

Data Science II

- Guidelines for Data Visualizations -



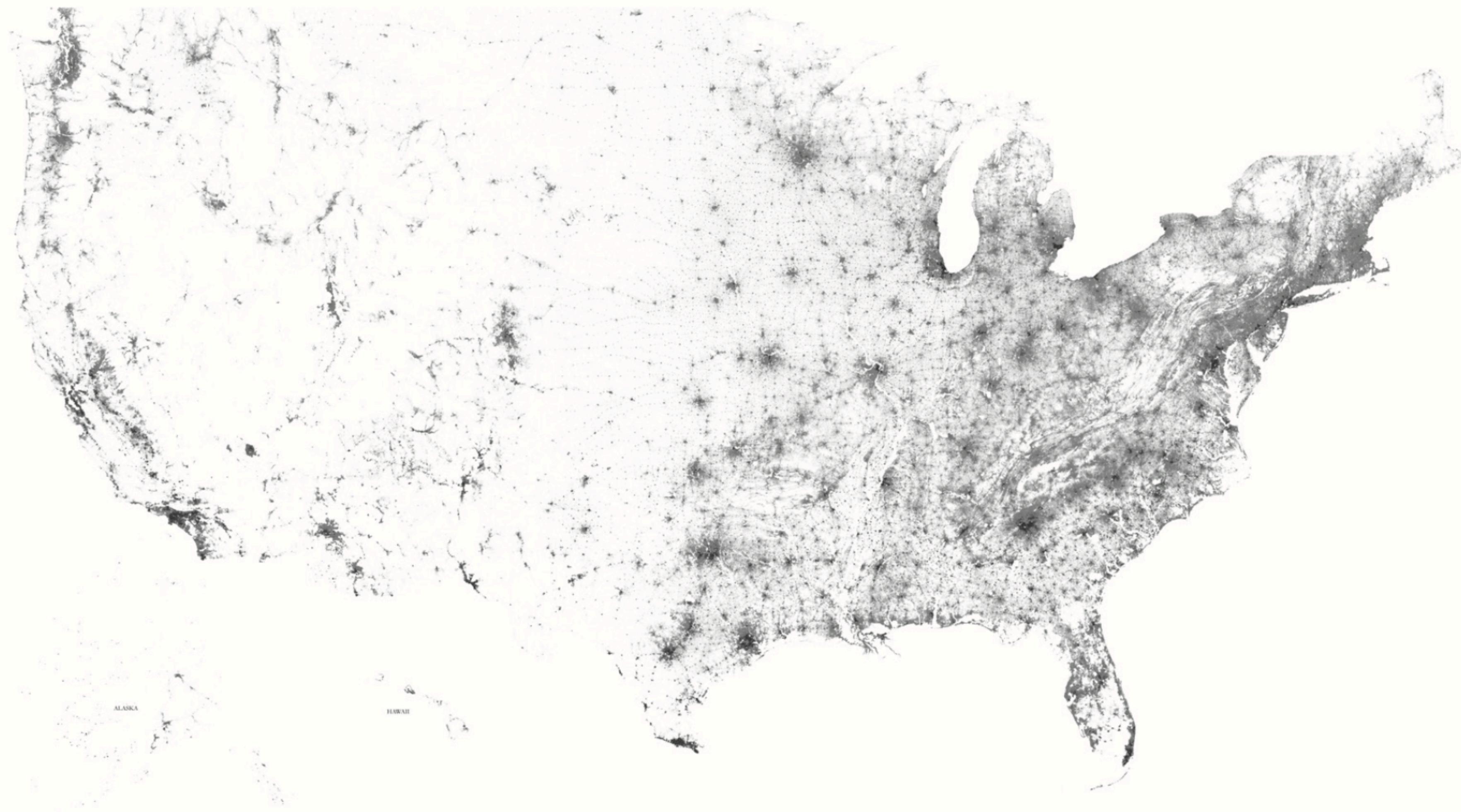
Prof. Dr. Eduard Kromer
Summer Semester 2024
University of Applied Sciences Landshut

Show the data

Show the data

- you don't have to show *all* the data, but you should highlight the values that are important to your argument
- the challenge: **how much** data to show and the **best way** to show it

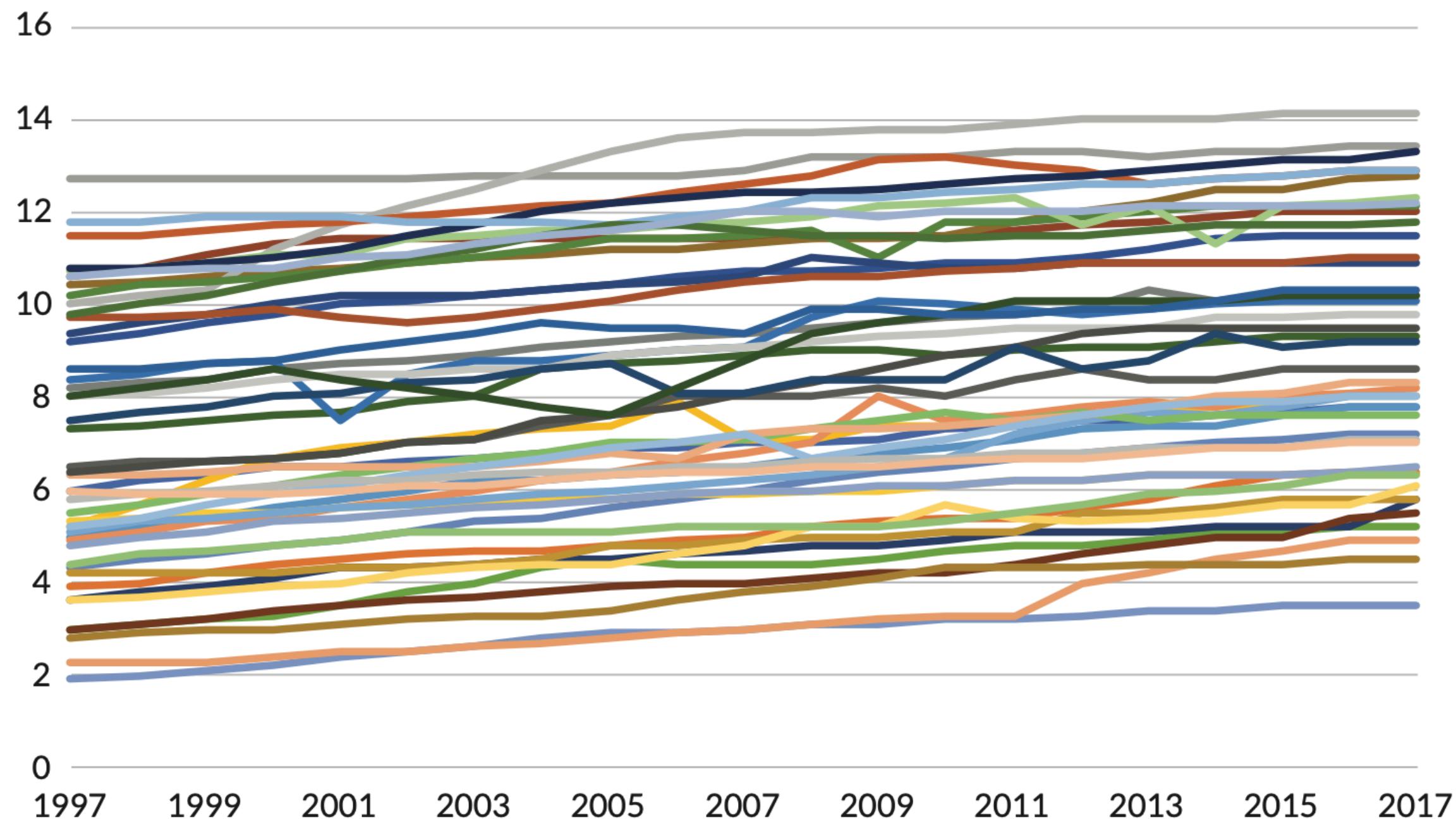
Show only the data



Show the data that matter most

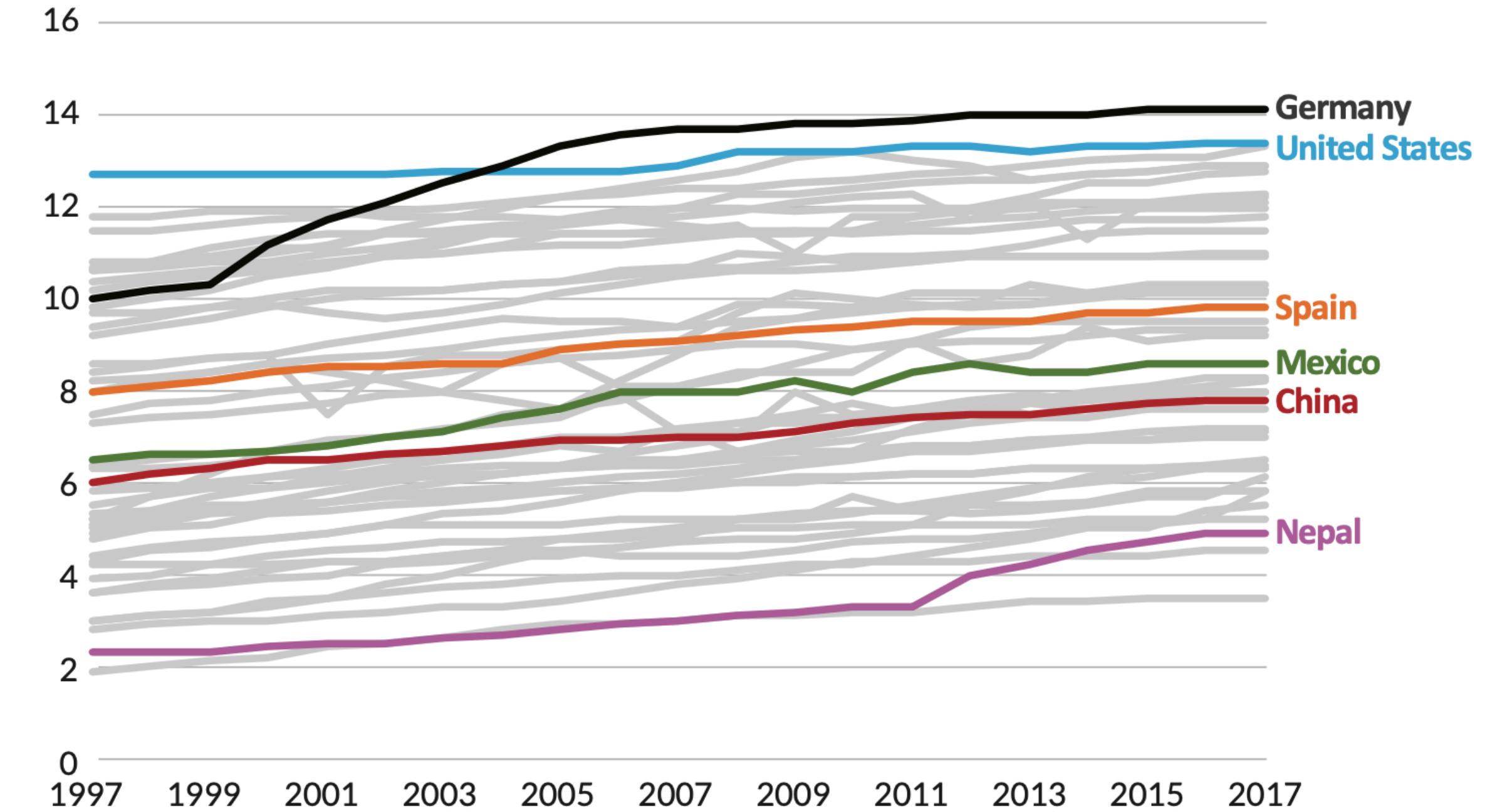
Average years of schooling has increased around the world

(Number of years)



Average years of schooling has increased around the world

(Number of years)

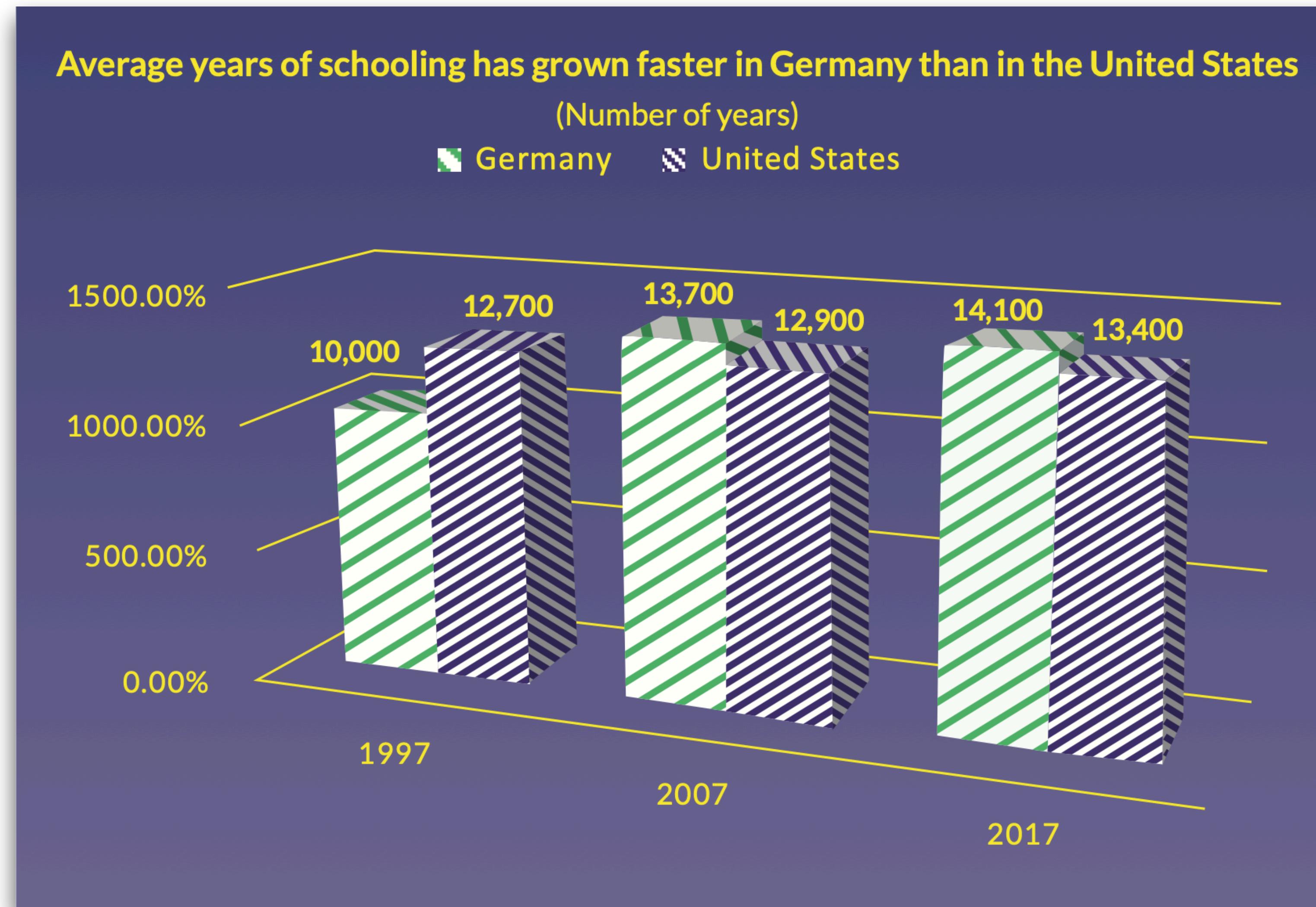


Reduce the clutter

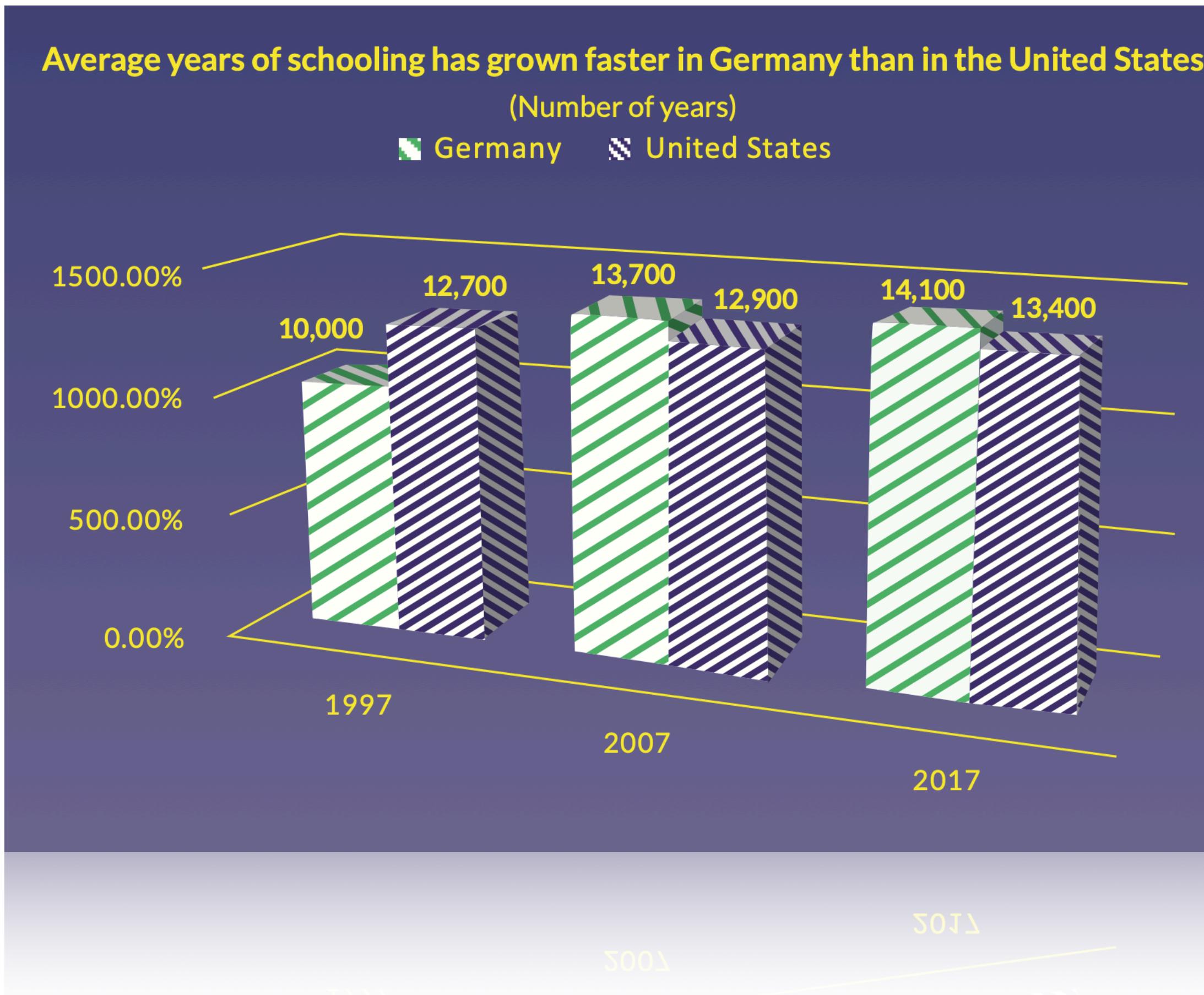
Reduce the clutter

- unnecessary visual elements distract your reader from the story you want to tell with data
- many different types of chart clutter you should avoid
 - e.g. basic elements, like heavy tick marks and gridlines should be removed in almost every case
 - overlapping markers
 - textured or filled gradients, when simple, solid shades of color are better

Reduce the clutter

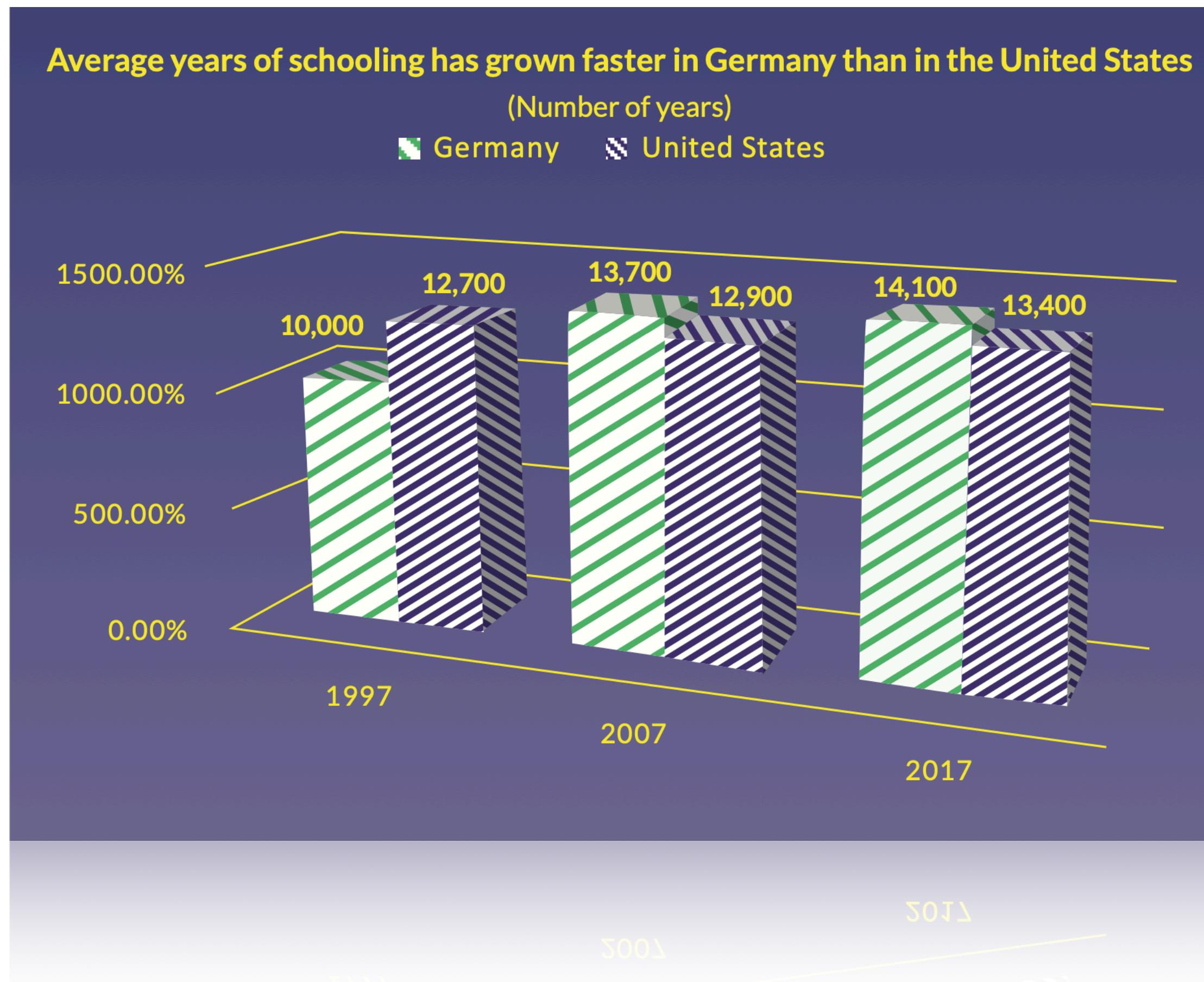


Reduce the clutter



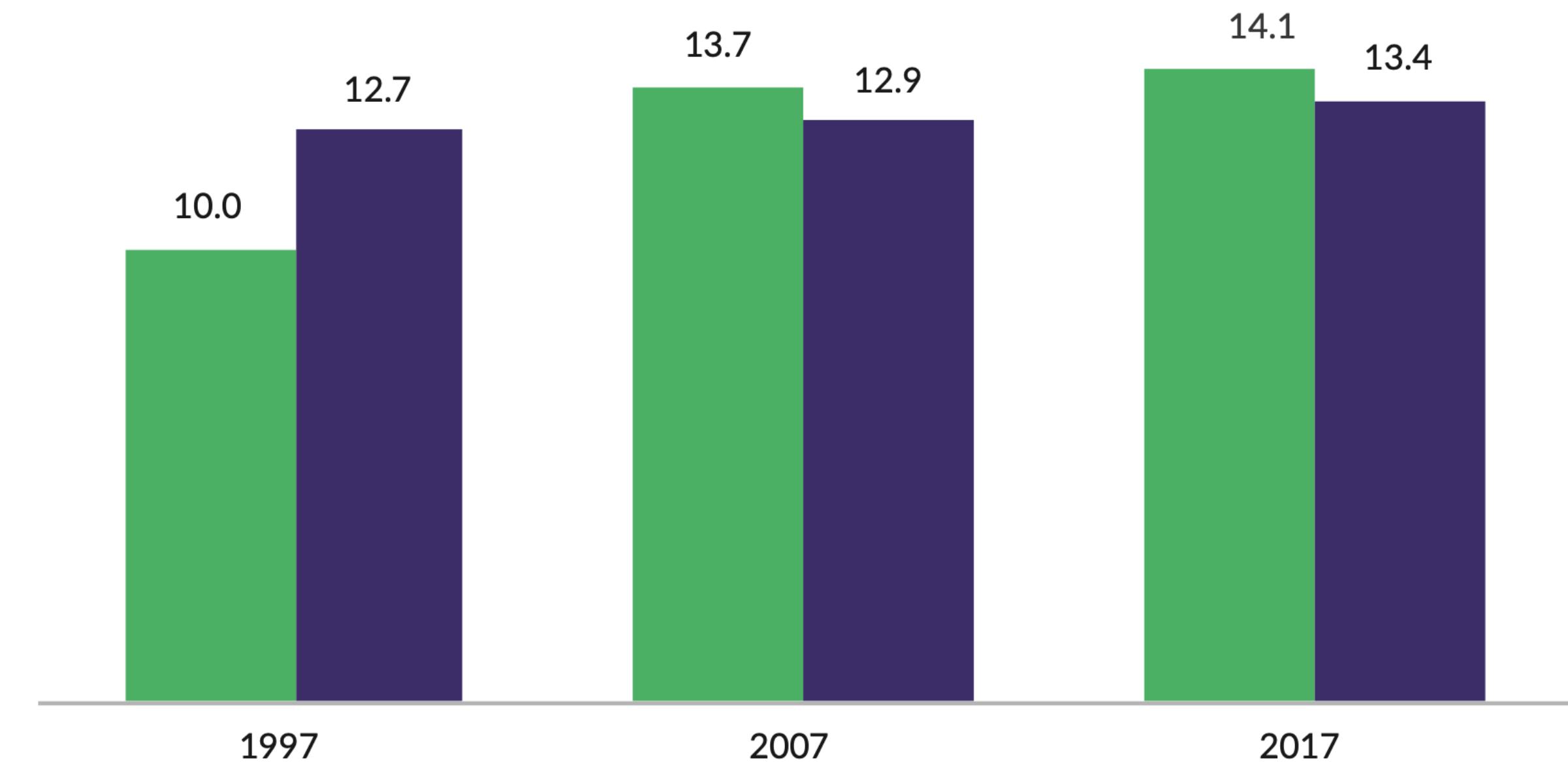
- the three-dimensional view distorts the data
- distortion occurs because the unnecessary third dimension requires adding perspective to the graph

Simplify the graph



Average years of schooling has grown faster in Germany than in the United States
(Number of years)

Germany United States



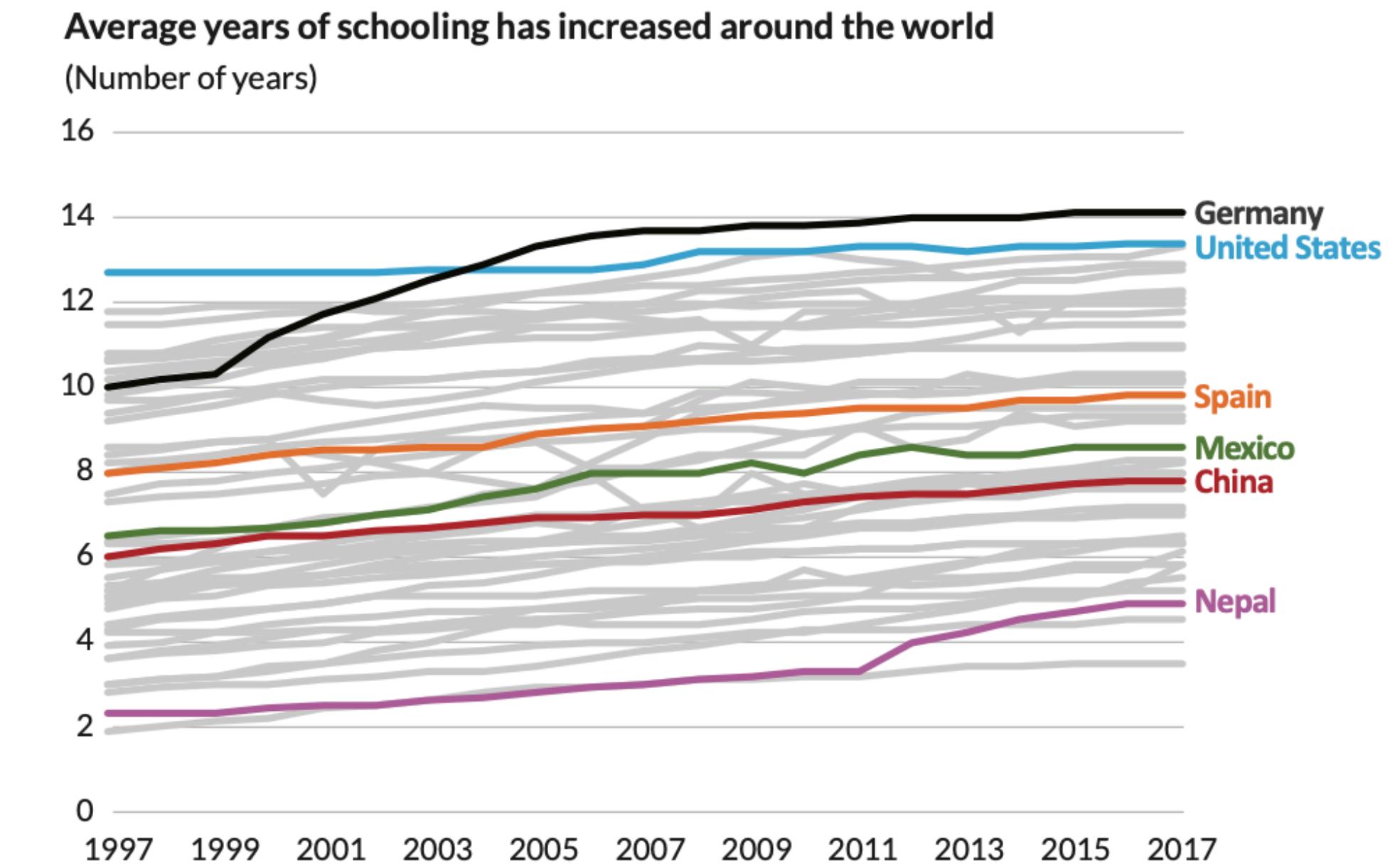
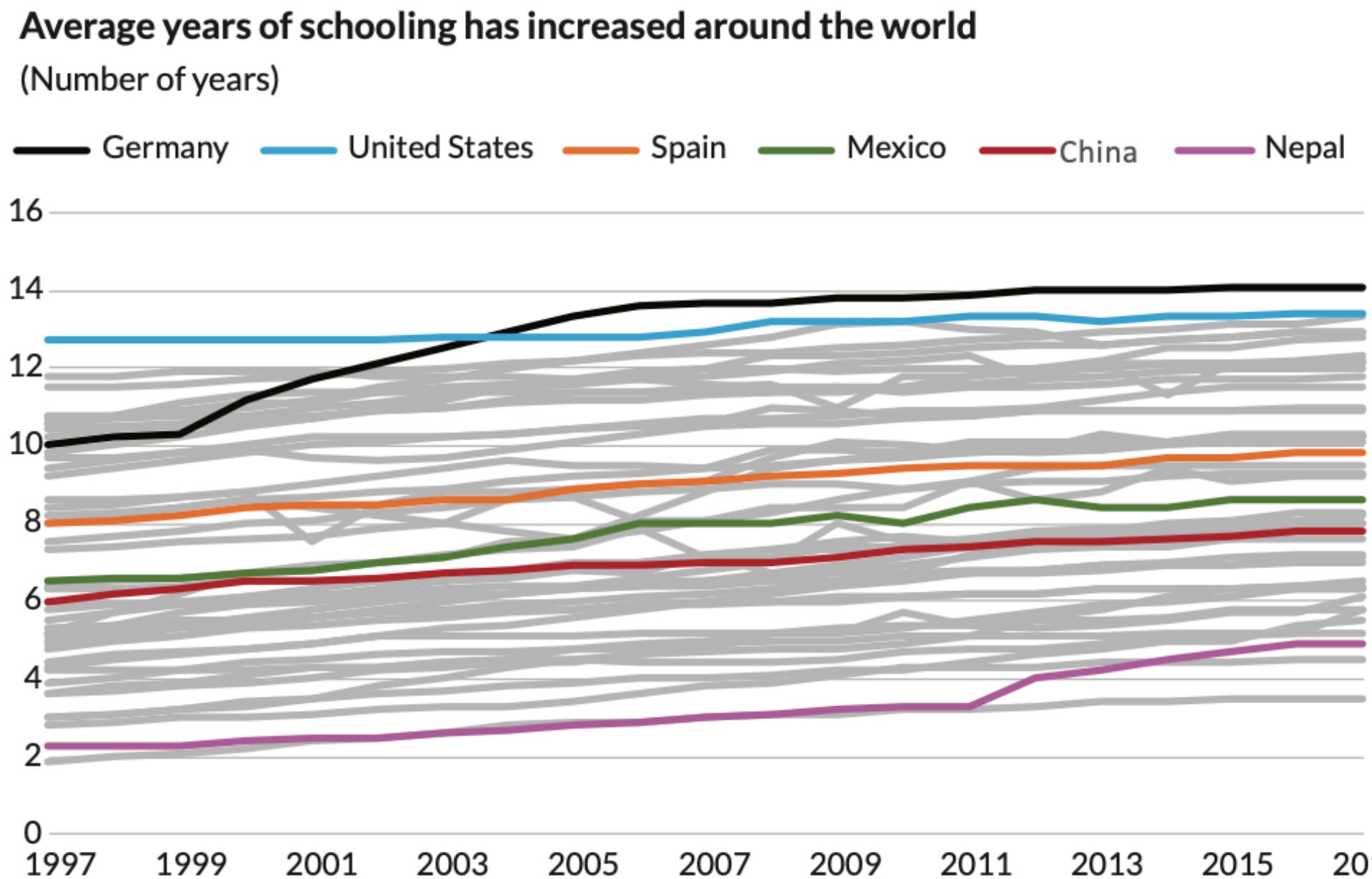
Integrate Graphics and Text

Integrate Graphics and Text

- the text we include in and around our graphs is equally important to the graphic elements in our visualizations
- text can be used to explain how to read the content in the graph and / or the graph itself

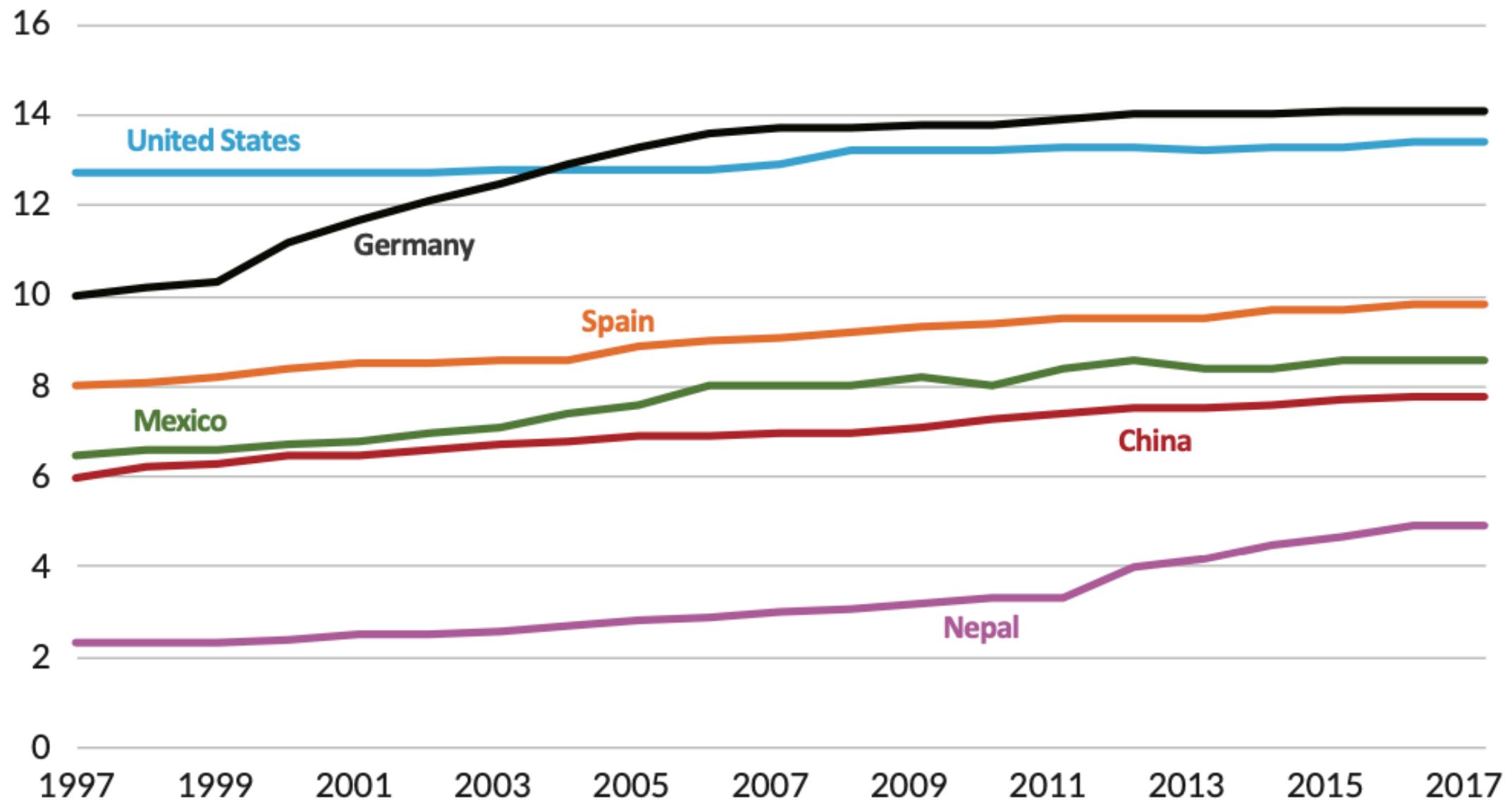
Remove legends when possible and label data directly

- many software tools default to create a data legend and place it around the chart, disconnected from the data
- this forces **more work upon the reader** to connect the information

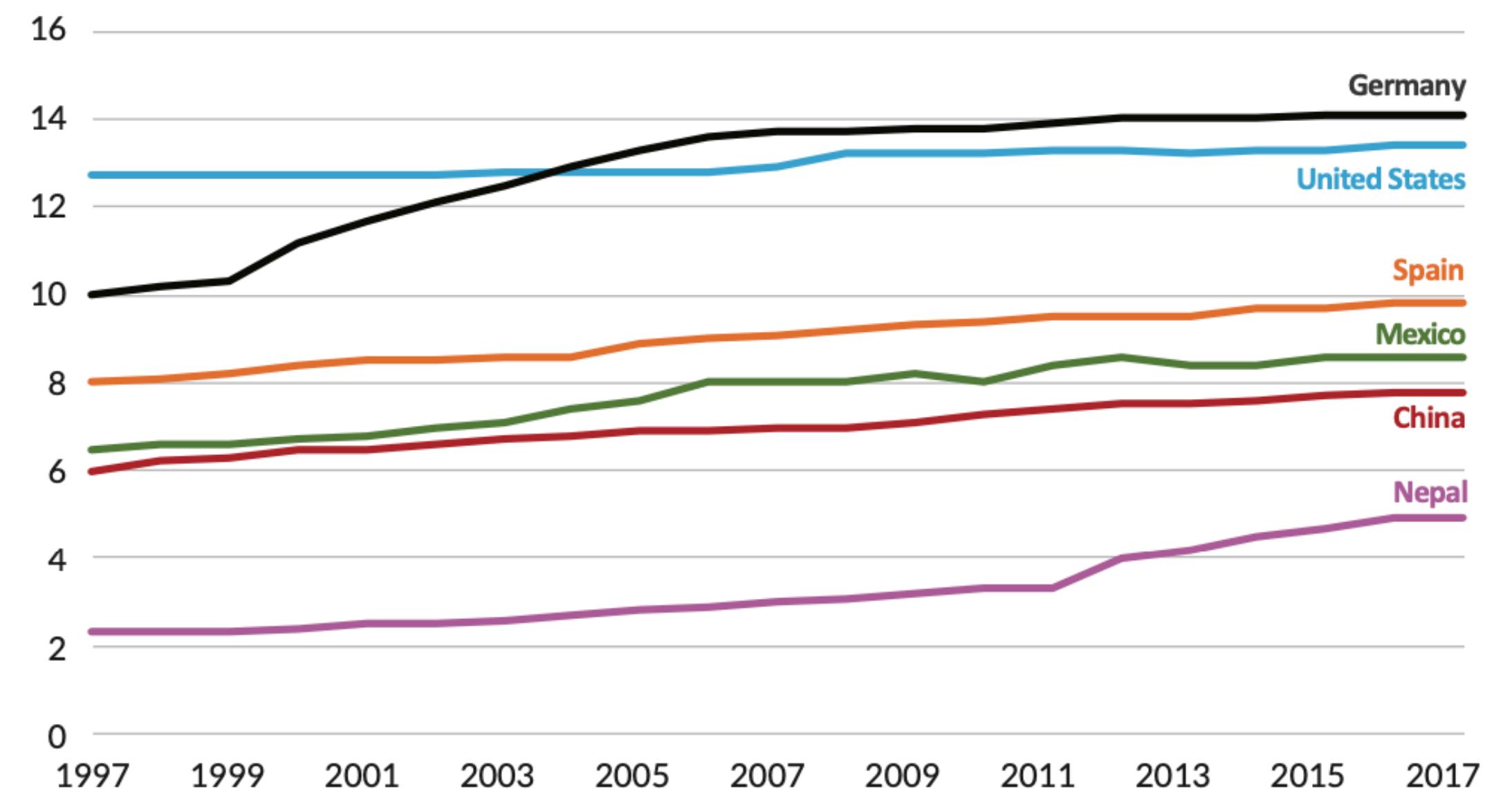


Align the labels and match colors

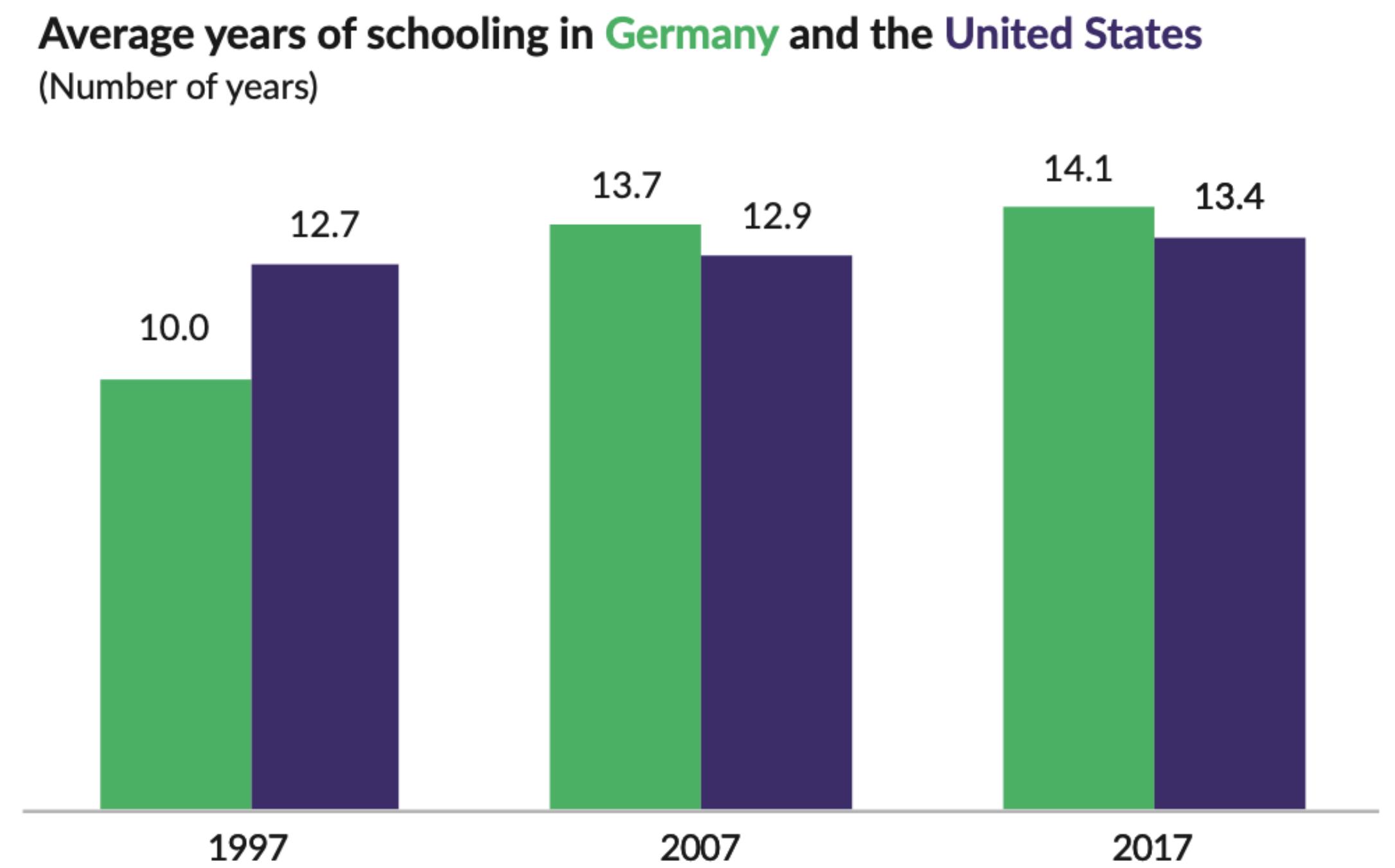
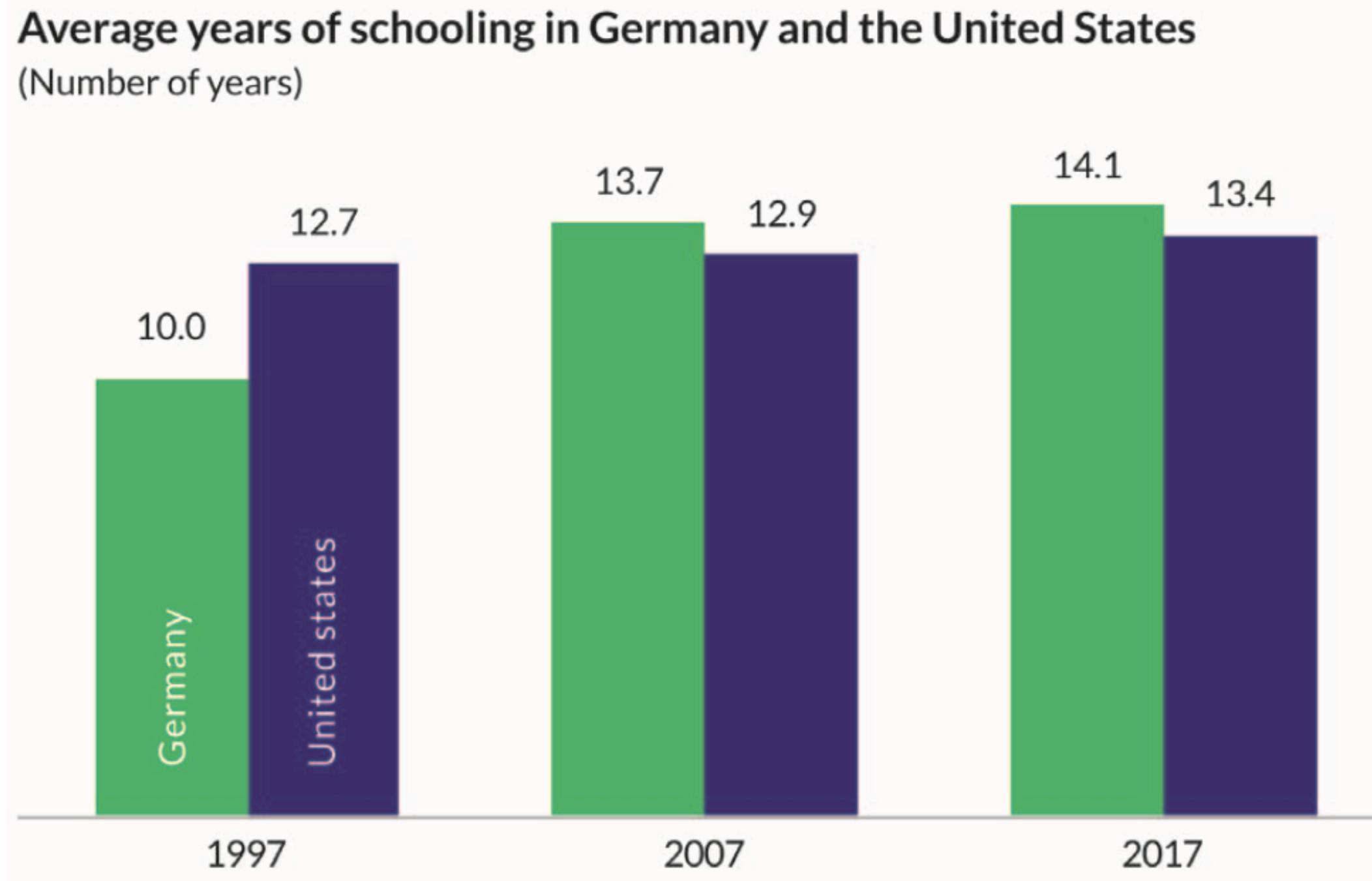
Average years of schooling has increased around the world
(Number of years)



Average years of schooling has increased around the world
(Number of years)

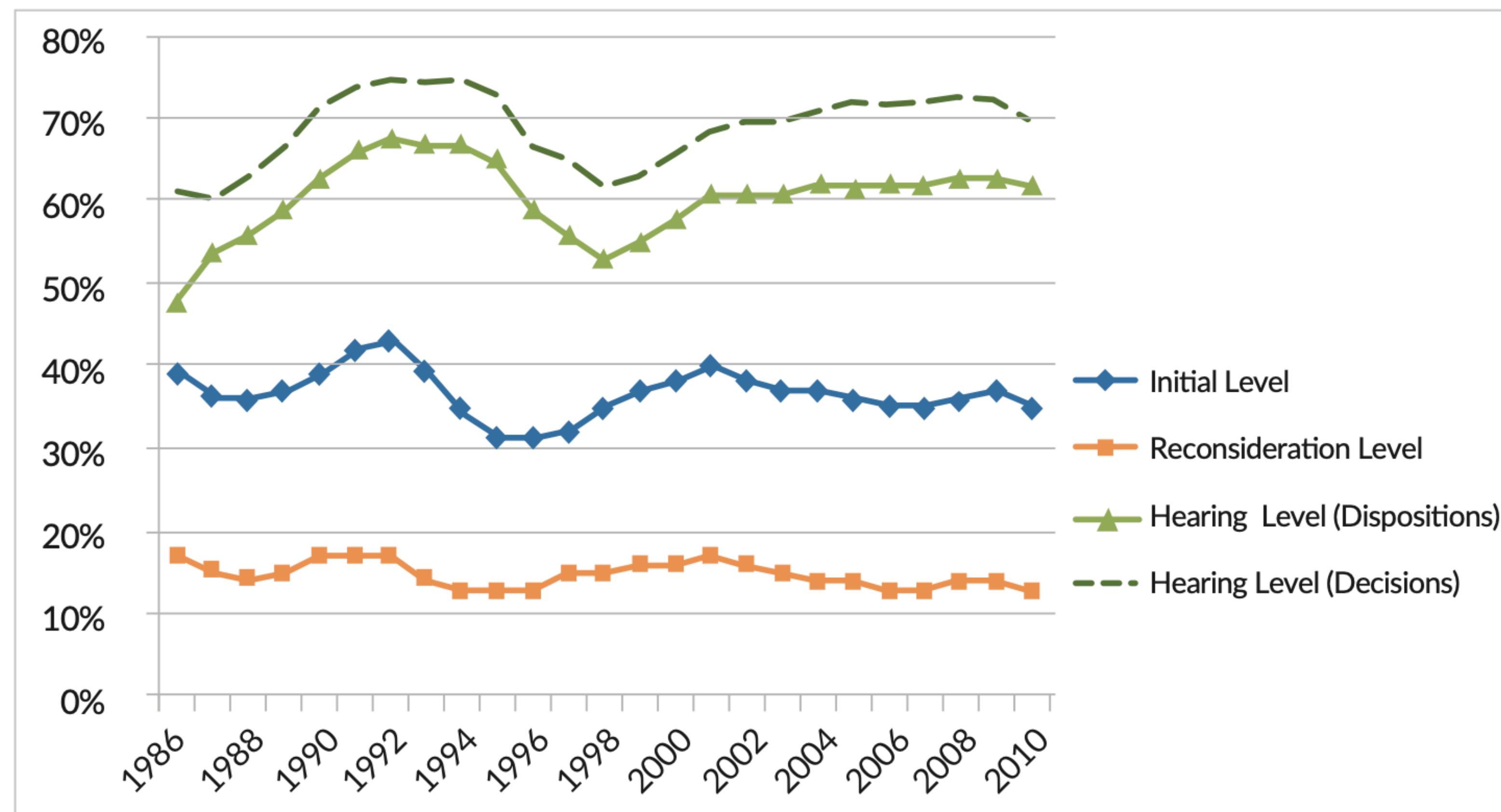


Integrate labels into the graph



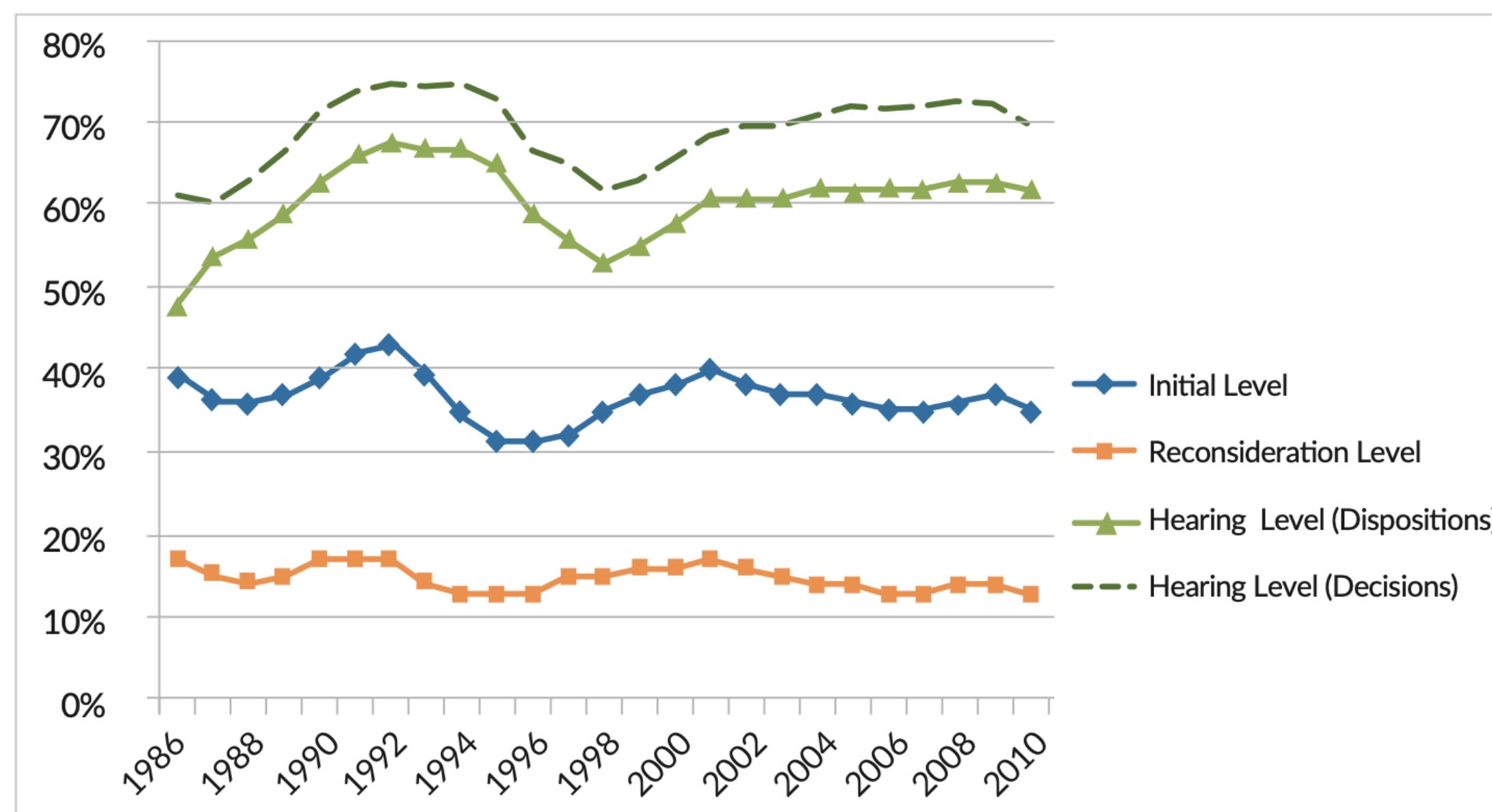
Avoid Inconsistencies

DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)

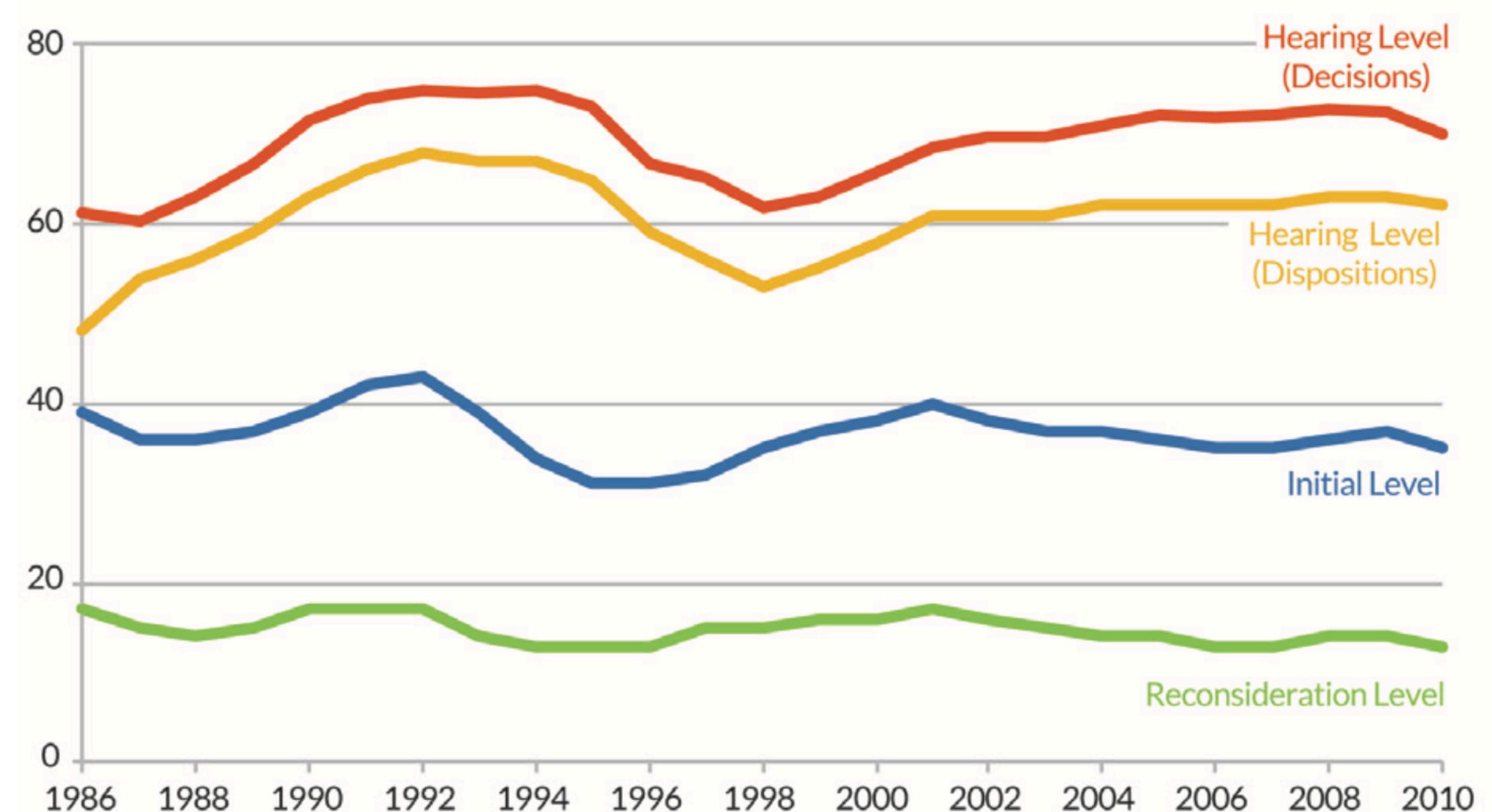


Remove Inconsistencies

DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)



DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)

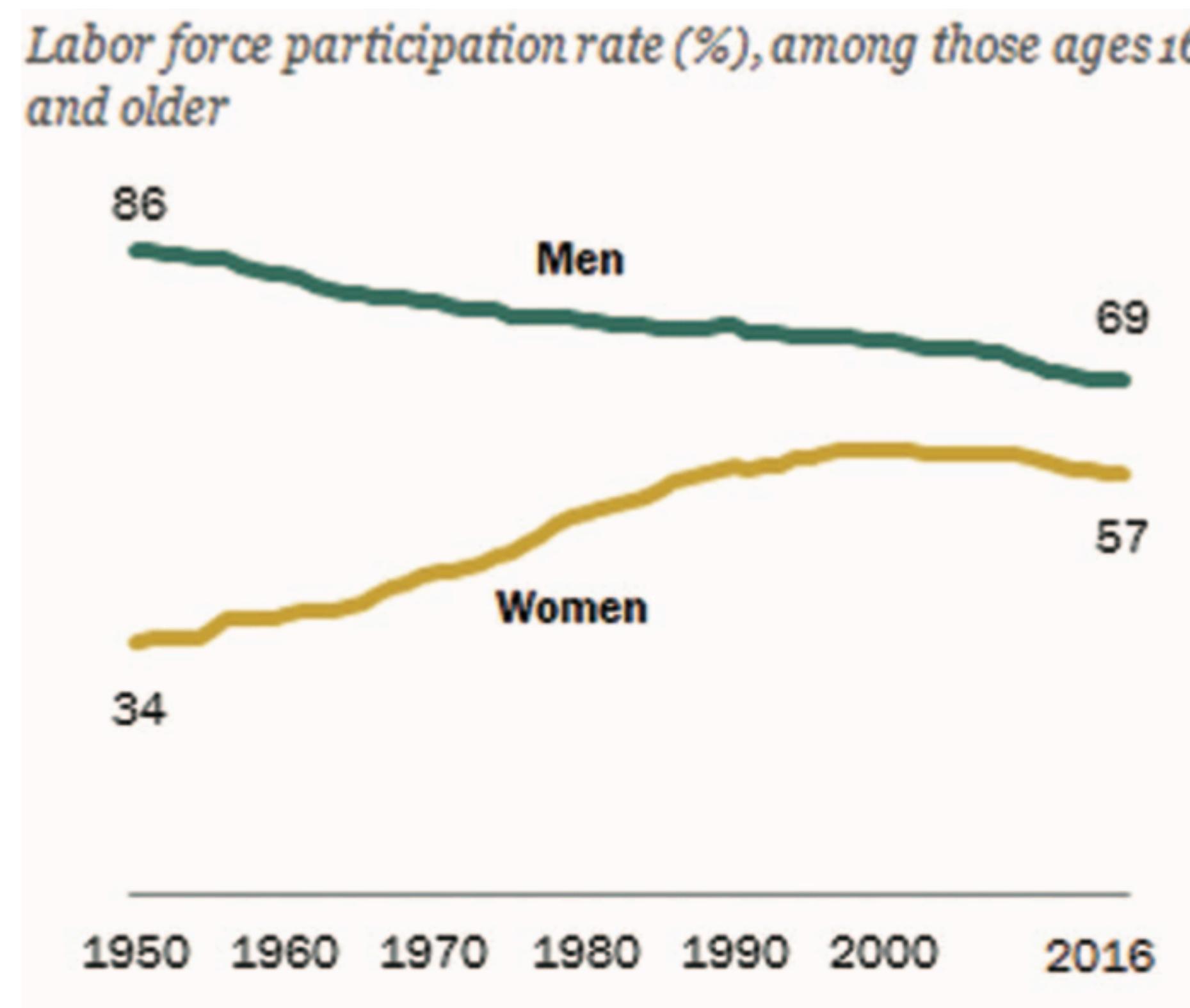


**Write the title like
a newspaper headline**

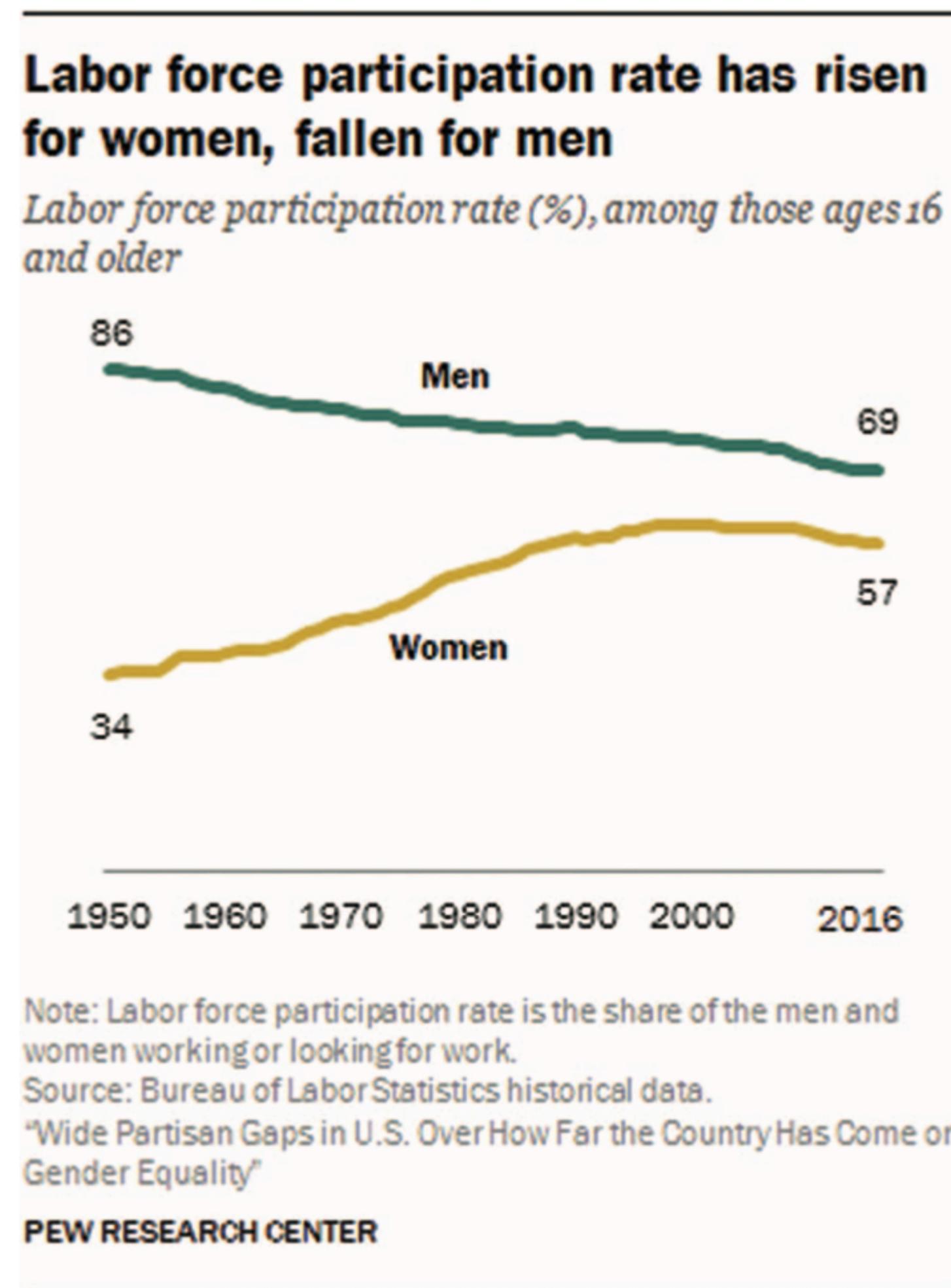
Write the title like a newspaper headline

- better titles capture the takeaway of the chart
 - tell the reader what conclusions can be drawn from the data
 - use concise, active phrases that make it easy to understand what the chart is aiming toward

What should the reader learn from this chart?



Titles and text attract peoples attention



- the active title in this chart tells you what you should learn from it
- you should treat chart titles as carefully as the chart itself

Add explainers

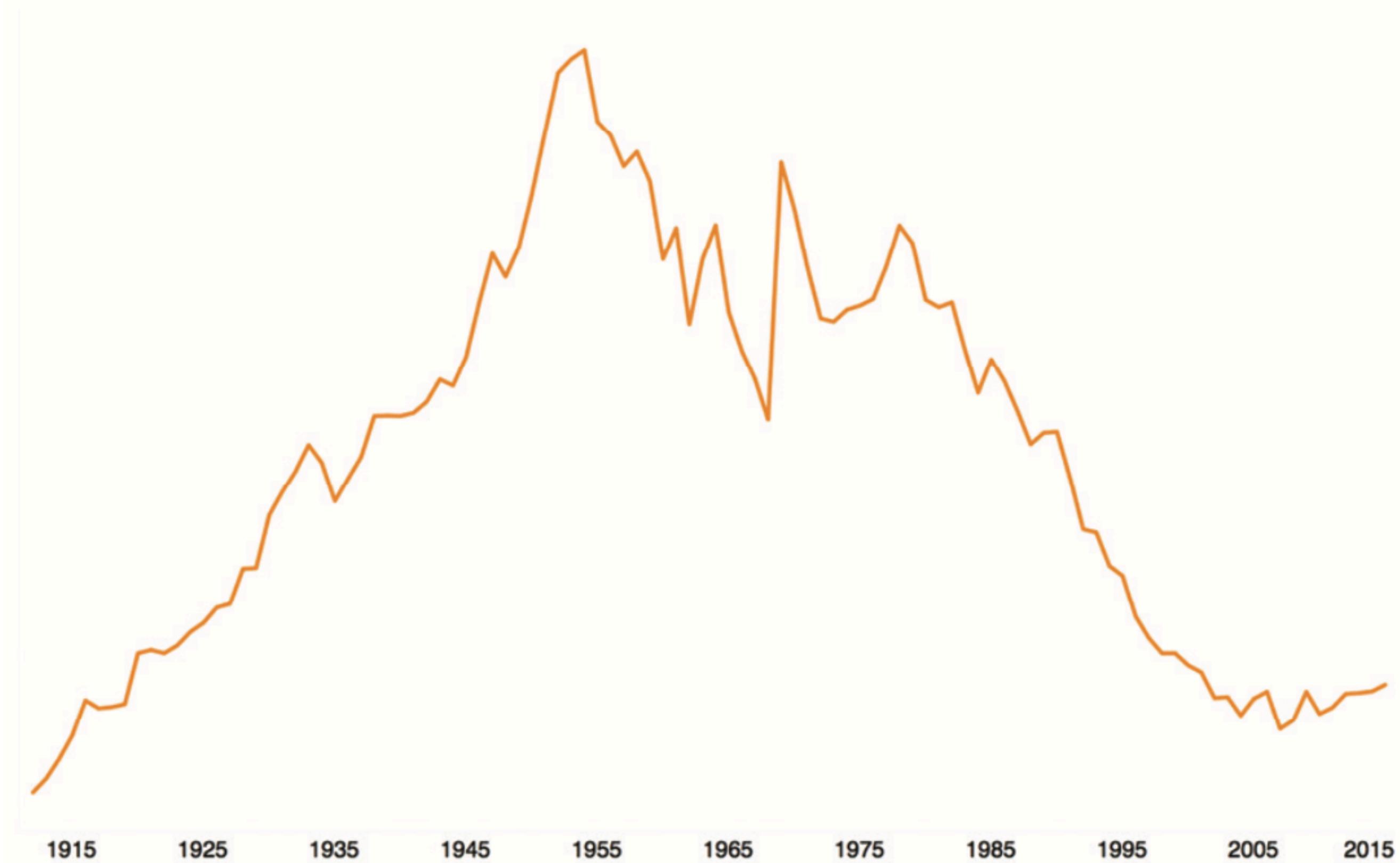
Add explainers

- sometimes data sets have points of interest that bear explanation
 - points of interest: e.g. peaks, valleys, outliers, variations
 - adding detail in graphs can push your argument
- annotation allows (less experienced) readers to grasp and understand content quickly; it can help navigate the format of the graph and its content

Questions?

Rise and Fall of the name **Neil** in the USA
Births 1912-2015

Source: data.gov

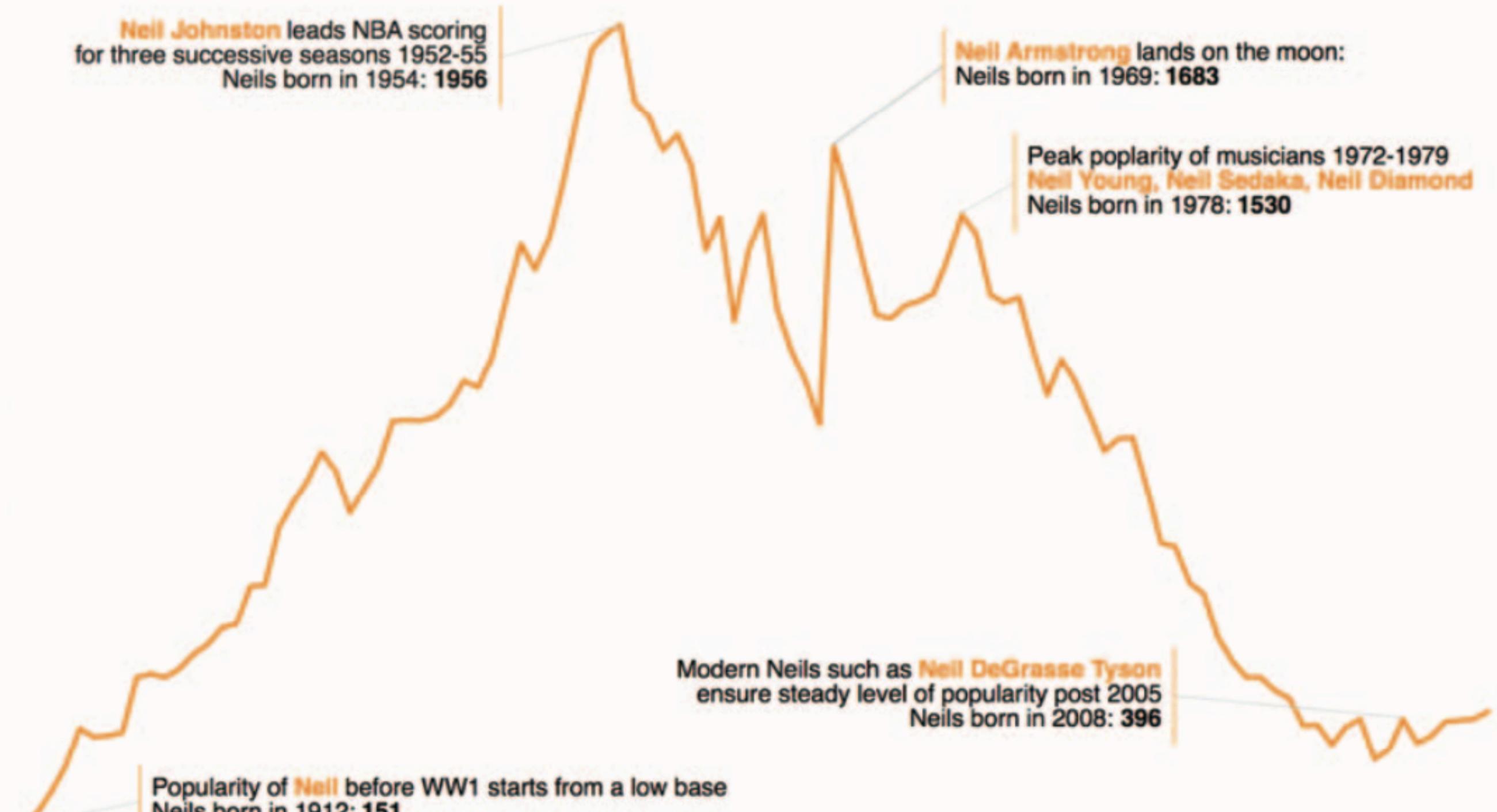


Visualisation: [@theneilrichards](#)

Line Chart with Explainers

Rise and Fall of the name **Neil** in the USA Births 1912-2015

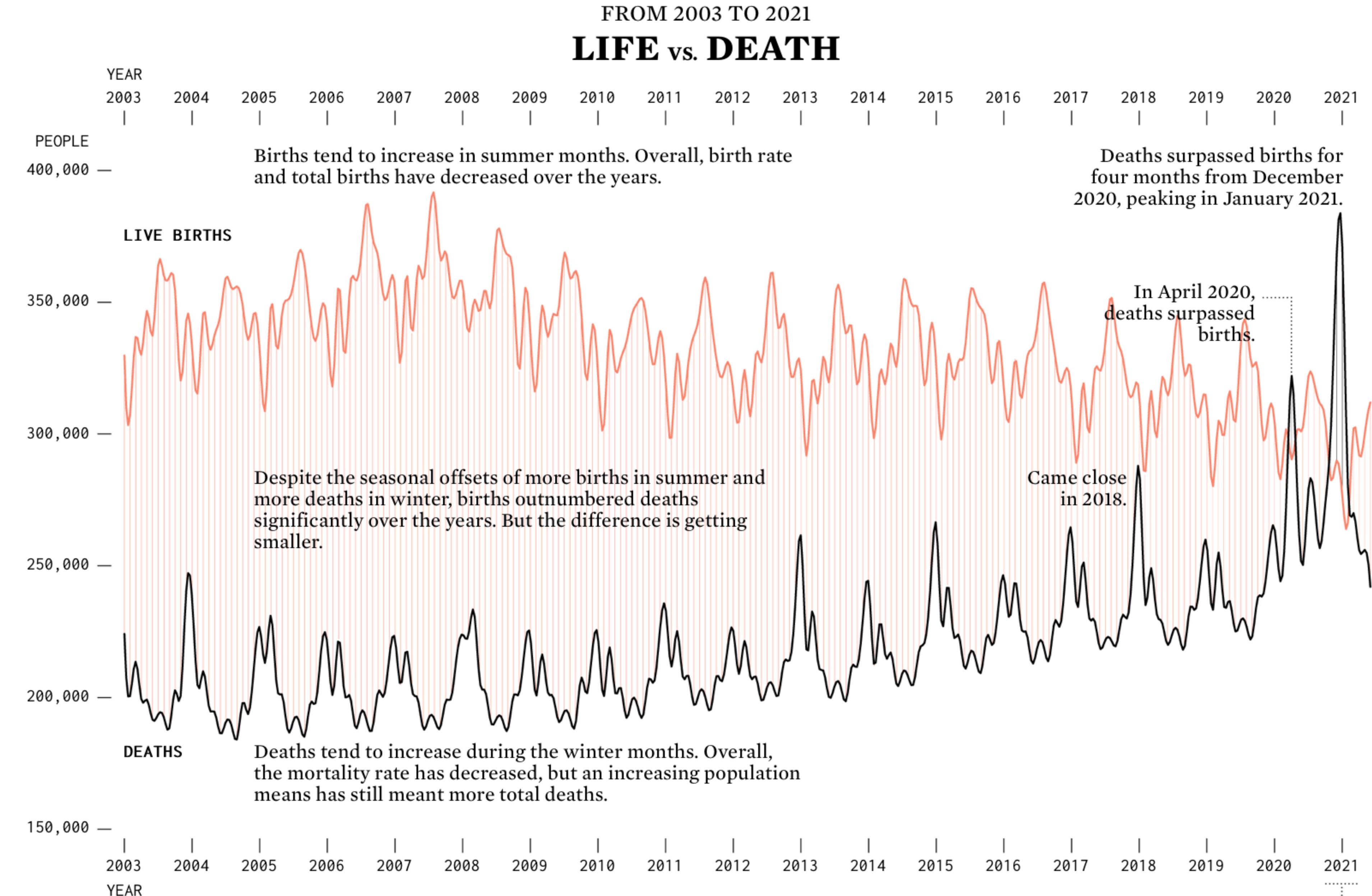
Source: data.gov



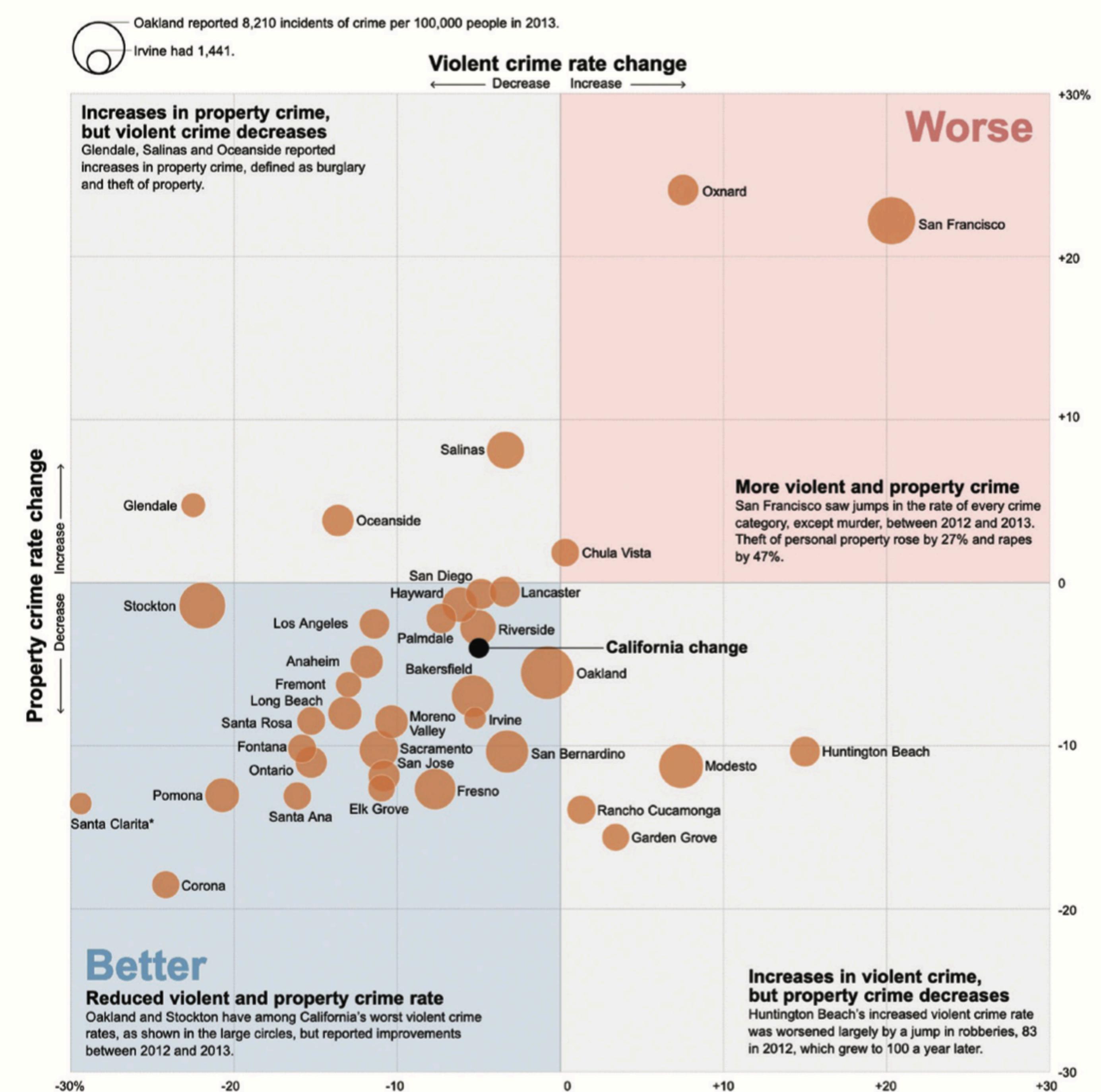
Visualisation: @theneilrichards

#SWDChallenge

Line Chart with Explainers



Explain how to read the graph



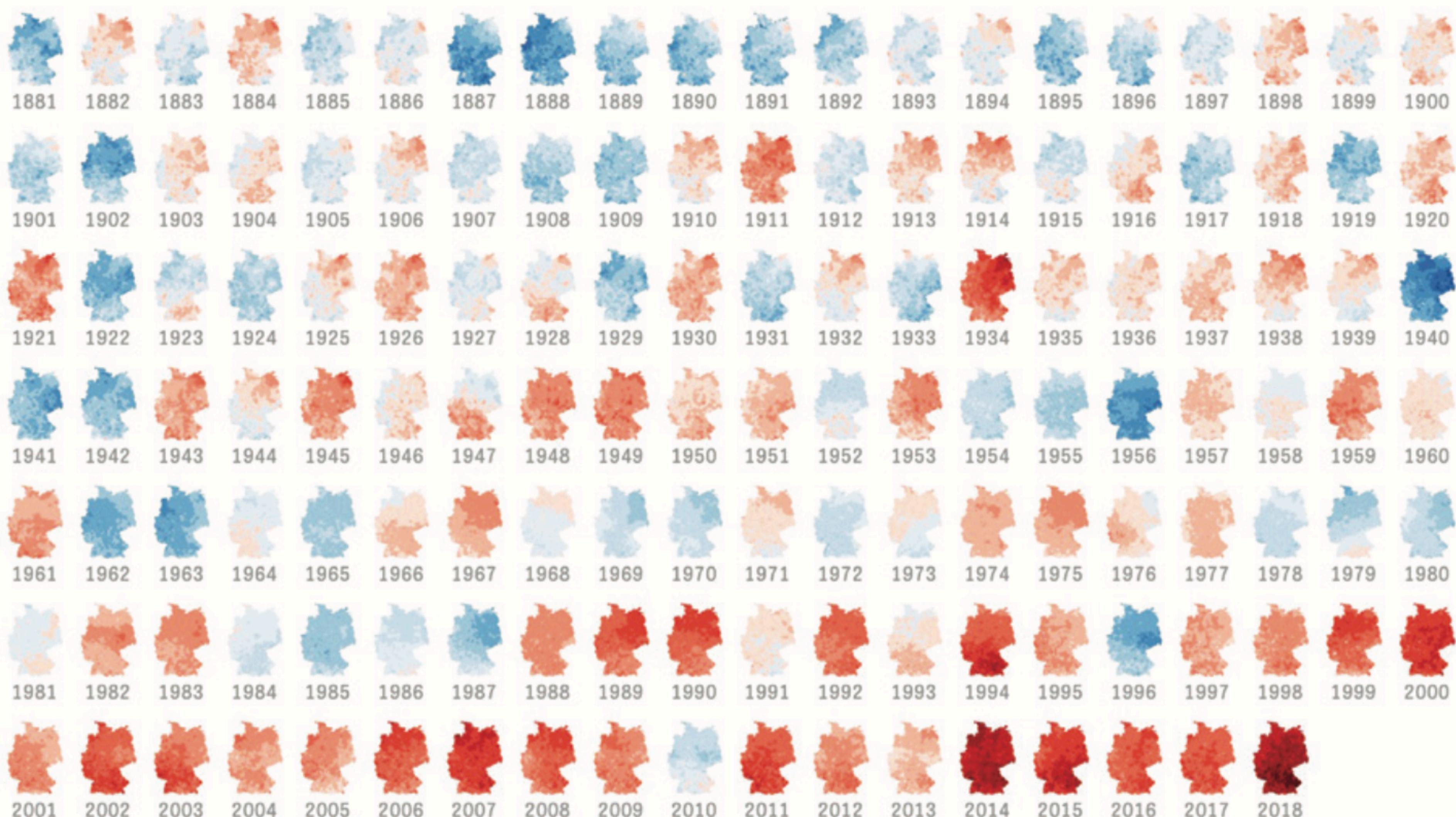
The small multiples approach -
or how to avoid the spaghetti chart

The small multiples approach

- when we face the challenge of including lots of data in a single graph there is risk that we end up with a “spaghetti chart”
 - ▶ we don't need to pack everything into a single graph
 - ▶ we can break it into smaller parts (grid charts, panel charts, facets, trellis charts or small multiples):
 - ▶ create multiple, smaller versions with variations on the basic data
 - ▶ smaller charts should use the **same scale, axes and scope**

Nine of the ten hottest years: all after the turn of the millennium

2018 was the hottest year since records started, with an average annual temperature of 10.5 ° C. 1940 the coldest with 6.6 ° C.



Example of the
small multiples
approach

Source: German Weather Service (DWD), own calculations

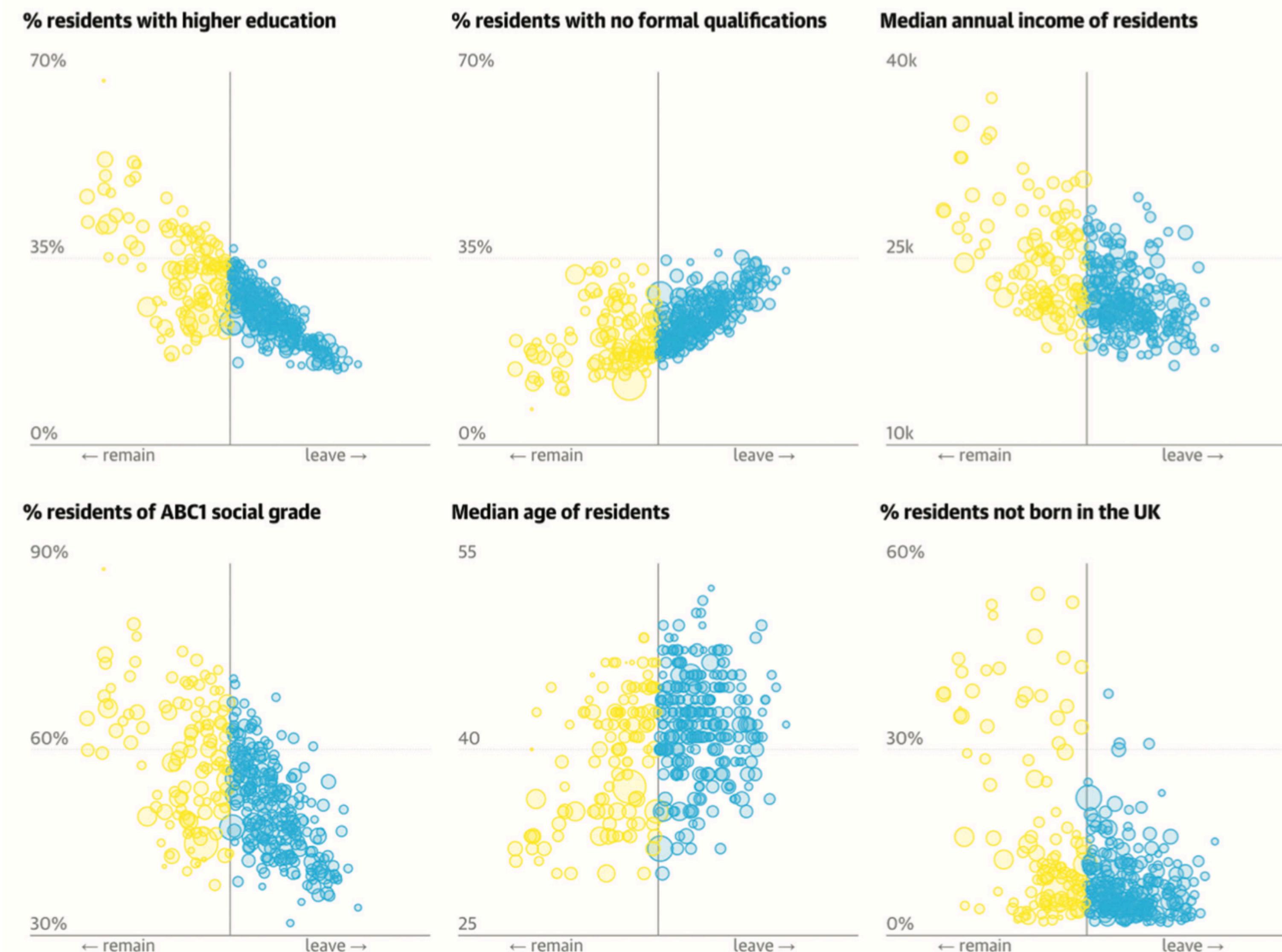
Advantages of the small multiples approach

1. once the reader understands how to read one chart, they know how to read all the charts
2. you can display lots of information without confusing your reader
3. it lets readers make comparisons across multiple variables

EU referendum for Brexit

Every area by key demographics

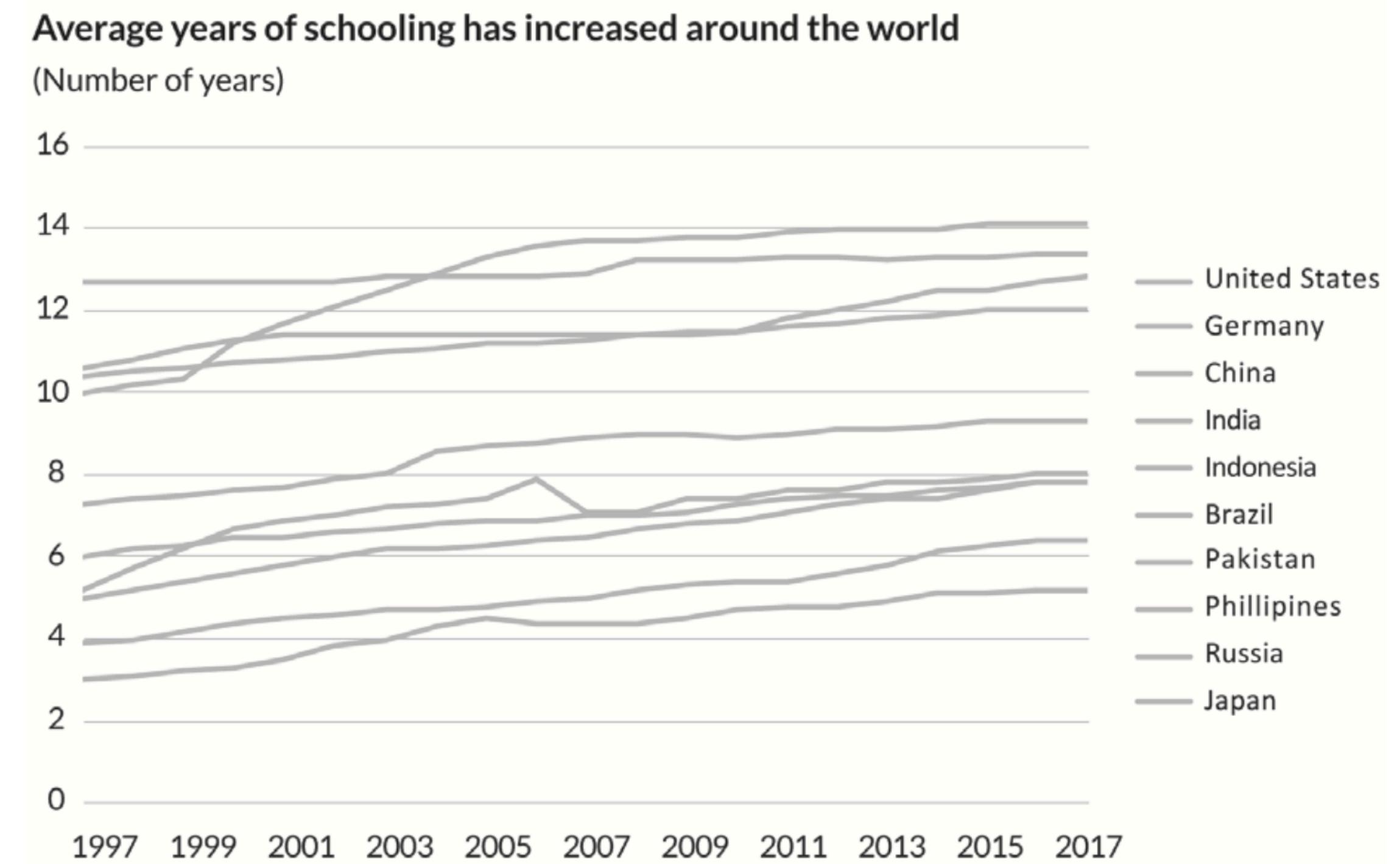
Comparing the results to key demographic characteristics of the local authority areas, some patterns emerge more clearly than others. The best predictor of a vote for remain is the proportion of residents who have a degree. In many cases where there are outliers to a trend, the exceptions are in Scotland.



Start with gray

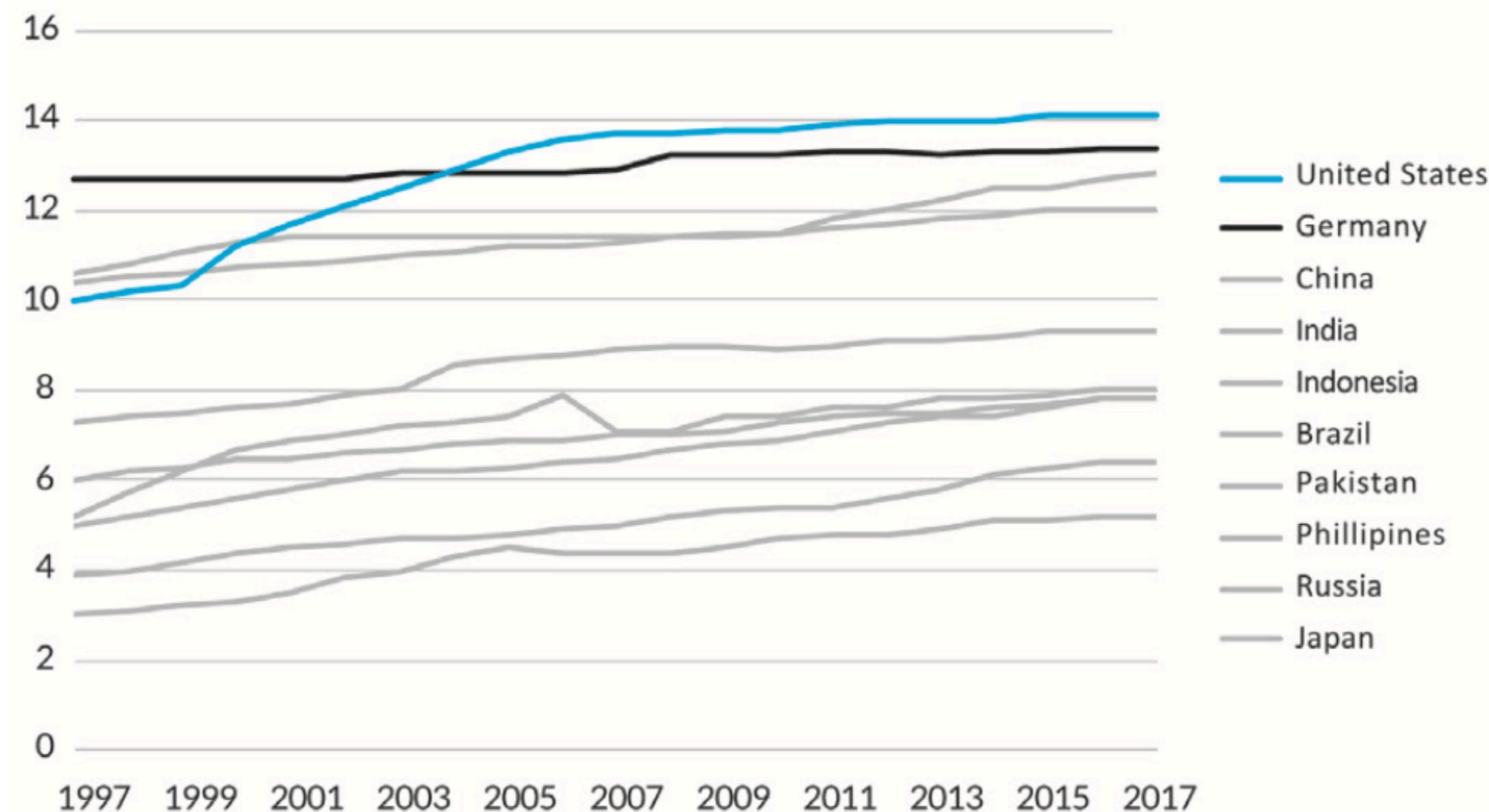
Start with gray

- whenever you make a graph, start with all-gray data elements
- by doing so, you force yourself to be purposeful and strategic in your use of color, labels, and other elements
- starting with gray forces us to deliberately choose what elements to put into the foreground

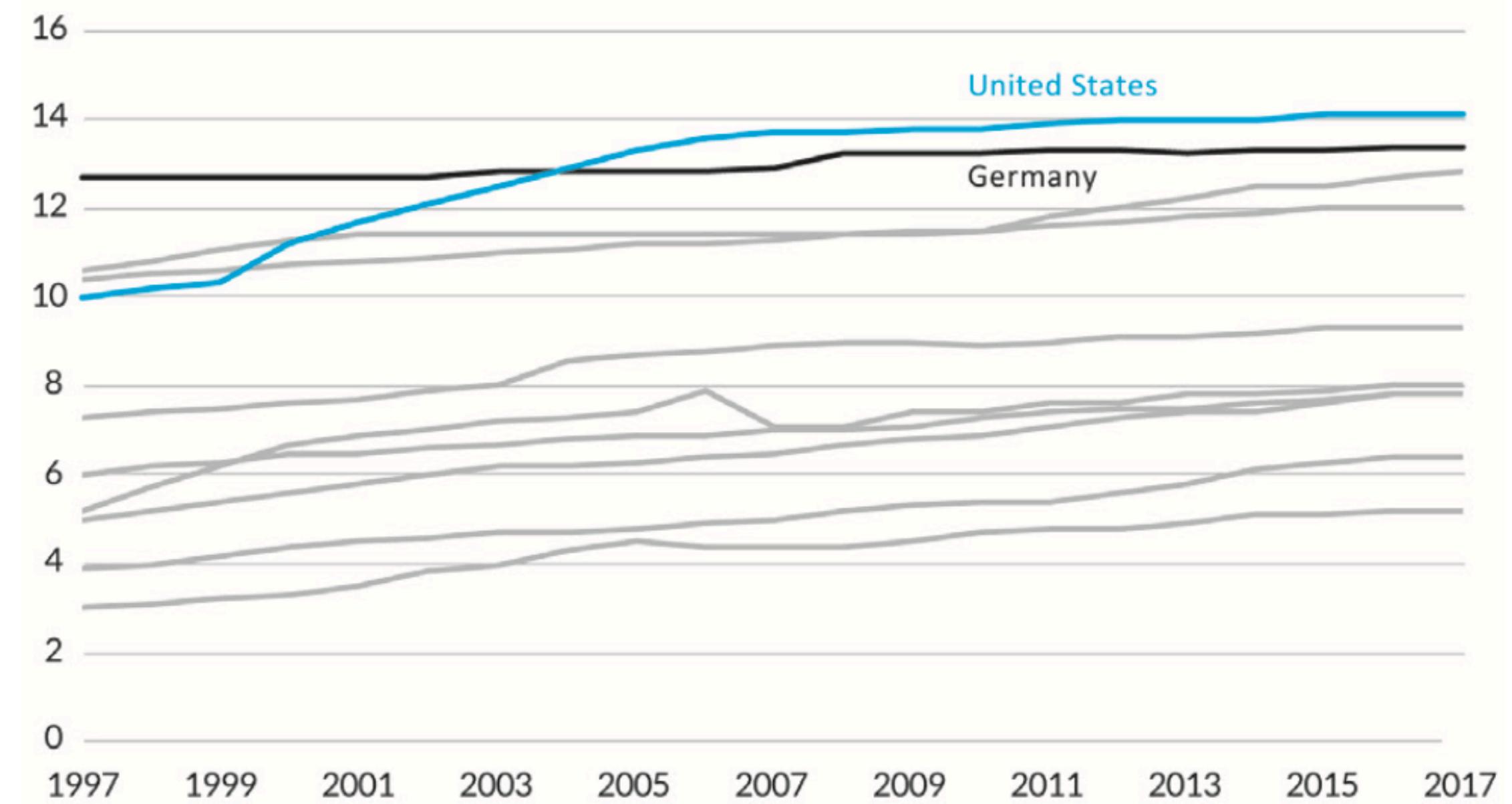


Direct your reader's attention

Average years of schooling has increased around the world
(Number of years)

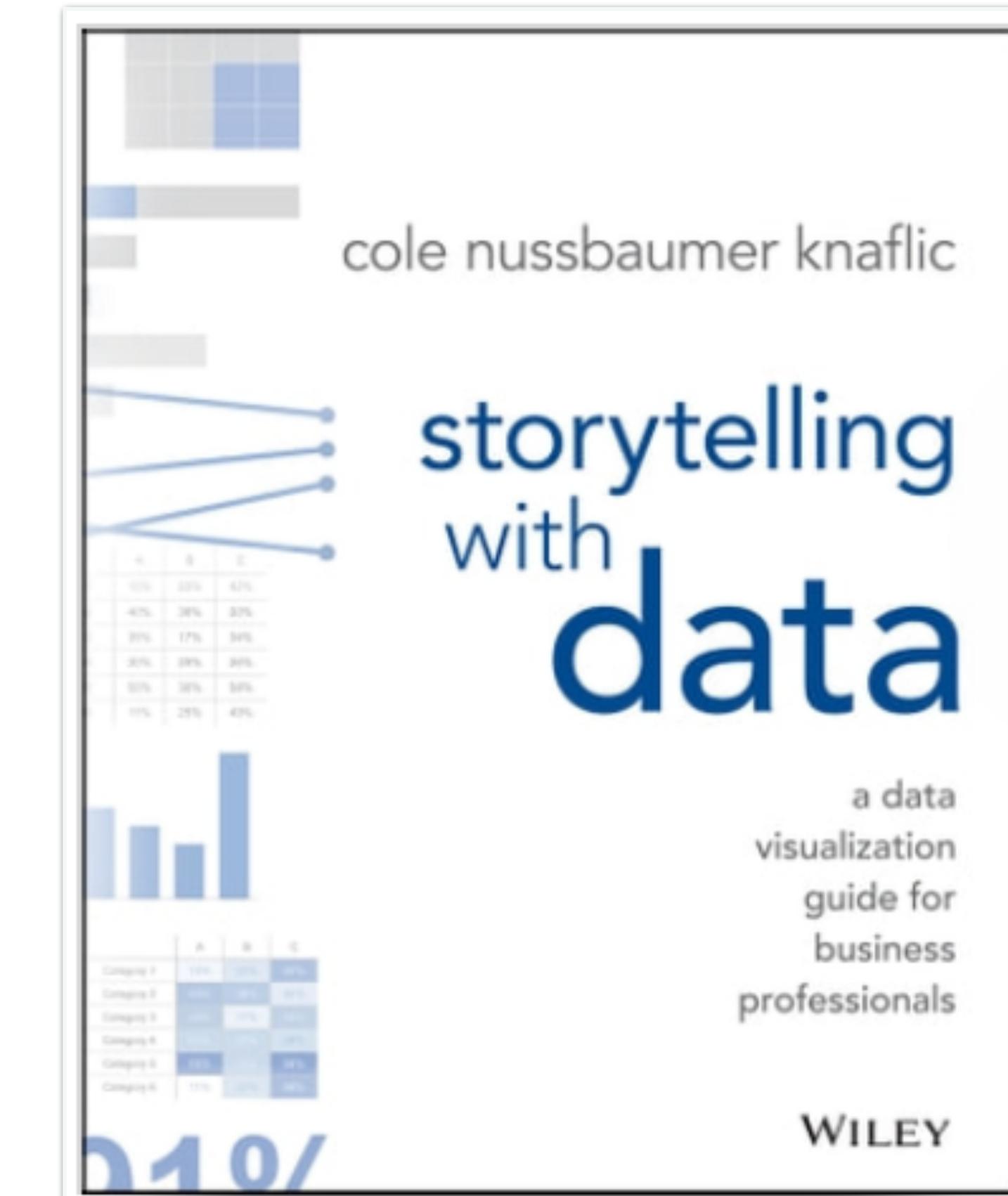
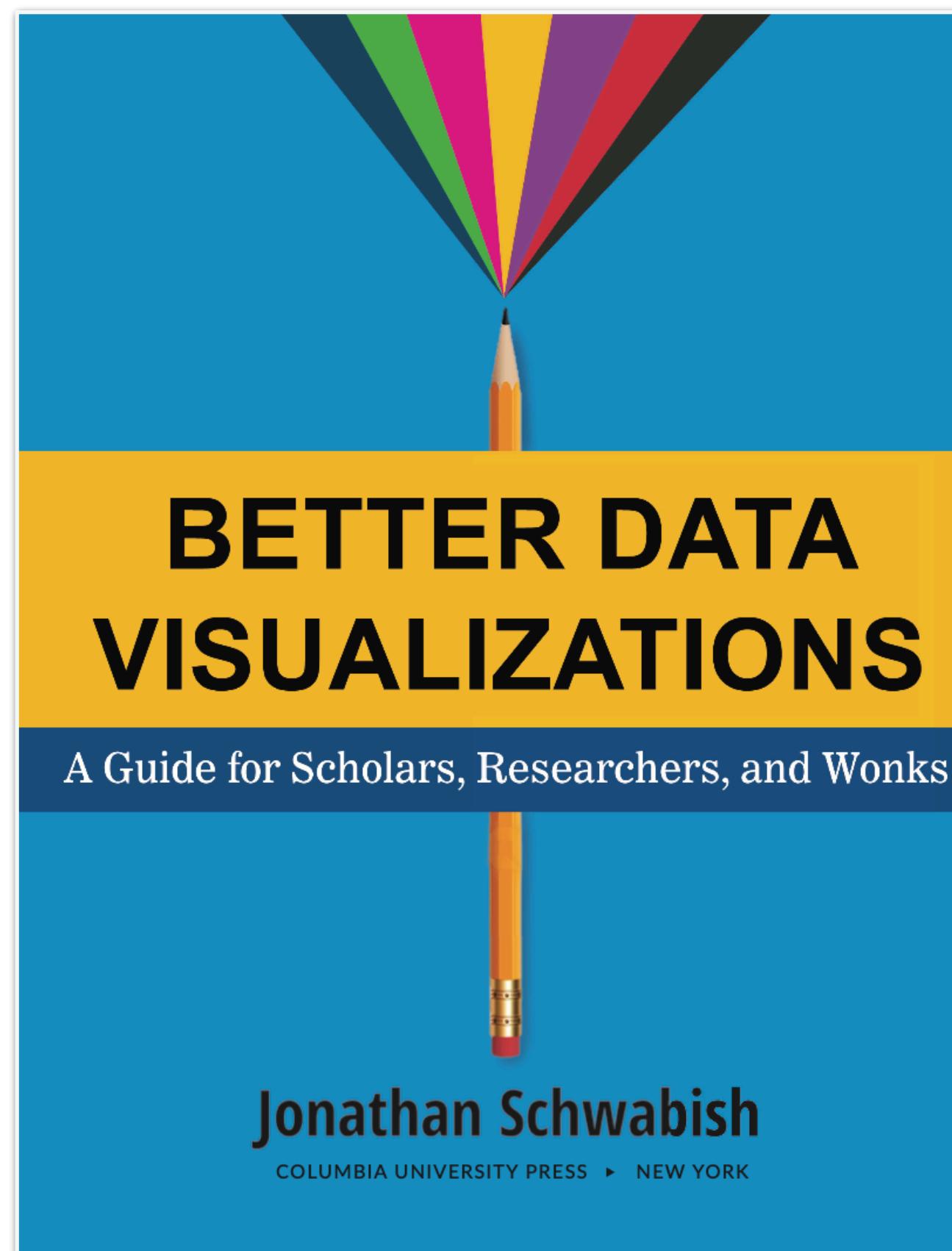


Germany and the United States have the highest average years of completed schooling
(Number of years)



Questions?

Literature



Sources

- **Slide 4** - Source: Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator).
- **Slide 5** - Source: Our World in Data; in Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 8-10** - Source: Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 13-17** - Source: Our World in Data; in Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 20,21** - Source: Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 24,25** - Source: Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 26** - Source: flowingdata.com: <https://flowingdata.com/2022/02/03/when-there-were-more-deaths-than-births-in-the-u-s/>
- **Slide 27** - Source: Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press
- **Slide 31:** Guardian: https://www.theguardian.com/politics/ng-interactive/2016/jun/23/eu-referendum-live-results-and-analysis?CMP=twt_b-gdndata
- **Slide 33,34** - Source: Our World in Data; in Schwabisch, Better Data Visualizations, A Guide for Scholars, Researchers and Wonks; Columbia University Press