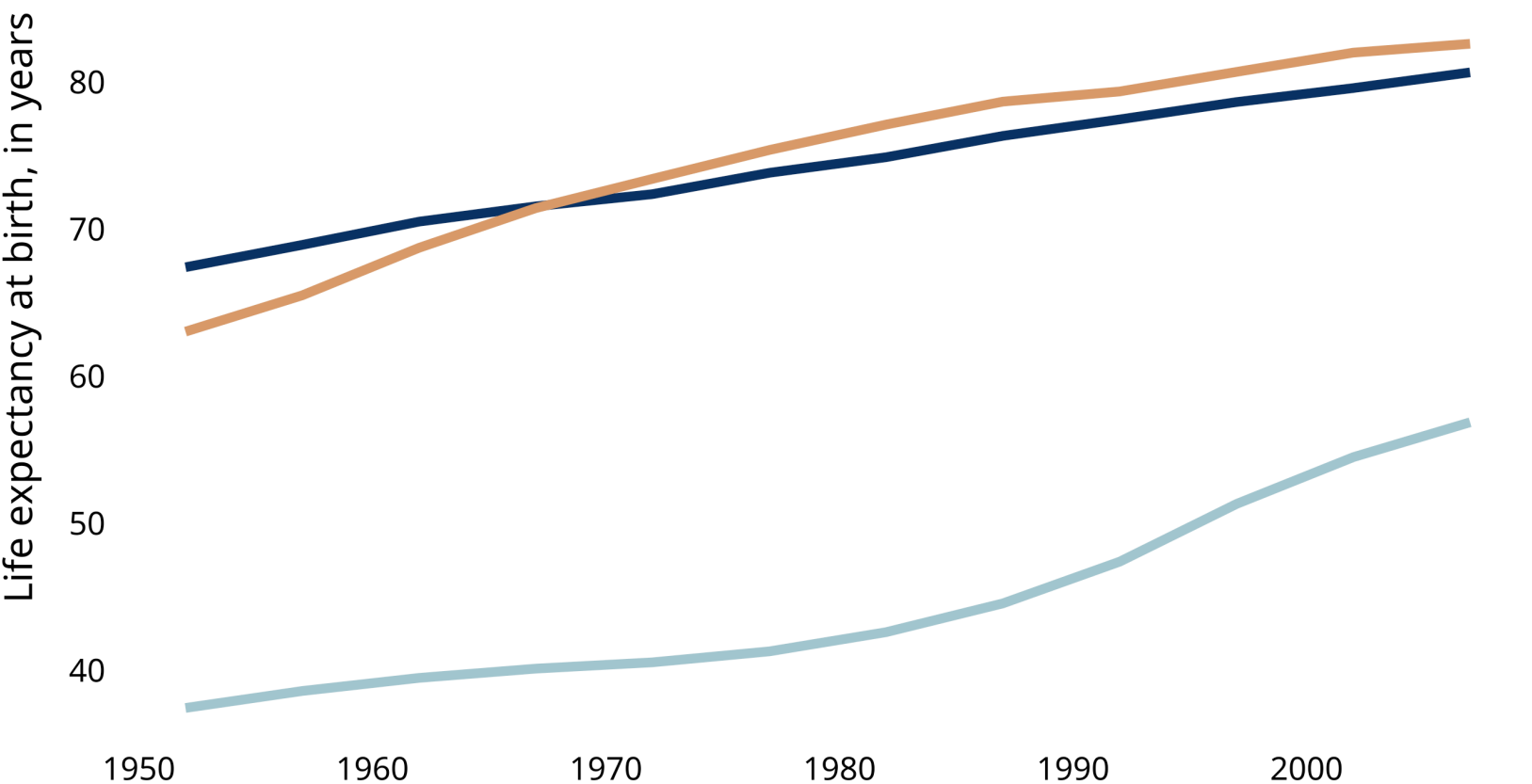


An informative title

Life expectancy remains substantially lower in Niger

Evolution of average life expectancy at birth

■ France ■ Japan ■ Niger



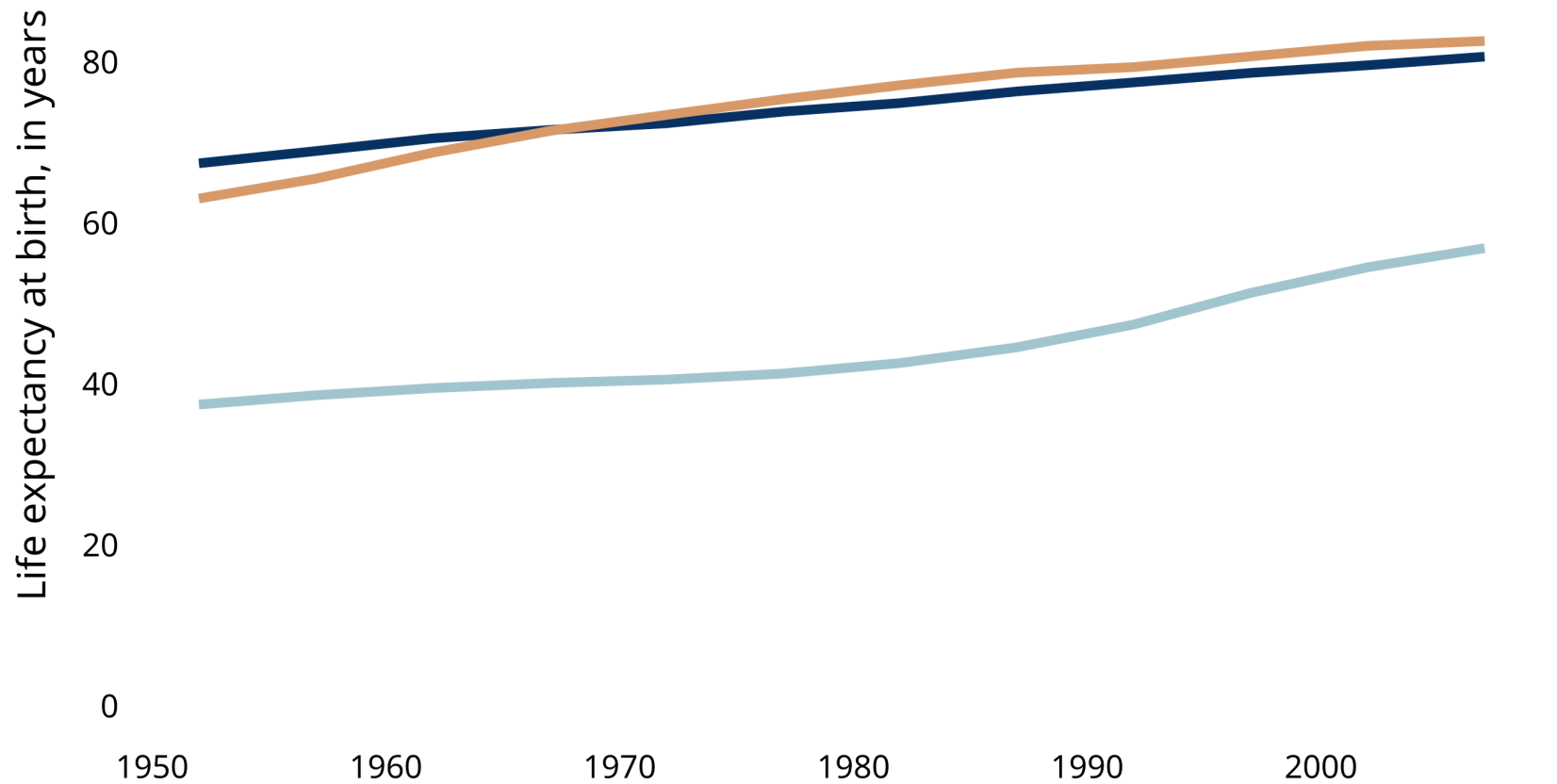
Source: Gapminder

Think about the scale of the y-axis

Life expectancy remains substantially lower in Niger

Evolution of average life expectancy at birth

■ France ■ Japan ■ Niger



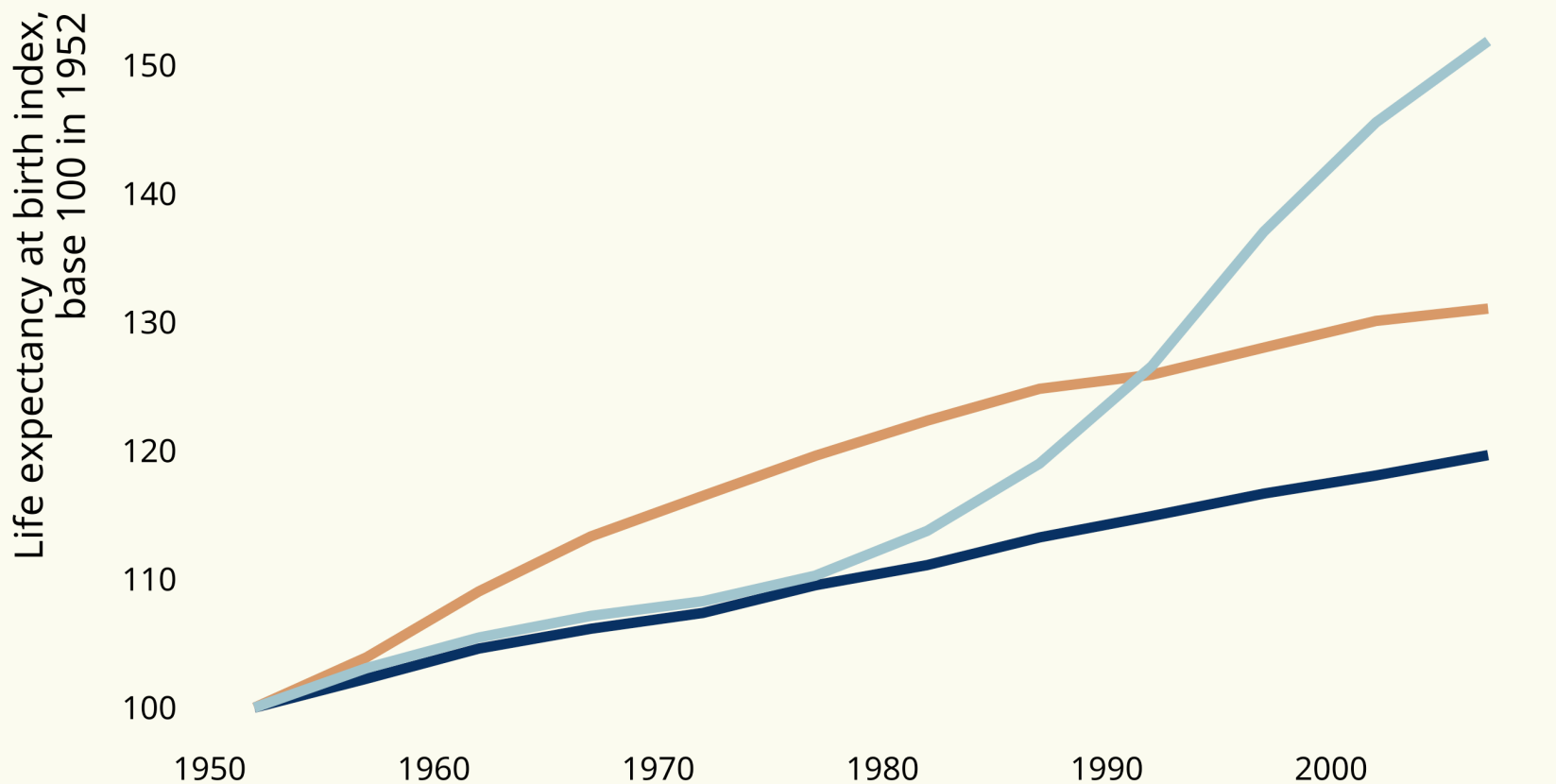
Source: Gapminder

An alternative story

Life expectancy at birth increased sharply in Niger

Evolution of an average life expectancy at birth index

■ France ■ Japan ■ Niger

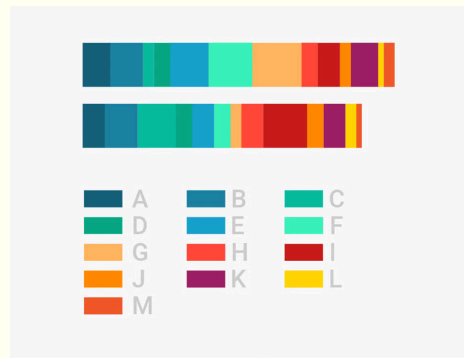


Source: Gapminder

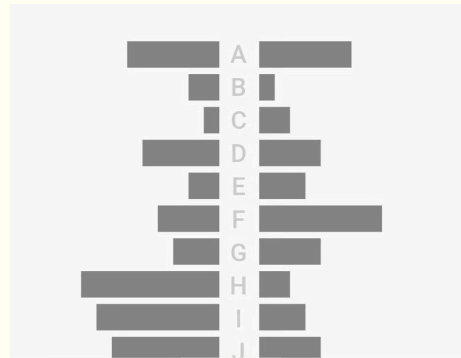
Avoid using many different colors

If need more than 7 colors or so:

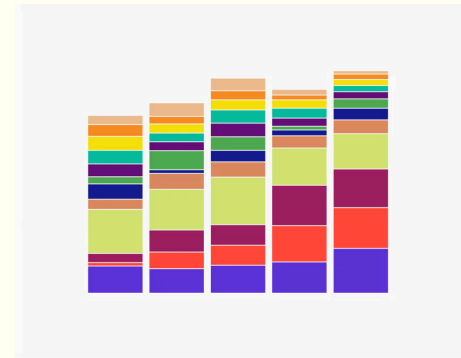
- Use another graph
- Group categories together



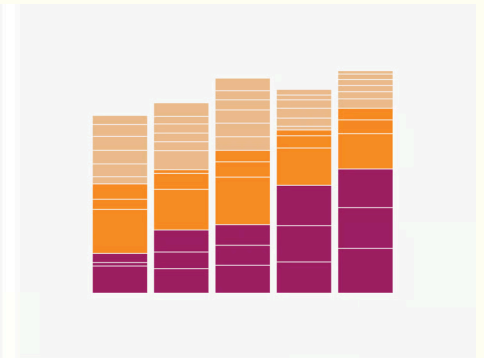
NOT IDEAL



BETTER



NOT IDEAL



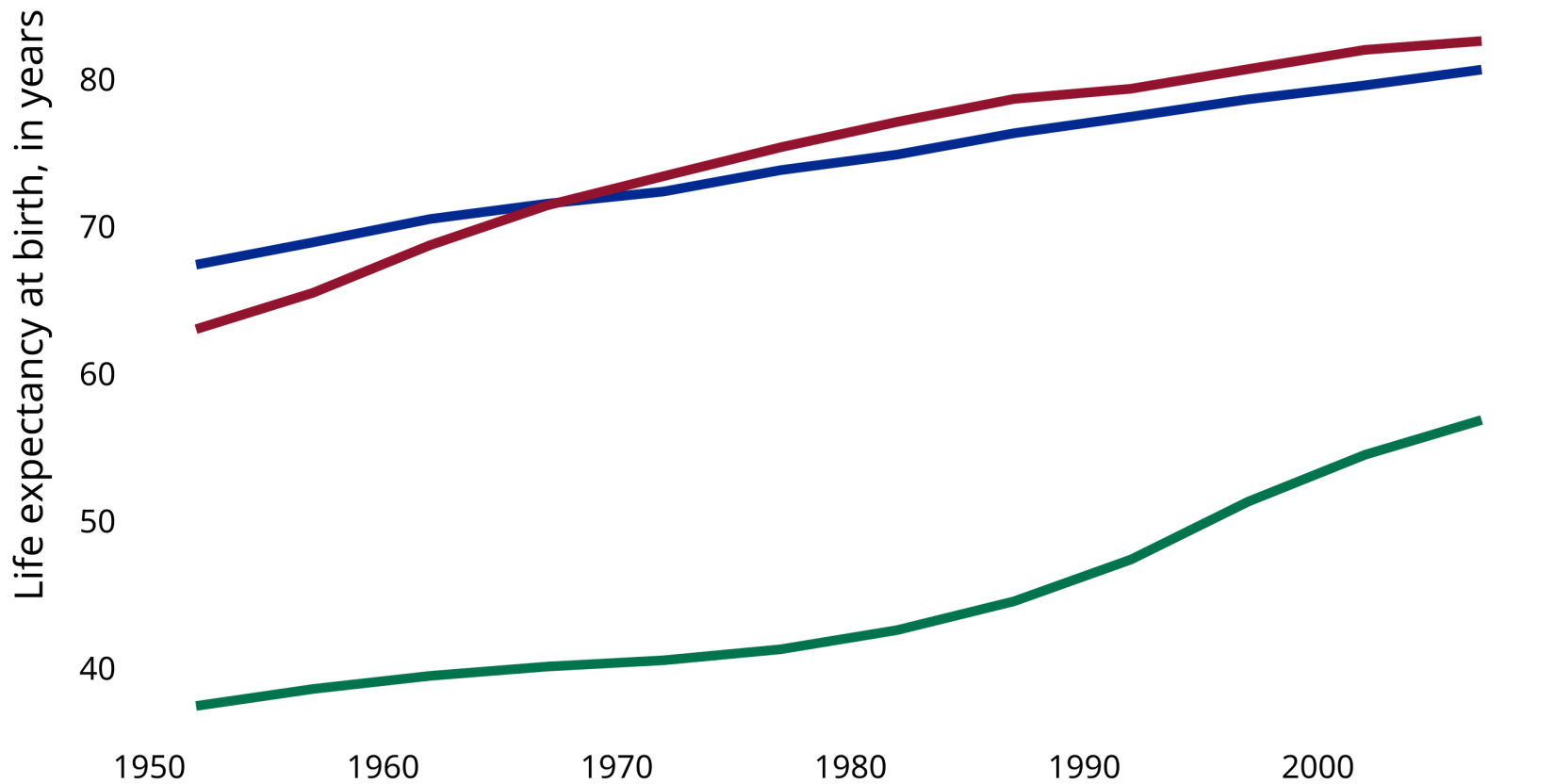
BETTER

Use intuitive colors

Life expectancy remains substantially lower in Niger

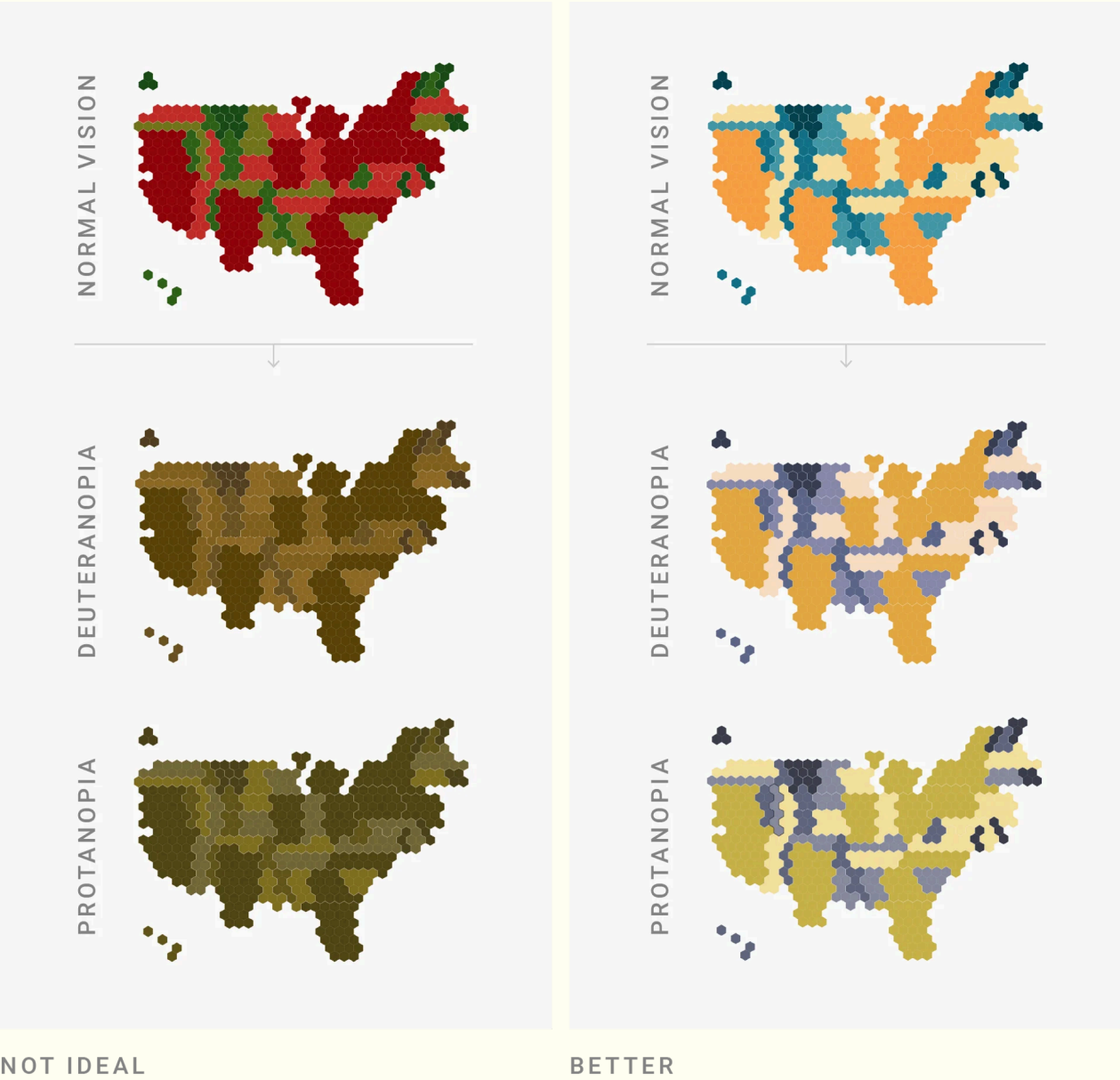
Evolution of average life expectancy at birth

■ France ■ Japan ■ Niger



Source: Gapminder

Colorblind-friendly visualizations

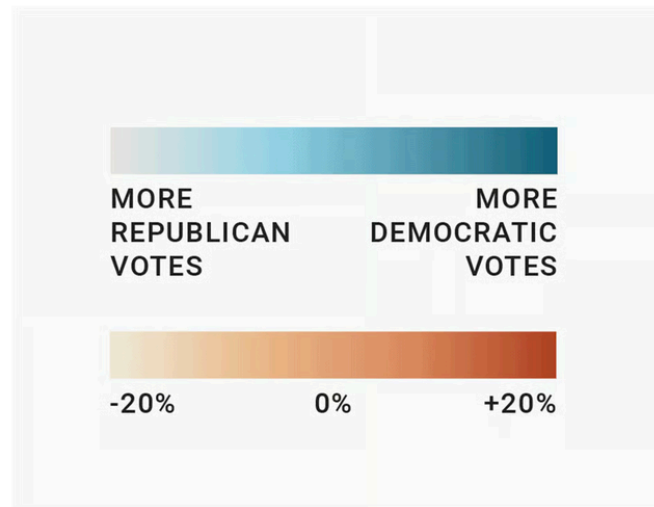


Diverging color scale (vs sequential)

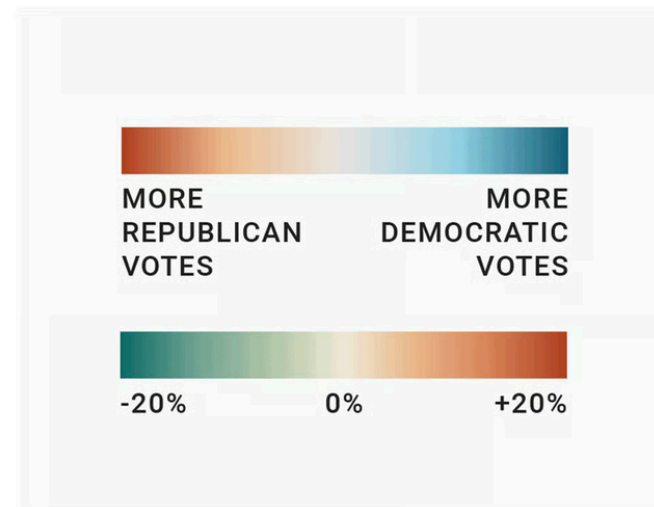
There is a mid-point

Emphasize extremes

See more difference



NOT IDEAL

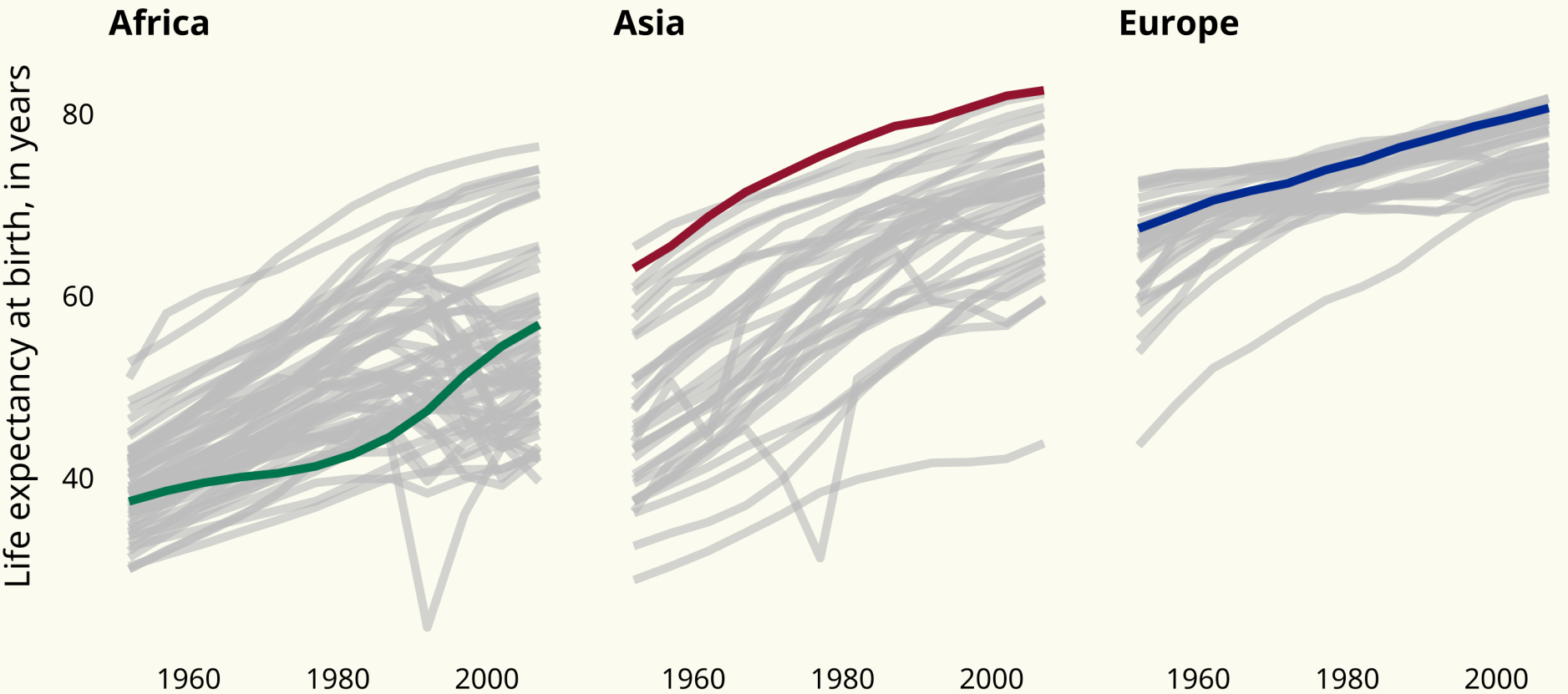


BETTER

Use gray. Emphasize.

Life expectancy in France, Japan and Niger as compared to their continent
Evolution of average life expectancy at birth, by continent

France Japan Niger

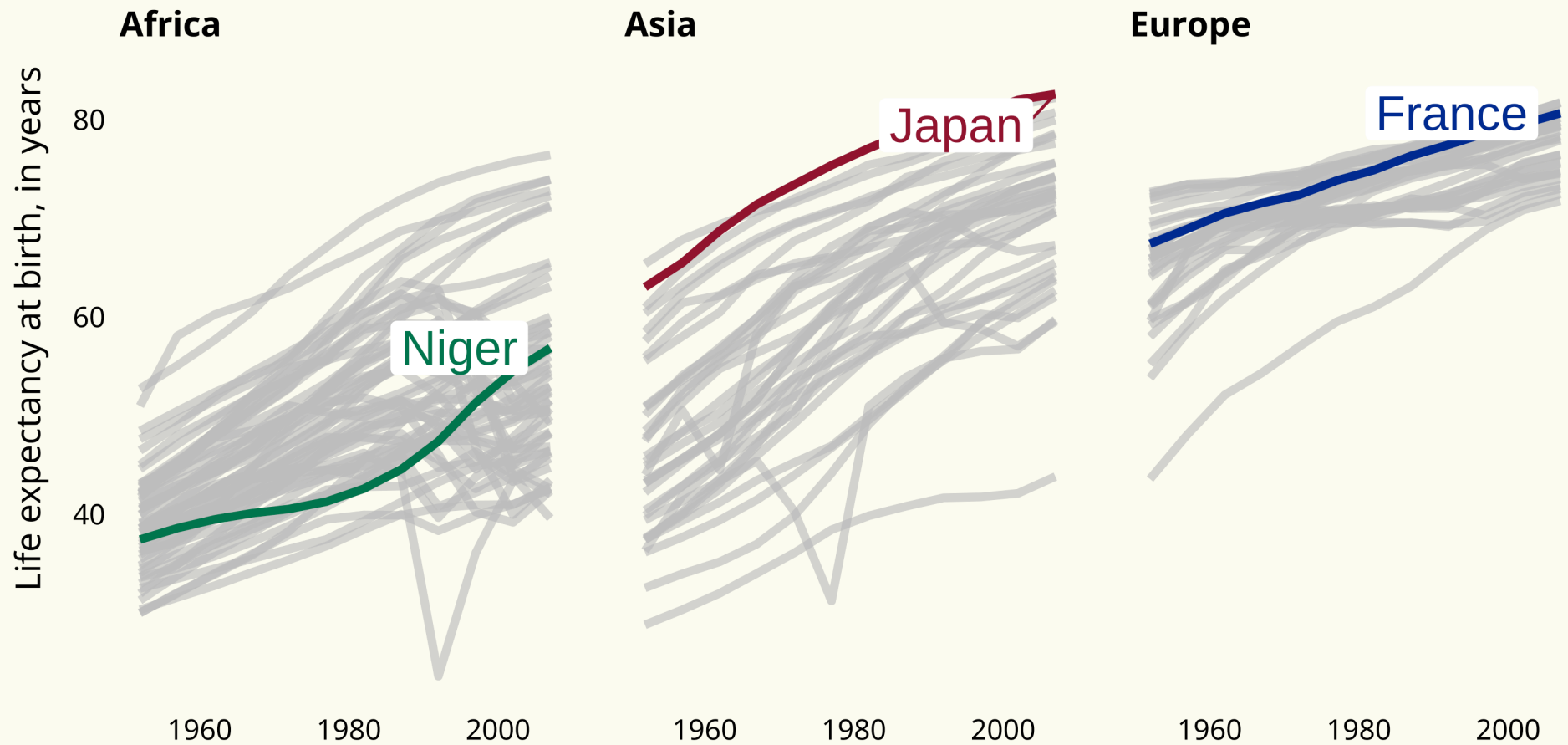


Source: Gapminder

Label directly

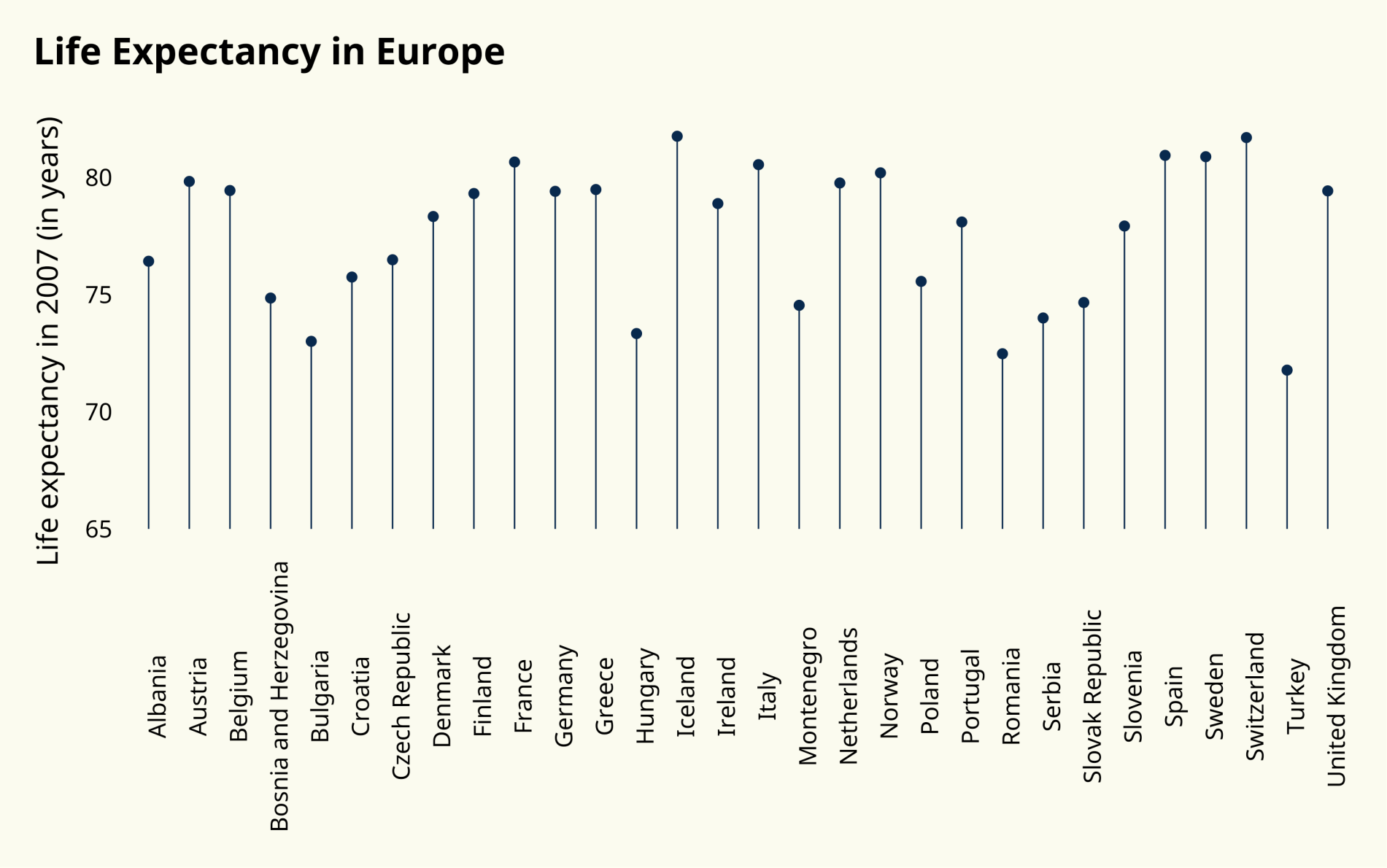
Life expectancy in France, Japan and Niger as compared to their own continent

Evolution of average life expectancy at birth, by continent



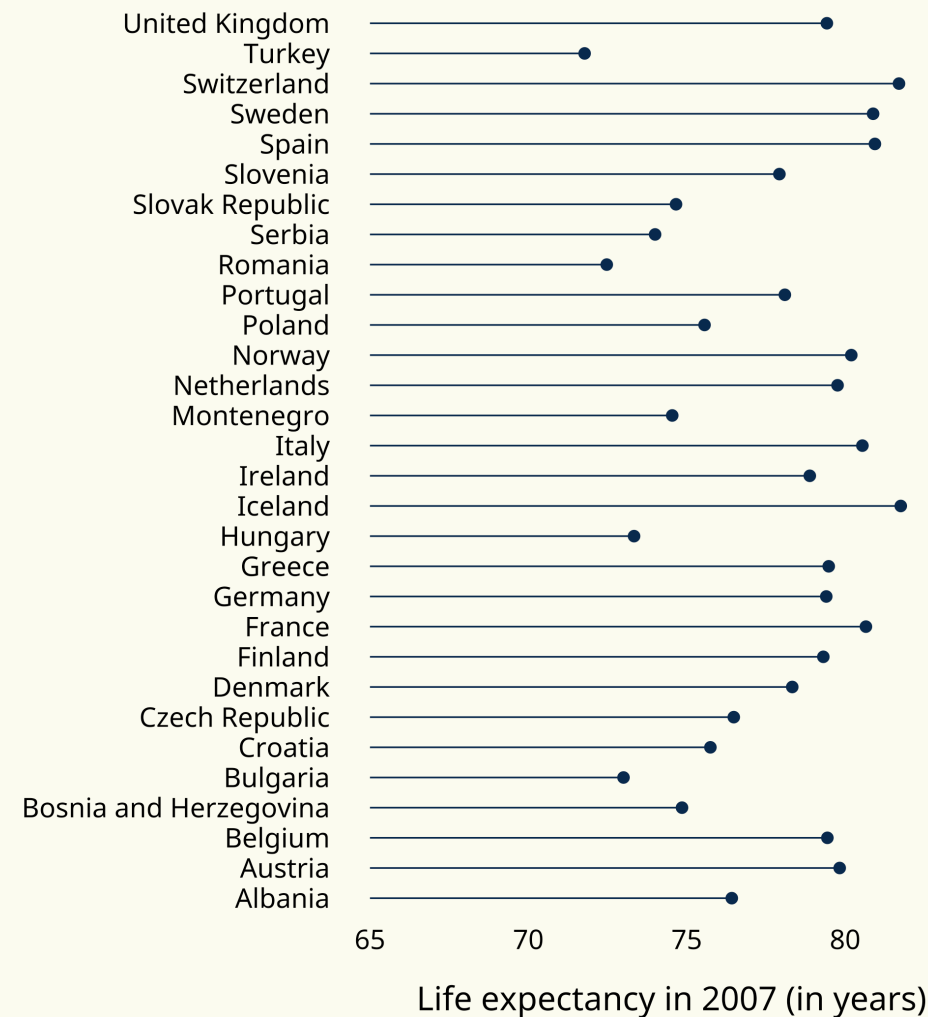
Source: Gapminder

Horizontal text

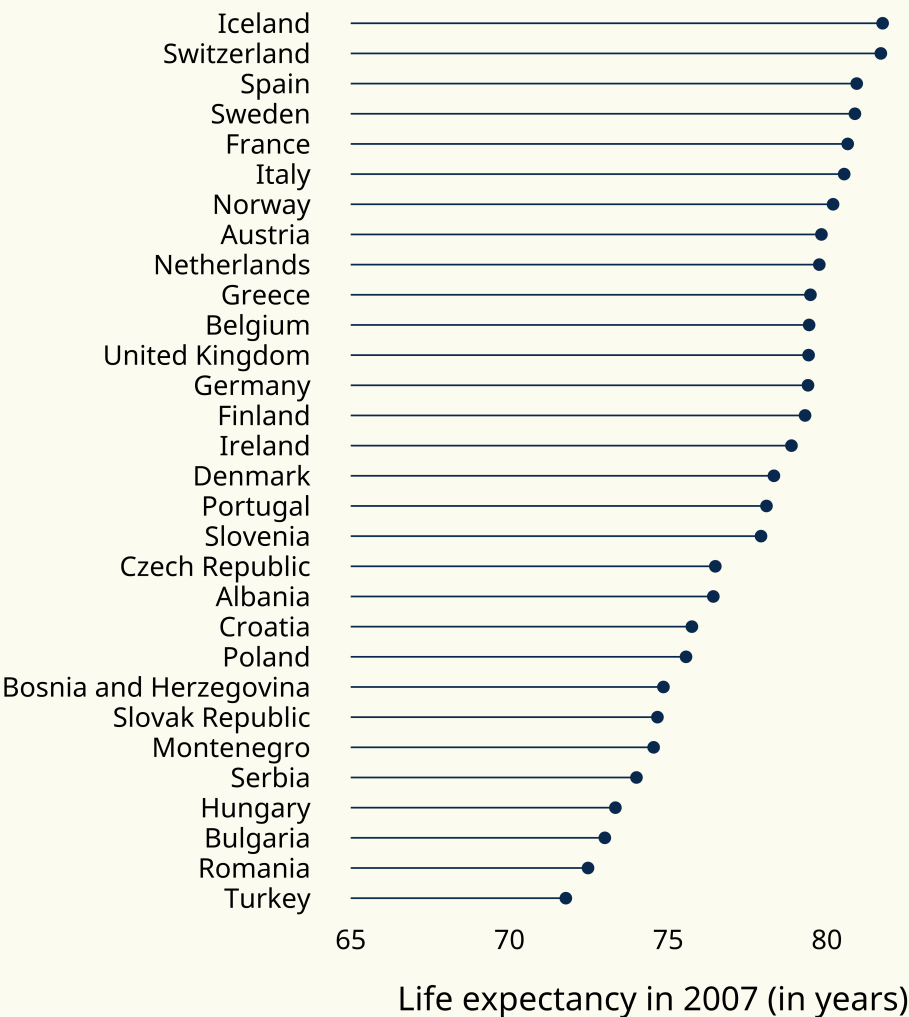


Order your data

Life Expectancy in Europe



Life Expectancy in Europe



Data viz caveats

CAVEATS

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

Show all

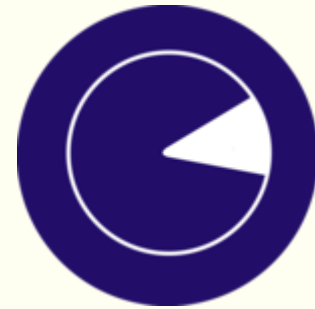
Top 10

Improvement

Misleading

Map

Bar



Sometimes you may break the rules

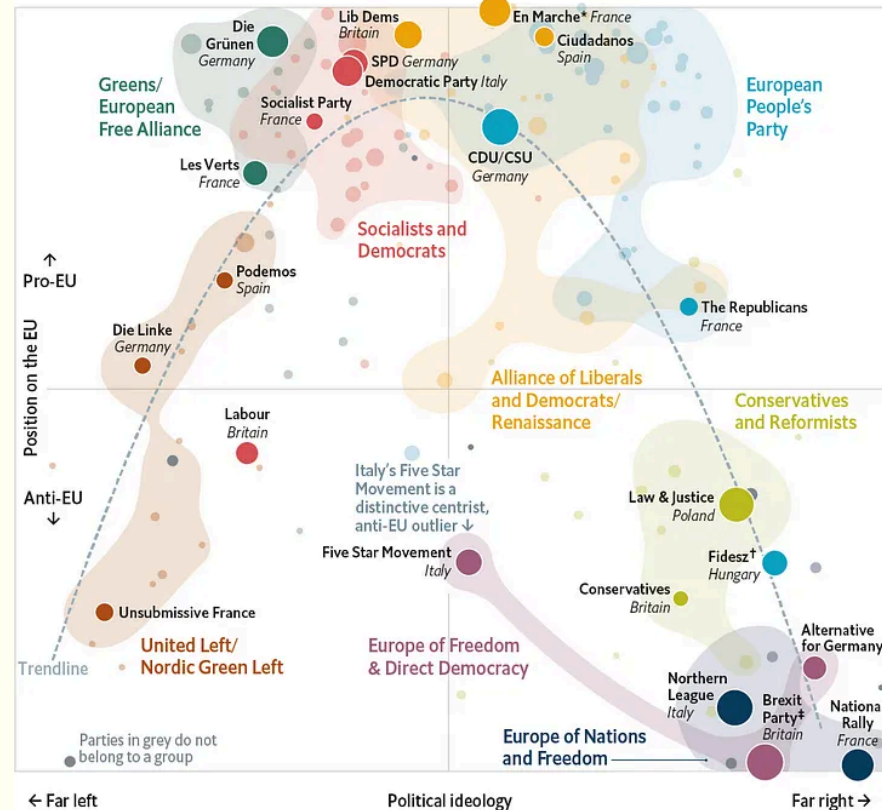
Anti-EU parties cluster at ideological extremes, whereas pro-EU ones are centrist

European Parliament political parties and groupings

By ideology and position on the EU

Party name — Parliamentary grouping

2019 election, provisional results, seats ○ 10 ○ 20



← Far left Political ideology Far right →
*Includes MoDem and UDI †Currently suspended from the EPP group ‡Ideology and difference based on UKIP 2014
Sources: Chapel Hill Expert Survey (2014/2017); ECFR; European Parliament

9 colors

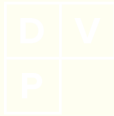
BUT

- Labeled directly
- Shade reinforces grouping

Choosing a type of 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/](https://datavizproject.com/))

FAMILY

INPUT

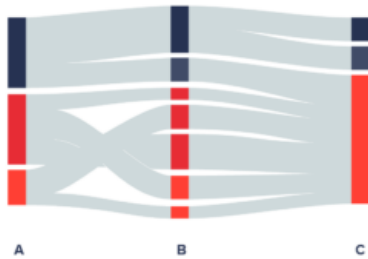
FUNCTION

SHAPE

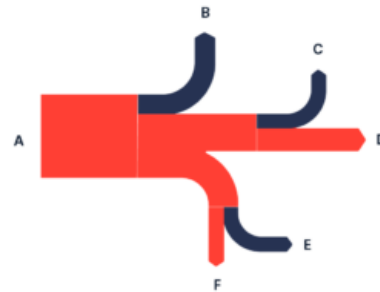
([HTTPS://DATAVIZPROJECT.COM/](https://datavizproject.com/))

(<https://datavizproject.com/>)

Alluvial Diagram



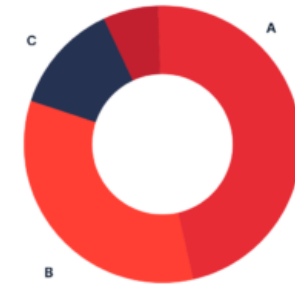
Sankey Diagram



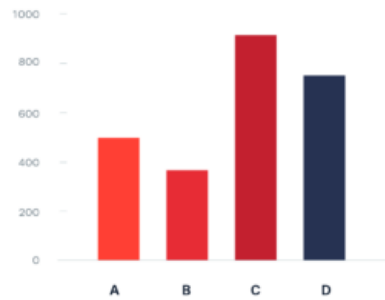
Line Graph



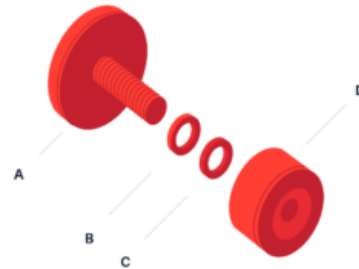
Donut Chart



(<https://datavizproject.com/data-> (<https://datavizproject.com/data-> (<https://datavizproject.com/data-> ([**Bar Chart \(Vertical\)**](https://datavizproject.com/data-</p></div><div data-bbox=)



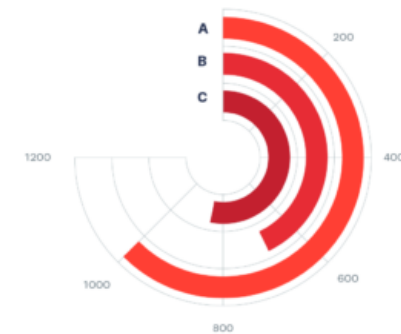
Exploded View Drawing



Polar Area Chart



Radial Bar Chart

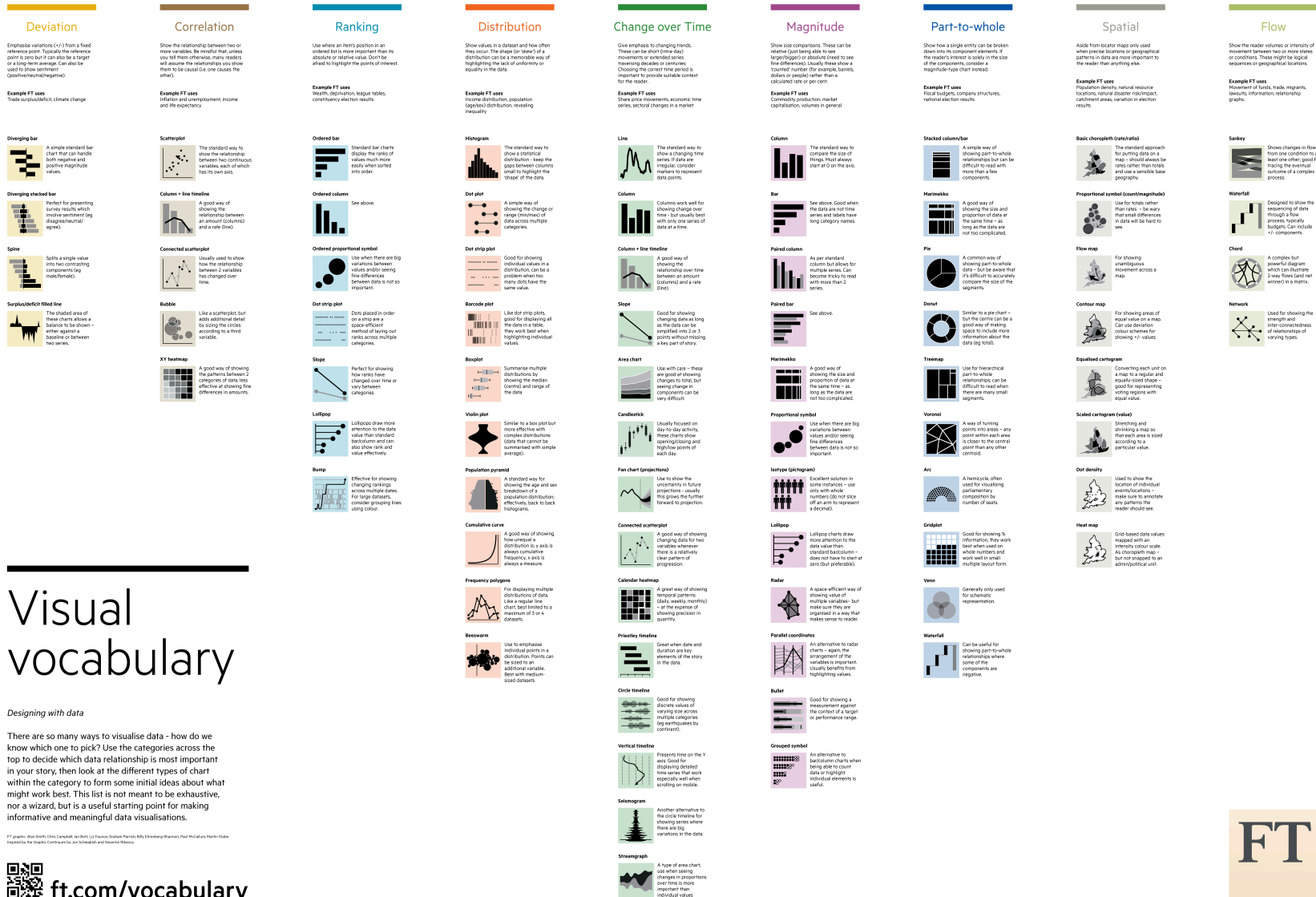


Type of relationship to show

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



Pay attention to good 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

“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

Communicating in economics

Specificities of communicating in economics

- Often have a specific type of output: **estimation results**
- There are **norms** for tables, in particular: deviating from them is costly
- And a **massive number of observations**
- Our analyses rest on **identifying assumptions** → can explore with graphs and tables
- We often use very complex models. Visualization can help **understand what we are actually estimating**

Graphs and tables in economics

When do we use them in econ?

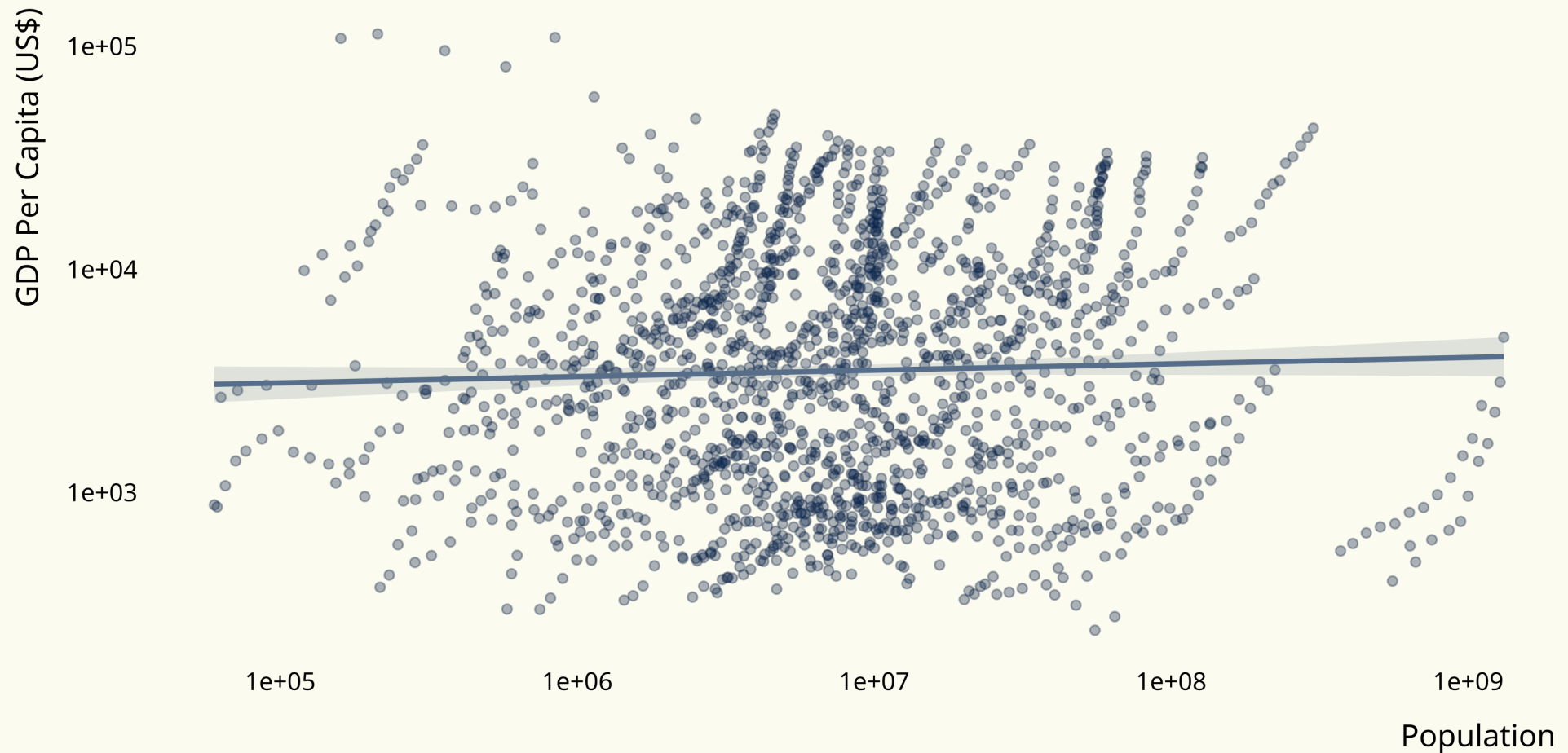
- As **rhetorical** visualization tools for models
- To **explore** our data
- To check the **validity** of our models and assumptions
- As **diagnostics**
- To **communicate** results

Rhetorical tool: FEs

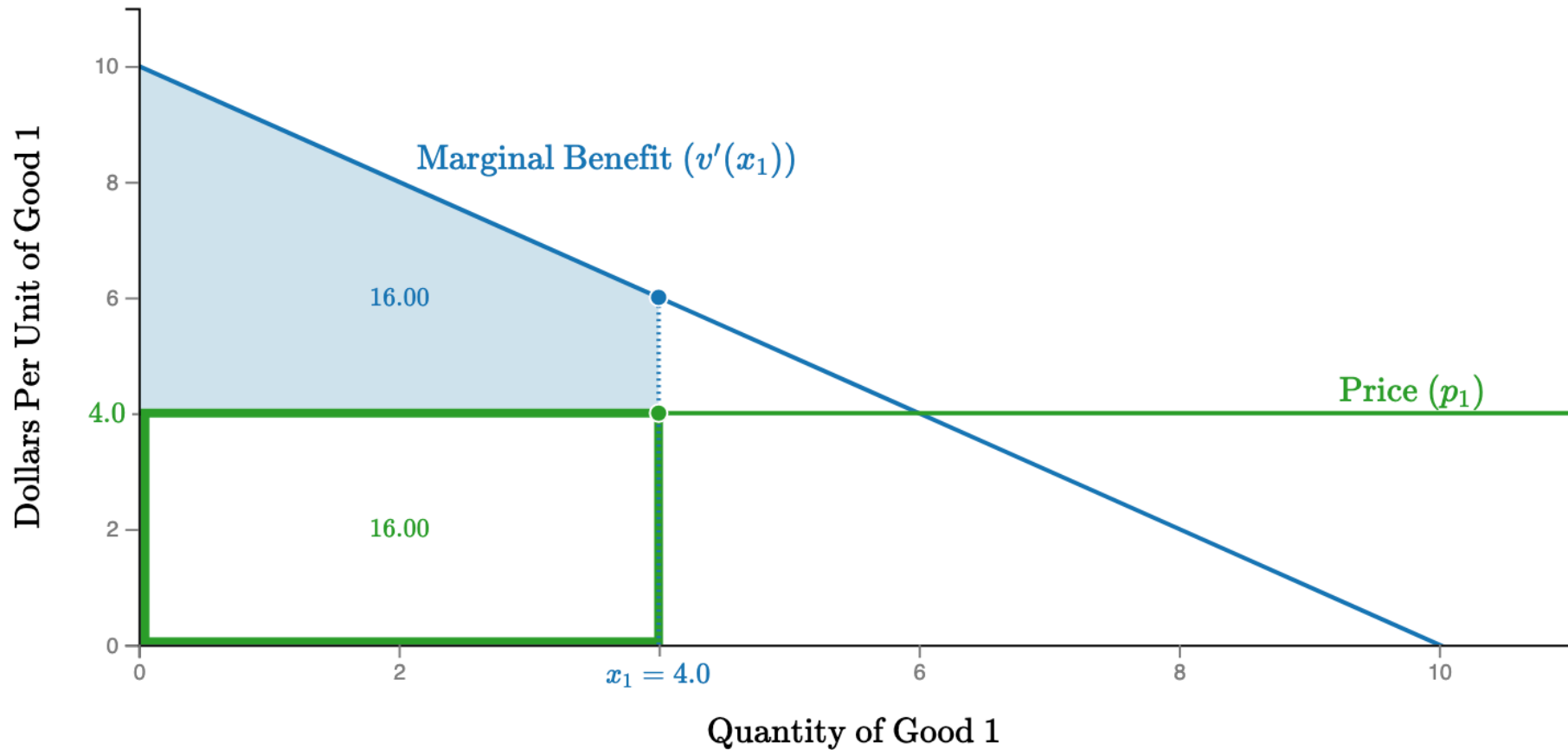
Pooled relationship

Within country

Population Against GDP per Capita



Rhetorical tool: models



Example

- We will run a **simple analysis**:
 - Link between **jail sentences** for drunk driving and **traffic deaths**?
- At each step, we ask the same two questions:
 - Which **graph** helps here?
 - Which **table** helps here?
- Side goal: provide practical tips to make these tables and graphs (eg chunks of code, packages, etc)

Setting of the running example

Data	What's in it?		Summary		Take-aways		Cleaning		Code		
state	year	spirits	unemp	income	emppop	beertax	baptist	mormon	drinkage	dry	youngdrivers
al	1982	1.37	14.4	10544.152	50.69204	1.5393795	30.35570	0.328290	19.00	25.006300	0.211572
al	1983	1.36	13.7	10732.798	52.14703	1.7889907	30.33360	0.343410	19.00	22.994200	0.210768
al	1984	1.32	11.1	11108.791	54.16809	1.7142856	30.31150	0.359240	19.00	24.042601	0.211484
al	1985	1.28	8.9	11332.627	55.27114	1.6525424	30.28950	0.375790	19.67	23.633900	0.211140
al	1986	1.23	9.8	11661.507	56.51450	1.6099070	30.26740	0.393110	21.00	23.464701	0.213400
al	1987	1.18	7.8	11944.000	57.50988	1.5599999	30.24530	0.411230	21.00	23.792400	0.215527
al	1988	1.17	7.2	12368.624	56.83453	1.5014436	30.22330	0.430180	21.00	23.792400	0.218328
az	1982	1.97	9.9	12309.069	56.89330	0.2147971	3.95890	4.919100	19.00	0.000000	0.209012
az	1983	1.90	9.1	12693.808	57.55363	0.2064220	3.89010	4.831070	19.00	0.000000	0.203855
az	1984	2.14	5.0	13265.934	60.37902	0.2967033	3.82260	4.744610	19.00	0.000000	0.209127
az	1985	1.86	6.5	13726.695	58.64853	0.3813559	3.75620	4.659710	21.00	0.000000	0.188428
az	1986	1.78	6.9	14107.327	60.28018	0.3715170	3.69100	4.576320	21.00	0.000000	0.171539
az	1987	1.72	6.2	14241.000	60.21506	0.3600000	3.62690	4.494420	21.00	0.000000	0.168724
az	1988	1.68	6.3	14408.085	60.49767	0.3464870	3.56400	4.413990	21.00	0.000000	0.161005
ar	1982	1.19	9.8	10267.303	54.47586	0.6503580	22.96720	0.328290	21.00	36.712799	0.204903
ar	1983	1.20	10.1	10433.486	53.81479	0.6754587	23.00090	0.343410	21.00	36.430099	0.194169
ar	1984	1.22	8.9	10916.483	54.67128	0.5989011	23.03460	0.359240	21.00	36.104000	0.186380

Use for data exploration

- Graphs and tables can help understand the **structure** of our data:
 - Distributions
 - Simple relationships
- Then, can help to:
 - Form hypotheses
 - Explore identifying assumptions
 - Build accurate models
- May want to plot the same graph many times → make **functions**, eg my balanced graphs below
- To communicate **descriptive** analyses
- To explore potential mechanisms

Explore data: raw data

Diff. outcome

Over time

Where

Treatment

Code

Difference in fatal accidents depending on jail penalty for drunk

Number of fatal accidents per state

3000

1000

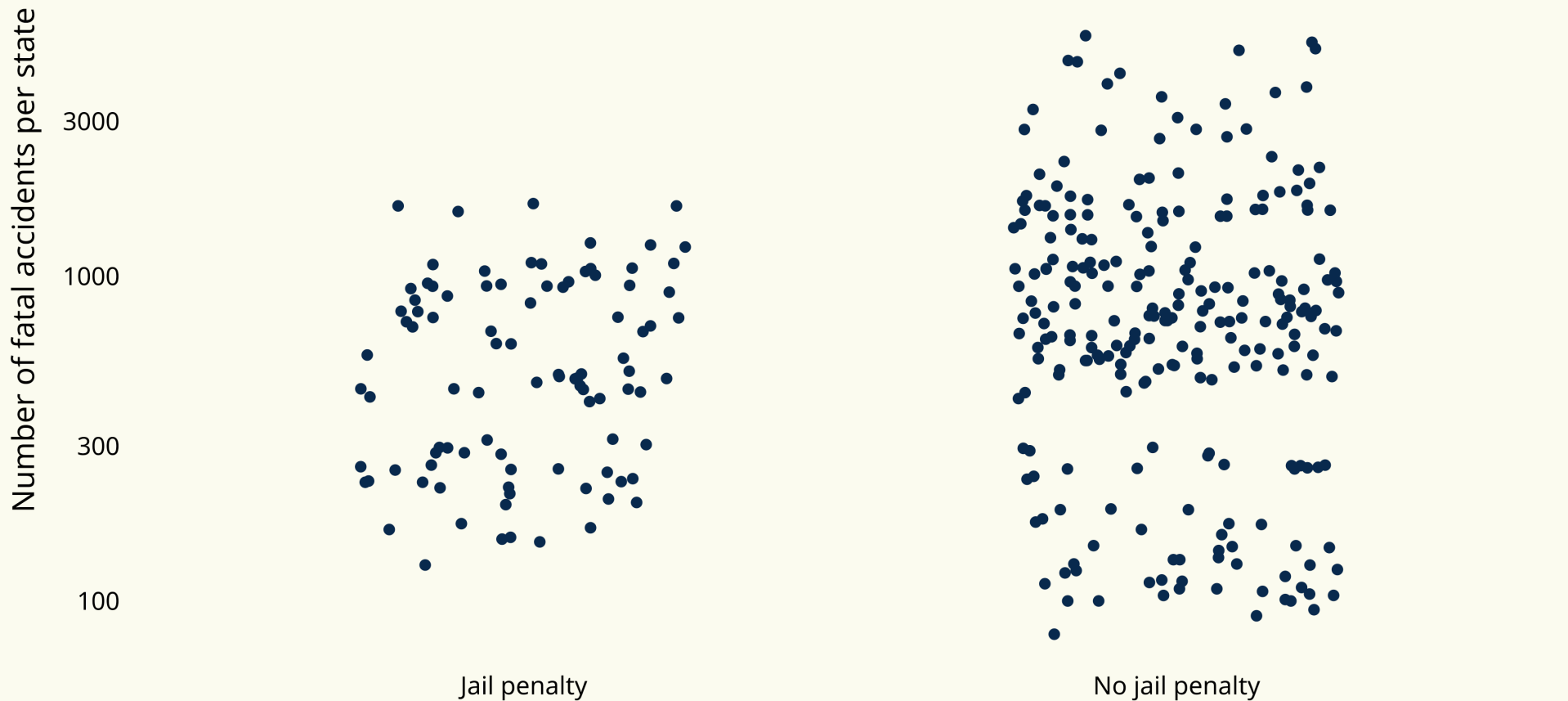
300

100

Jail penalty

No jail penalty

State law on drunk driving



Handling large numbers of observations

Raw

Opacity

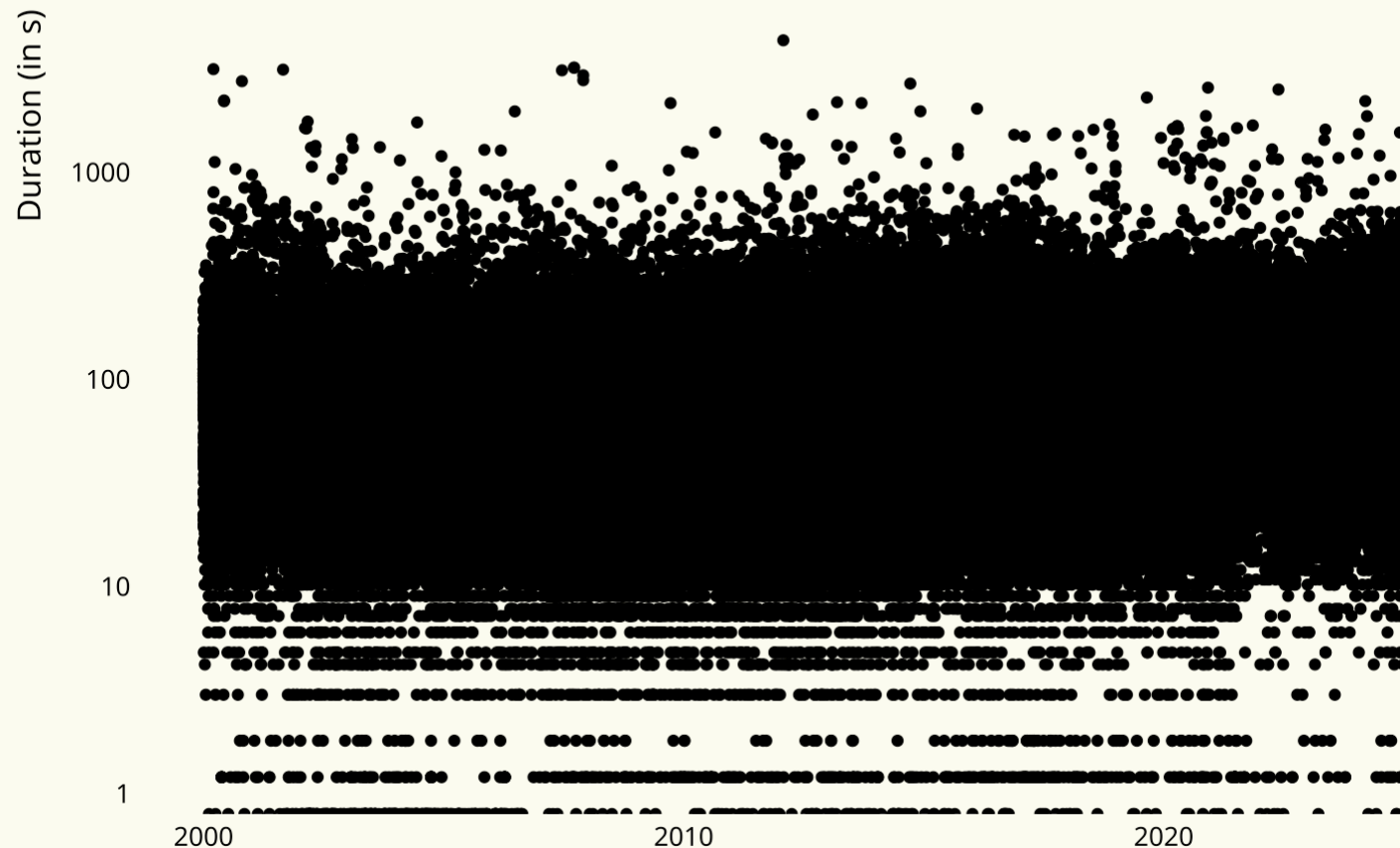
Heatmap

Binscatter

Code

Duration of evening news items French TV

For TF1 and France 2



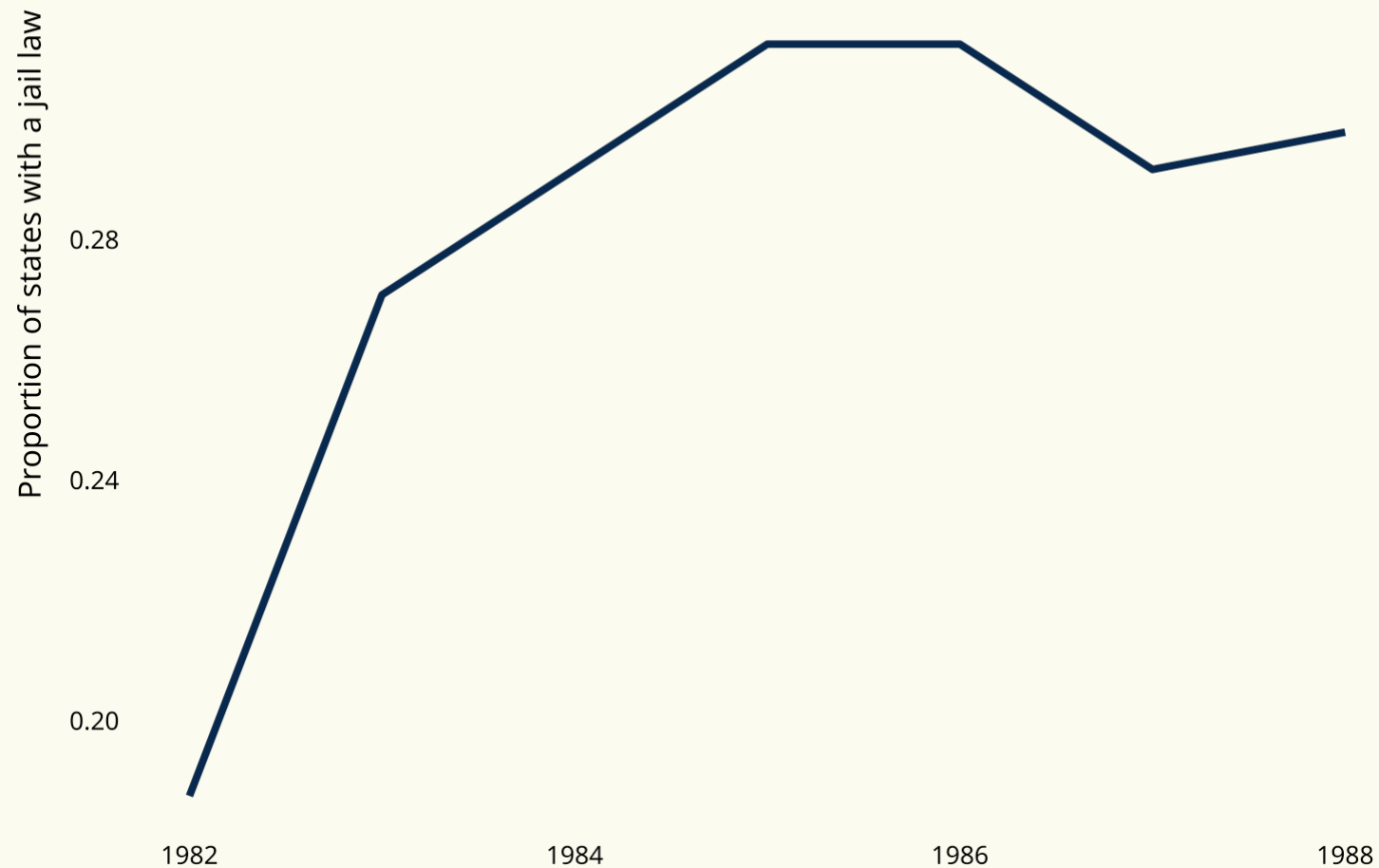
Explore data: finding sources of variation

Evolution

Map

Code

Adoption of jail sentence for drunk driving



Explore data: balance

Unemployment

Income

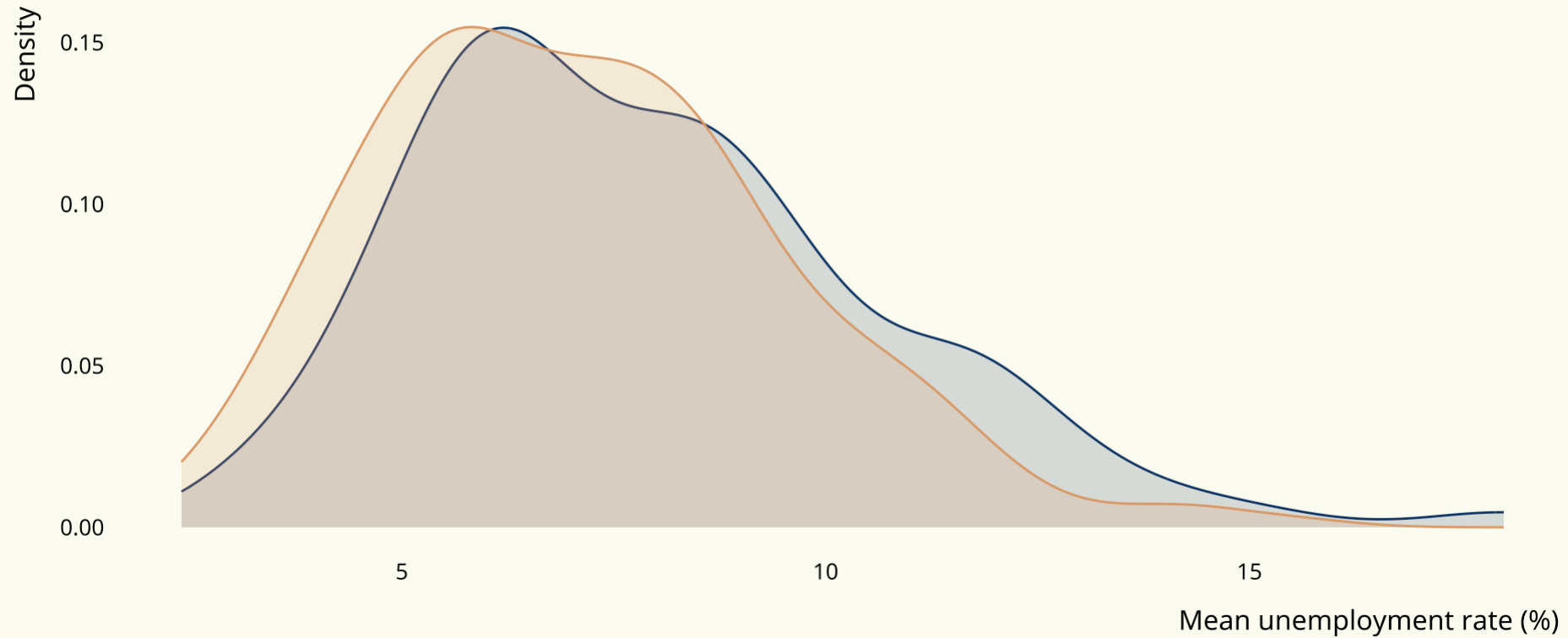
Spirits consumption

Table

Code

Balance plot for unemp

■ Jail penalty ■ No jail penalty



Use to test validity

- Identifying assumptions are usually **not testable**
- But most have **testable implications**, and those are usually **visual**
- The graph shows the raw data underlying the claim
- But important to combine it with a formal statistical test

Graphs and identifying assumptions

Find an identification strategy

- The setting calls for a TWFE approach (staggered roll-out)
- Identifying **assumptions**?
- **Threats** to identification?
 - Trends in number of fatalities before adoption
 - Other shocks at the time of adoption
 - Anticipation: decrease in fatalities before the implementation
 - Change in composition of the states
 - Spillover effects
- How to explore these? Are graphs helpful?

Check: who is treated, and when

Roll-out

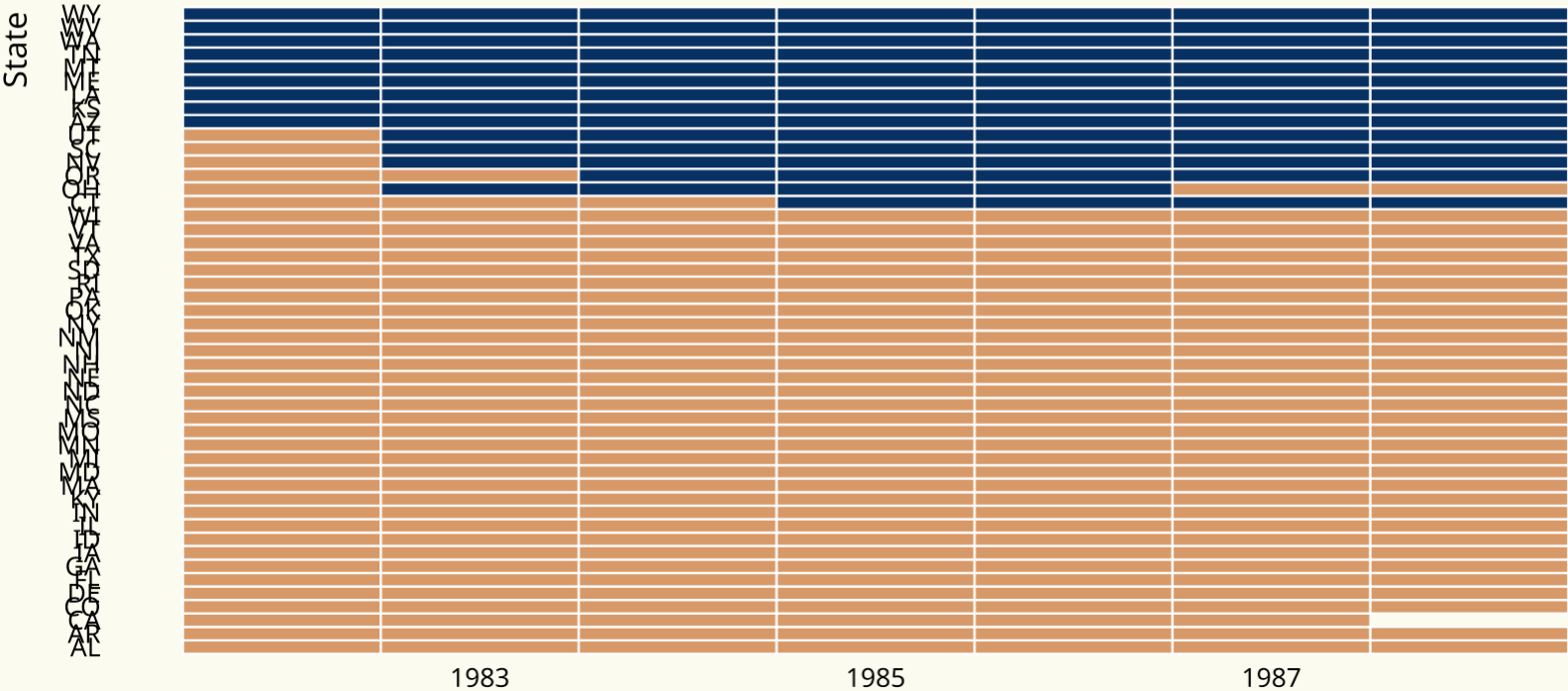
Who identifies it?

Code

Only 6 of the 48 states ever change their law

Mandatory jail sentence for drunk driving, by state and year

Jail penalty No jail penalty

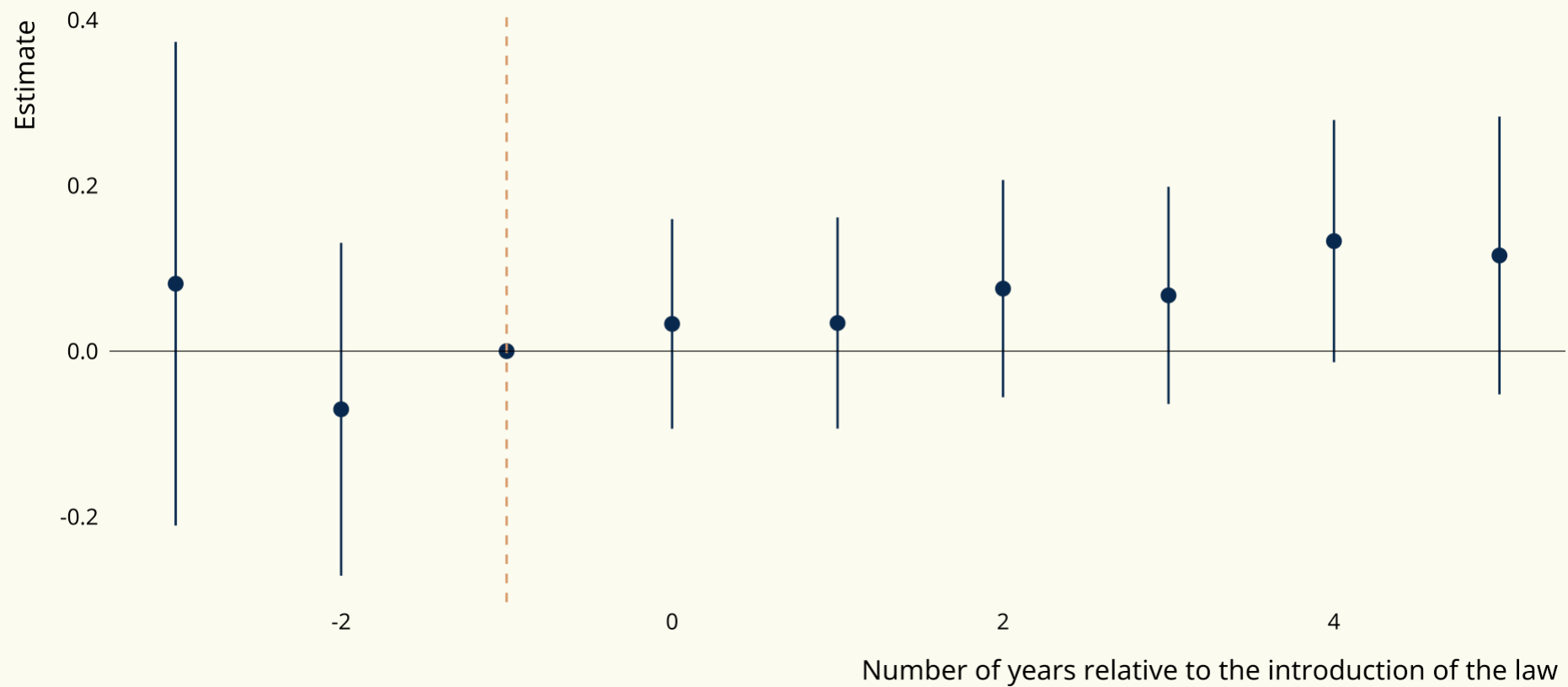


Identifying assumptions: parallel trends

Event study graph

Code

Event study graph: law and impact on fatalities



Estimation

- Let's estimate a couple of models
- Three specifications: raw, with fixed effects, with controls

```
1 reg_raw <- feols(log(fatal) ~ jail, data = fatalities)
2
3 reg_fe <- feols(log(fatal) ~ jail | state + year, data = fatalities)
4
5 reg_ctrl <- feols(
6   log(fatal) ~ jail + log(unemp) + log(income) | state + year,
7   data = fatalities,
8   cluster = ~state
9 )
```

- How do we **communicate** them?

Example table

TABLE 6—THE BIRTH OF A NATION AND DISCRIMINATION AGAINST AFRICAN AMERICANS

	Supremacism			Discrimination		
	OLS (1)	RF (2)	2SLS (3)	OLS (4)	RF (5)	2SLS (6)
<i>Panel A. Equation (1)</i>						
<i>Birth of a Nation</i>	0.036 (0.005)		0.093 (0.014)	0.037 (0.005)		0.093 (0.015)
<i>Million Dollar Mystery</i>		0.035 (0.005)			0.035 (0.006)	
First-stage <i>F</i> -statistic	—	—	218	—	—	223
<i>Panel B. State × month fixed effects</i>						
<i>Birth of a Nation</i>	0.037 (0.006)		0.102 (0.018)	0.031 (0.006)		0.070 (0.020)
<i>Million Dollar Mystery</i>		0.031 (0.006)			0.021 (0.006)	
First-stage <i>F</i> -statistic	—	—	125	—	—	126
Dep. var. mean	0.11	0.11	0.11	0.12	0.12	0.12
Dep. var. st. dev.	0.31	0.31	0.31	0.32	0.32	0.32
Observations	89,325	89,325	89,325	89,325	89,325	89,325

Notes: The table reports OLS (columns 1 and 4), reduced form (columns 2 and 5) and 2SLS (columns 3 and 6) estimates. The dependent variable is an indicator variable for the presence of White supremacism in the public discourse ($Supremacism_{c,t}$) in columns 1–3, and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in columns 4–6. See Section VI for further details. The unit of observation is the county (c) in the month-year (t). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section IC for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section IIB for further details). All regressions control for county and month-year. Results in columns 1–3 include percentile fixed effects based on the frequency of pages that contain the word *Americans* in the corresponding county-month. Results in columns 3–6 include frequency of job ads percentile fixed effects. Results presented in panel B also include fixed effects at the state-period (month) level. The reported standard deviation and number of observation refer to estimates in panel A. Standard errors are clustered at the county level.

- Explicit **title**
- **Column headers** name the outcome and what changes across specifications
- **Coefficient block:** only the interpreted estimates
- **Bottom block:** N , fixed effects, controls, mean of the outcome
- **Notes** give units, sample, and what the SE are clustered on

Notable points

- Self contained can read it **without the paper**
- We can understand the **magnitude** of the effect
- We can see what changes across columns
- It shows only the coefficients the authors discuss
- Numbers are rounded

To include in a table

- Coefficients and their associated standard error
- Number of observation the number of clusters
- **Mean (and SD) of the outcome** to get a sense of the magnitude of the effect
- Yes/no rows for each set of FEs and each set of controls
- Column headers that name the outcome or the specification
- Clean variable names
- An **informative title**
- Notes that make the table **self-contained**

Not to include in a table

- Coefficients on **controls we never interpret**
- The intercept, almost always
- Some goodness-of-fit statistics
- **Six decimals** on a coefficient whose standard error is 0.11
- **Too many zeros** for a coef (eg 0.00001) → change unit
- Vertical lines, boxes, shading, and color for its own sake

Default table

Default output

Problems

Code

	(1)	(2)	(3)
(Intercept)	6.506 (0.059)		
jailTRUE	-0.303 (0.111)	0.025 (0.034)	0.004 (0.051)
log(unemp)			-0.132 (0.053)
log(income)			1.075 (0.341)
Num.Obs.	335	335	335
R2	0.022	0.992	0.995
R2 Adj.	0.019	0.991	0.993
R2 Within		0.002	0.308
R2 Within Adj.		-0.002	0.300
AIC	893.9	-617.9	-736.5
BIC	901.5	-408.1	-519.1
RMSE	0.91	0.08	0.07
FE: state		X	X
FE: year		X	X

A better table

Table	Formatting	Other packages	Code
Outcome: log of annual traffic fatalities.			
	No controls	State & year FE	FE + controls
State with a mandatory jail sentence	-0.303 (0.111)	0.025 (0.034)	0.004 (0.051)
Unemployment rate (log)			-0.132 (0.053)
Income per capita (log)			1.075 (0.341)
Observations	335	335	335
Mean fatalities	915	915	915
Clusters (states)	48	48	48
State fixed effects		X	X
Year fixed effects		X	X

Notes: OLS estimates of the association between a mandatory jail sentence for drunk driving and traffic fatalities. The outcome is the log of the annual number of traffic fatalities in a state. The sample covers the 48 contiguous US states, observed annually from 1982 to 1988. Column (1) includes no controls. Column (2) adds state and year fixed effects. Column (3) further controls for the log unemployment rate and log income per capita. Mean fatalities is the average annual number of traffic fatalities per state over the period. Standard errors, clustered at the state level, are reported in parentheses.

Visualizing estimation output

Coef plot

Standardized

Why standardize?

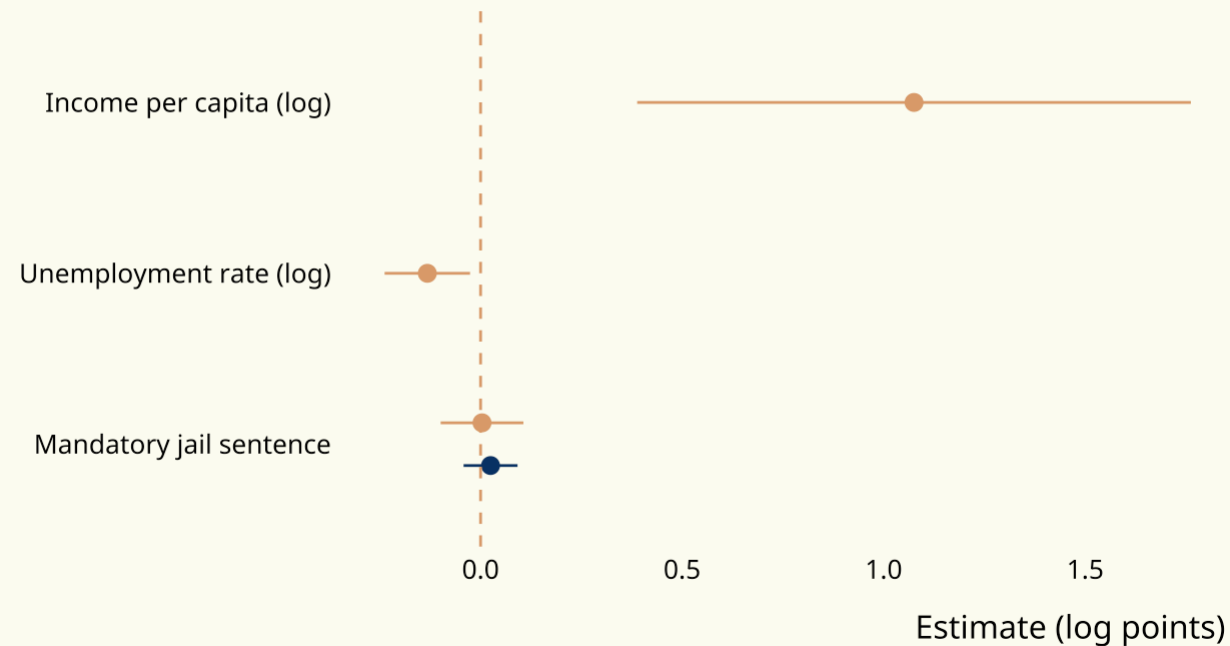
Distribution

Code

Regression coefficients of log traffic fatalities

Dot-whisker plot, 95% confidence intervals, FE model

• No controls • With controls



R package: modelsummary

Use for diagnostic plots

- Residuals against fitted values, QQ plots
- **Influence and leverage**: is one observation carrying the whole result?
- **Overlap** and common support before matching
- Distribution of **implicit weights**
- Audience of diagnostic plots: most often you

Diagnostic plots

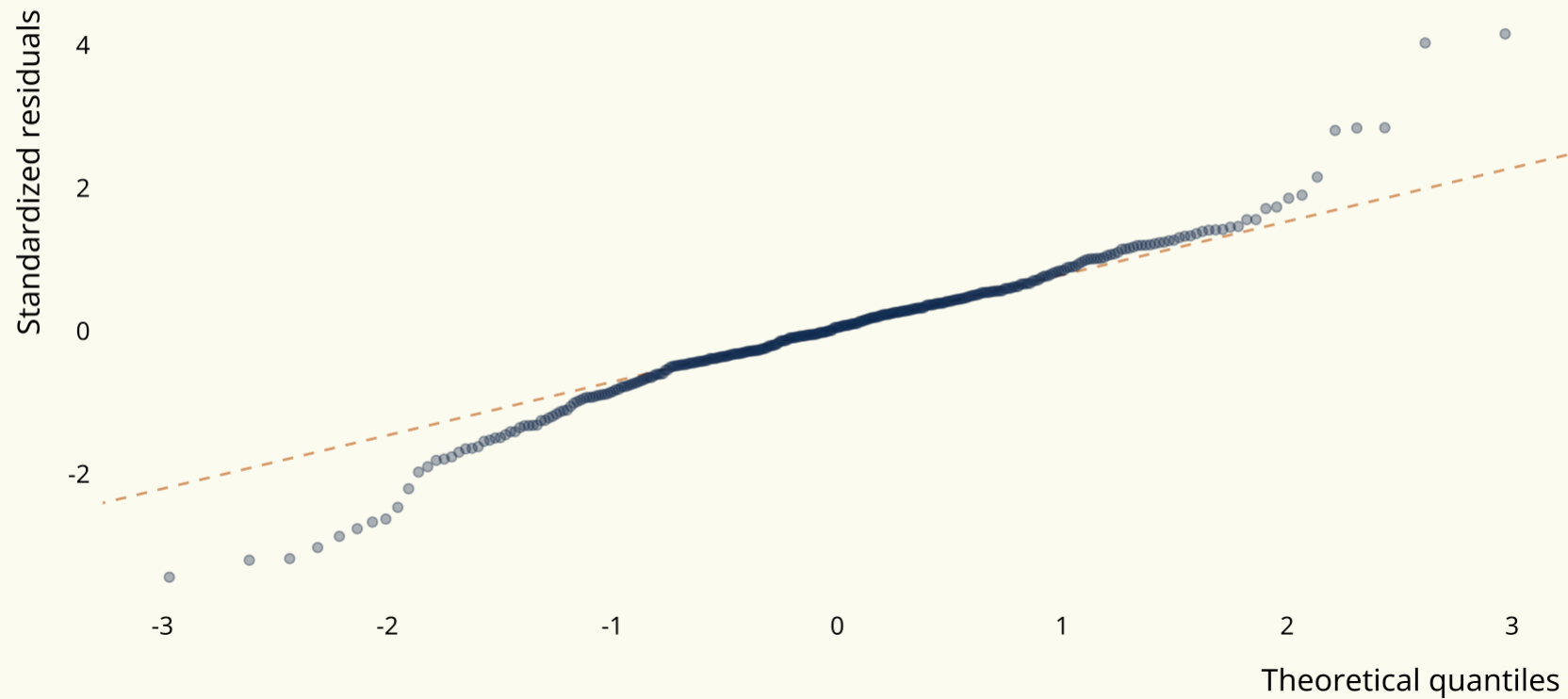
Normal Q-Q

Res. vs fitted

Code

Normal Q-Q

Look for: departures in the tails (here: fat lower tail)



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?

i Take-away messages

1. Identify **message**
2. Identify **audience** and objective
3. Find the right **type of graph** for your story and audience
4. Keep **cognitive load** in mind
5. Think about **aesthetics**
6. Follow “**good-practice**” rules

Graphs: concrete take-aways

Concrete take-away messages

- Overall, facilitate the retrieval of information and **reduce cognitive load**
- 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.

Tables: take-aways

i Take-away messages

- One table, **one question**. Name the outcome and the treatment in words
- Report the coefficient you interpret, its **uncertainty**, N , clusters and the **mean of the outcome**
- Yes/No rows for FE and controls: the reader must see **what changes across columns**
- Round to what you can defend; delete everything you will not discuss
- Write the notes so the table **stands alone**
- Build it from **code**, never by hand

Thanks