

Data Visualization and Visual Analytics

Time

Study Program Data Science
Prof. Dr. Tillmann Schwörer

Data Domains

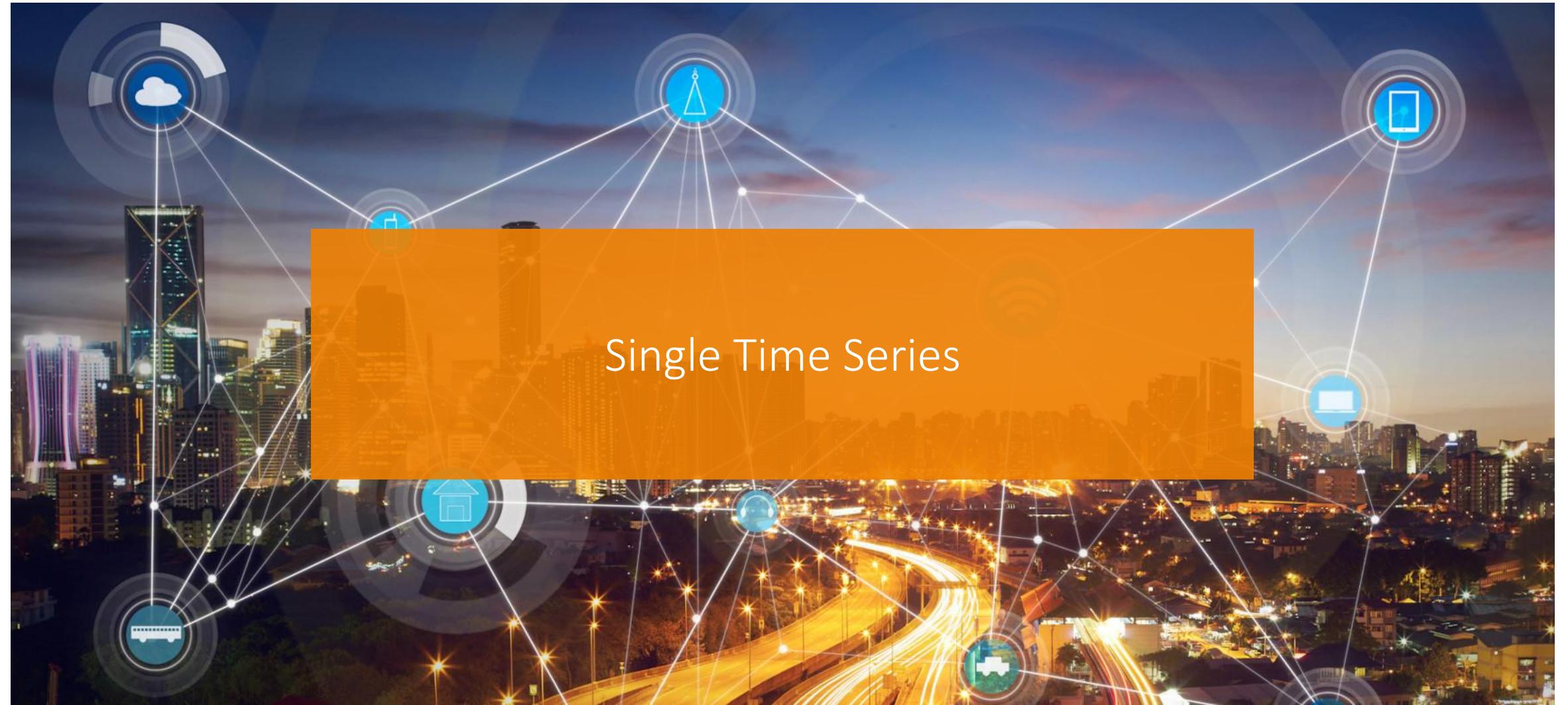
Comparing Categories | Relationships | Geospatial | Time |
Part-to-whole | Distributions | Uncertainty | ...

Storytelling

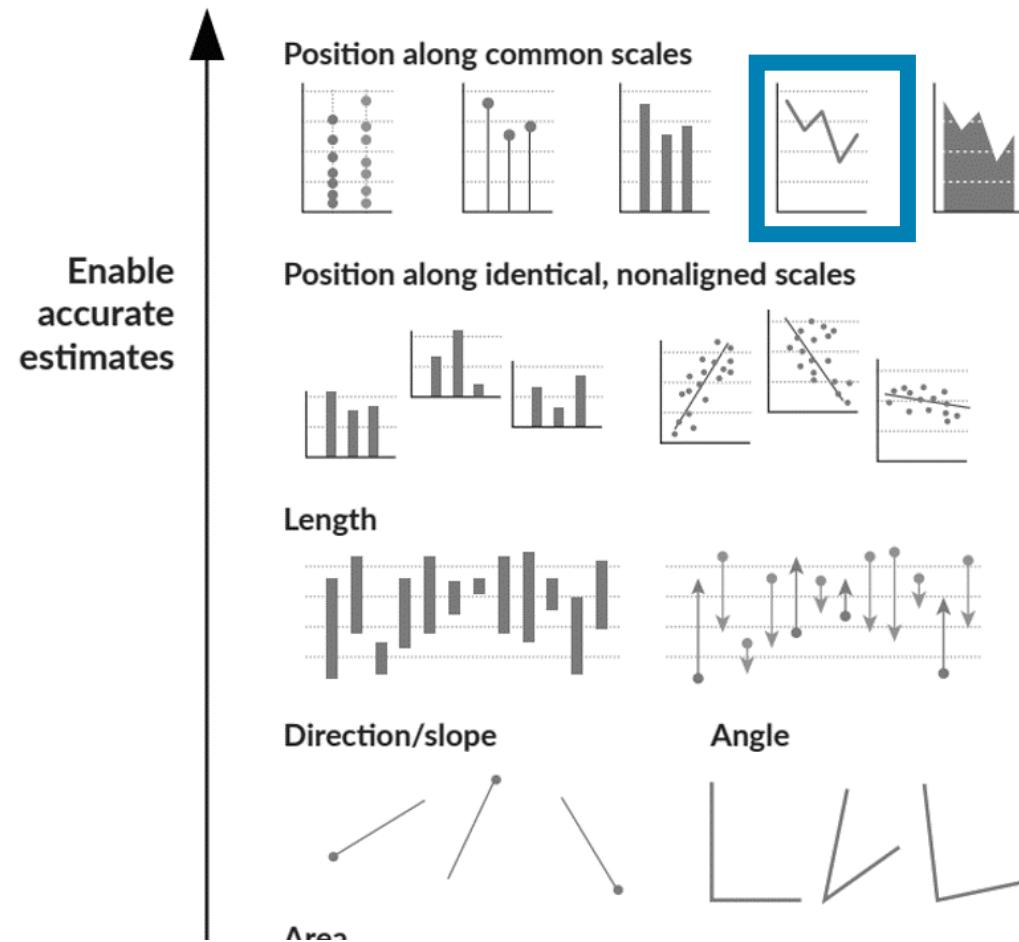
Perception +
Visualization Design

Python + Tools

Interactive
Visualization



Line charts

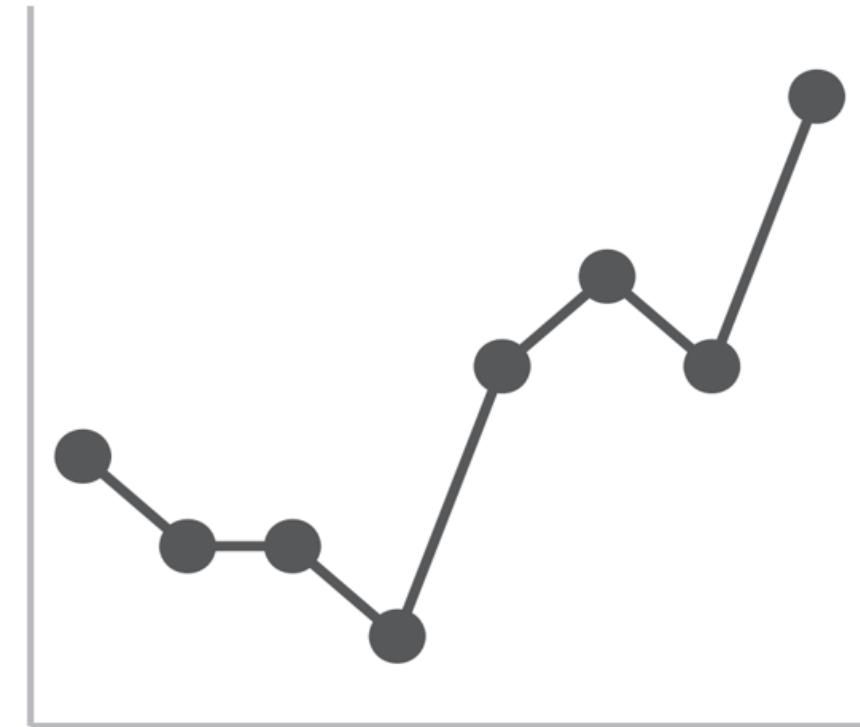
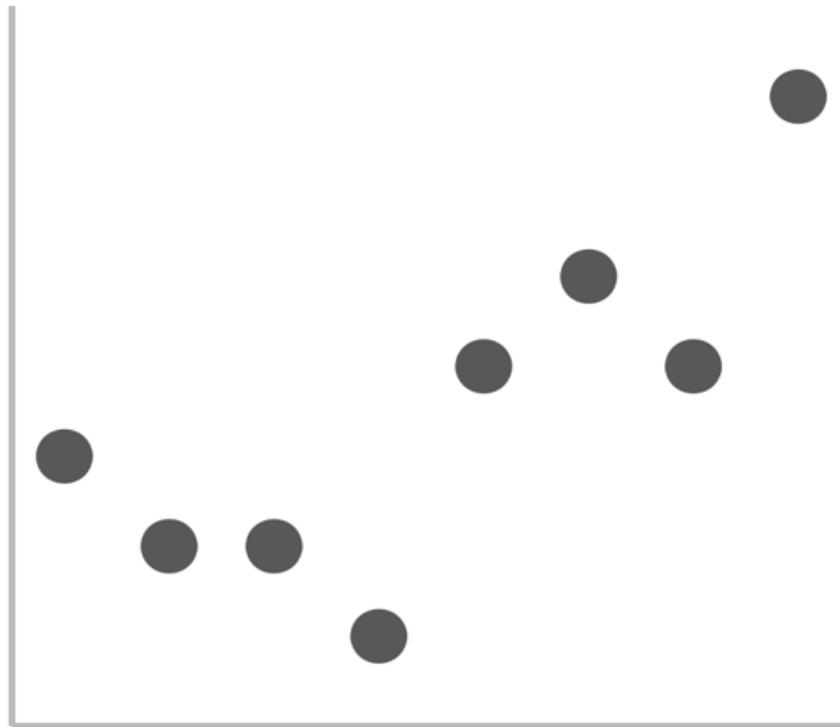


Line plots are the natural default choice:

- ▶ Accurate perception
- ▶ Viewers are familiar with it

Lines vs. Markers

Gestalt principle of **connection**: line plots help perceive the natural order in the data



Handling missing values

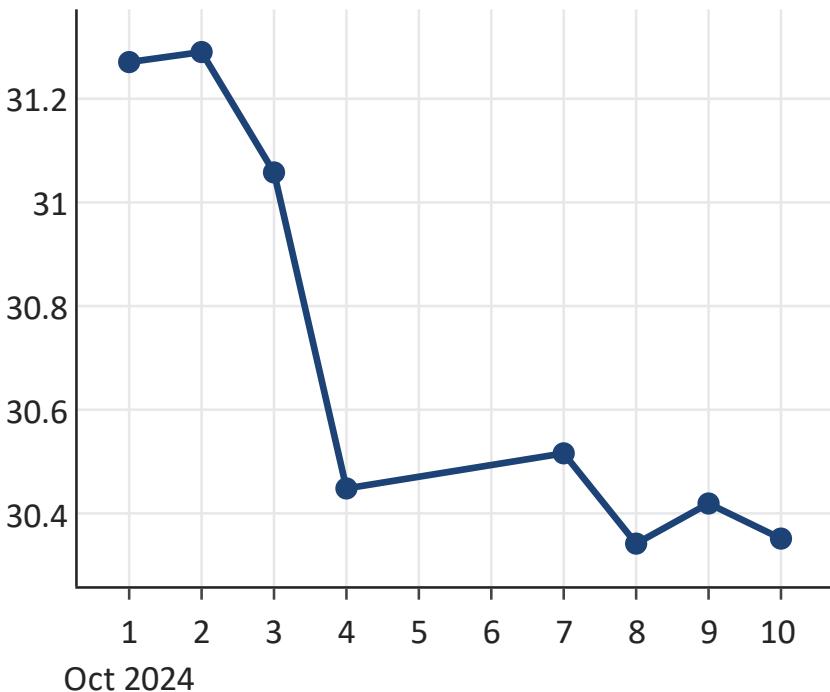
RWE share price in €



- ▶ No data available on October 5 and 6 (weekend)
- ▶ The line gives a **wrong impression of smoothness** and of full data availability

Handling missing values

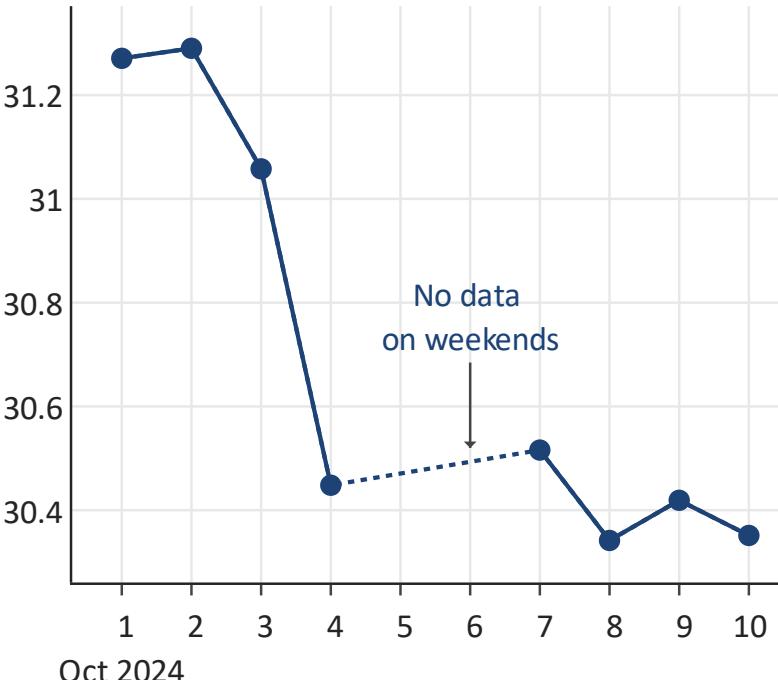
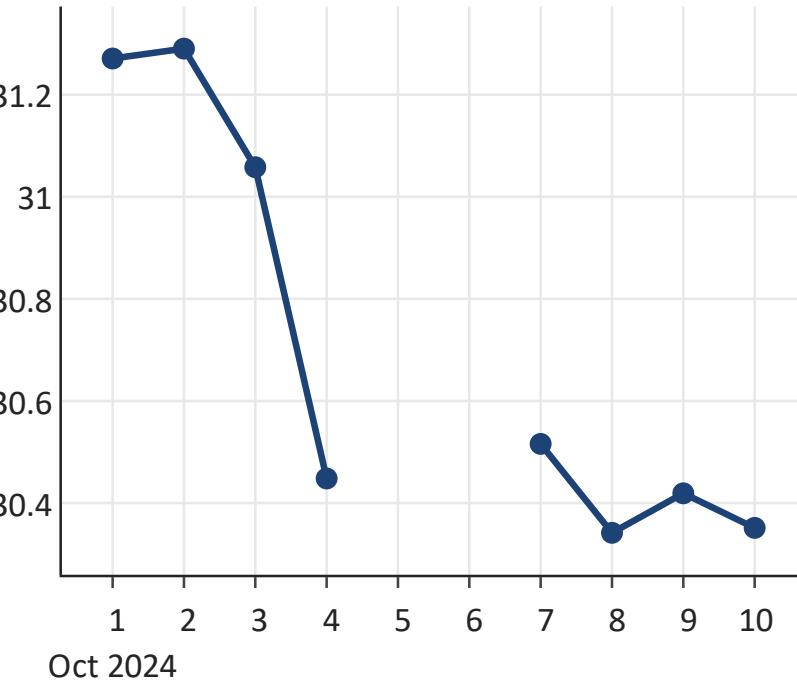
RWE share price in €



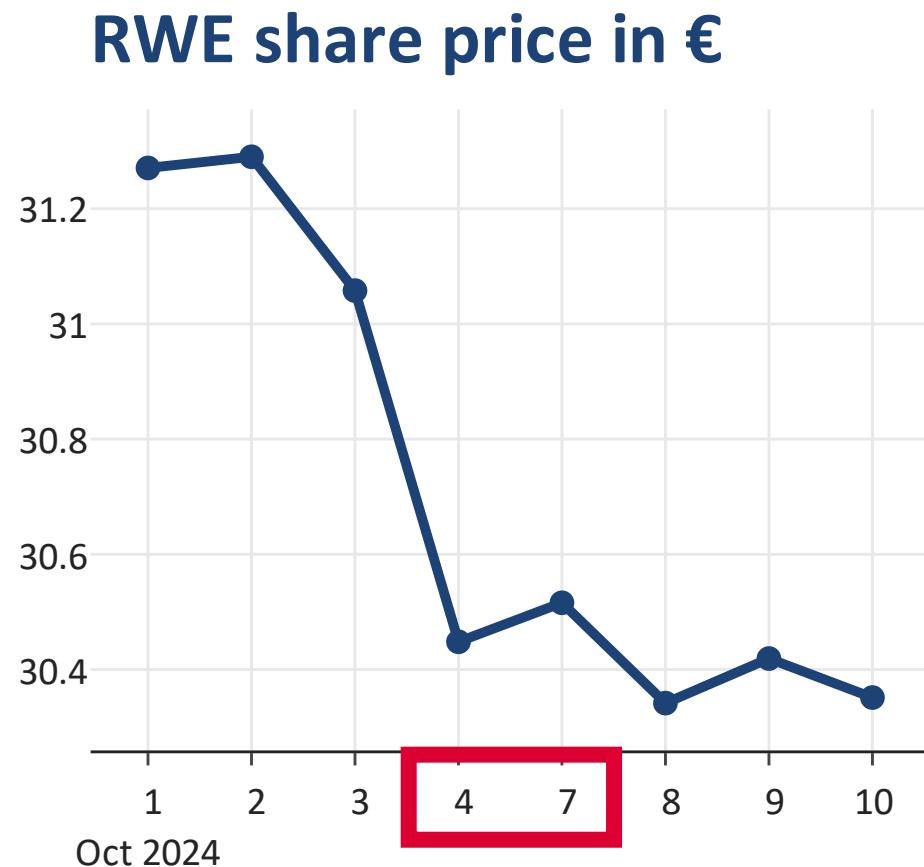
- ▶ Combining lines with markers is useful if missing values are present
- ▶ **Markers** show for which of the dates we have actual data points

Handling missing values

- We can make the presence of missing values even more explicit by using **gaps**, **dotted lines**, and **annotations** (→ truthfulness)



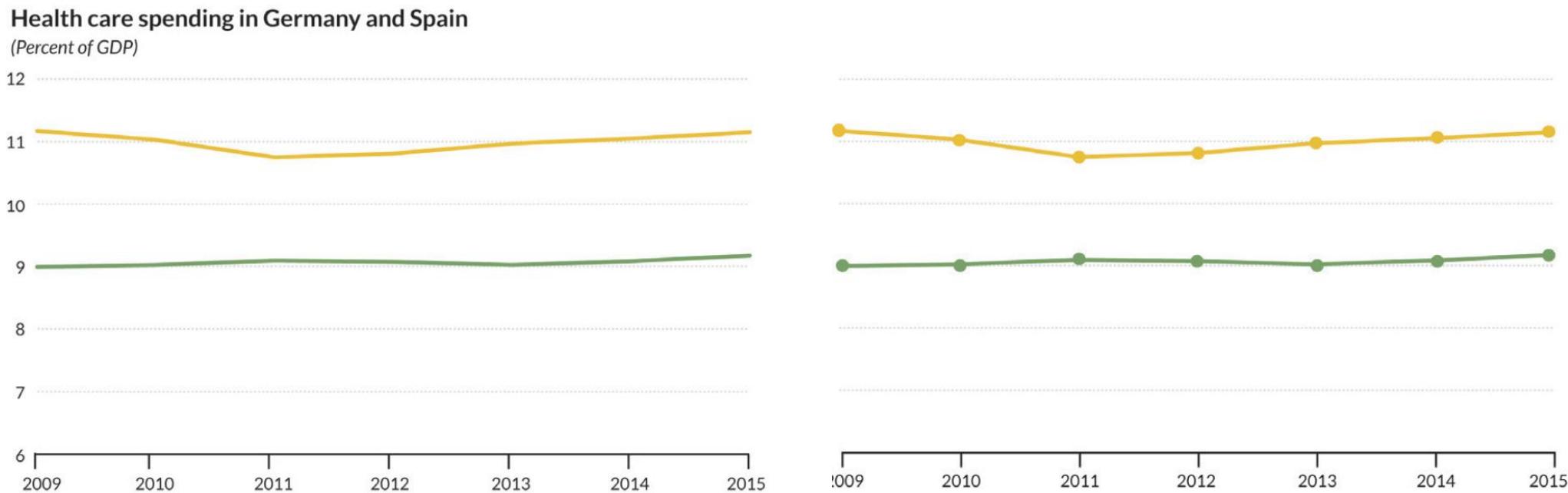
Handling missing values



- ▶ Usually, visualizations of stock price data **cut out weekends, holidays, and other times without stock trading**
- ▶ Note: this is fine here, because the missing values are justified (it does not represent a data error)

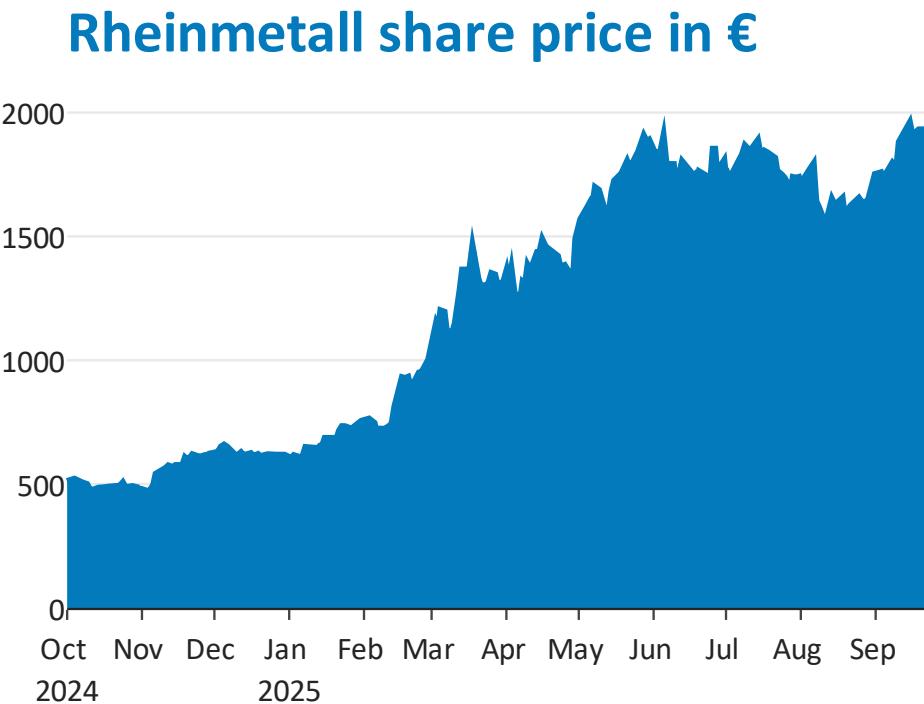
Adding markers can be visually helpful ...

- ▶ if we have few measurements and small fluctuations
- ▶ if we have data points at irregular intervals
- ▶ to highlight specific dates of interest



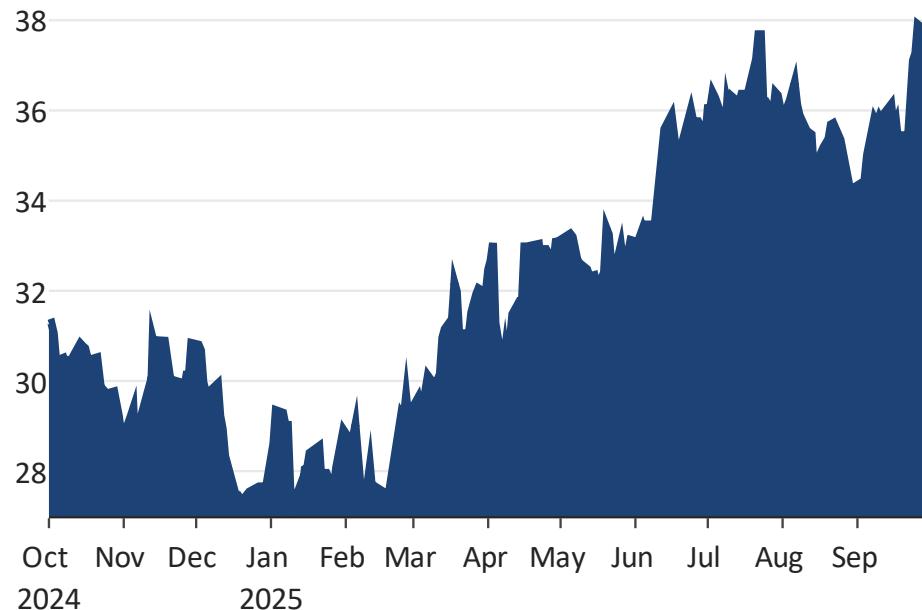
Area plot?

- ▶ For a single time series, the area plot can be a possible alternative
- ▶ Attaches a strong visual weight to the data (as opposed to axes, gridlines, ...)
- ▶ But: unnecessary ink?



Principle of proportional ink

RWE share price in €



Violates Principle of Proportional Ink



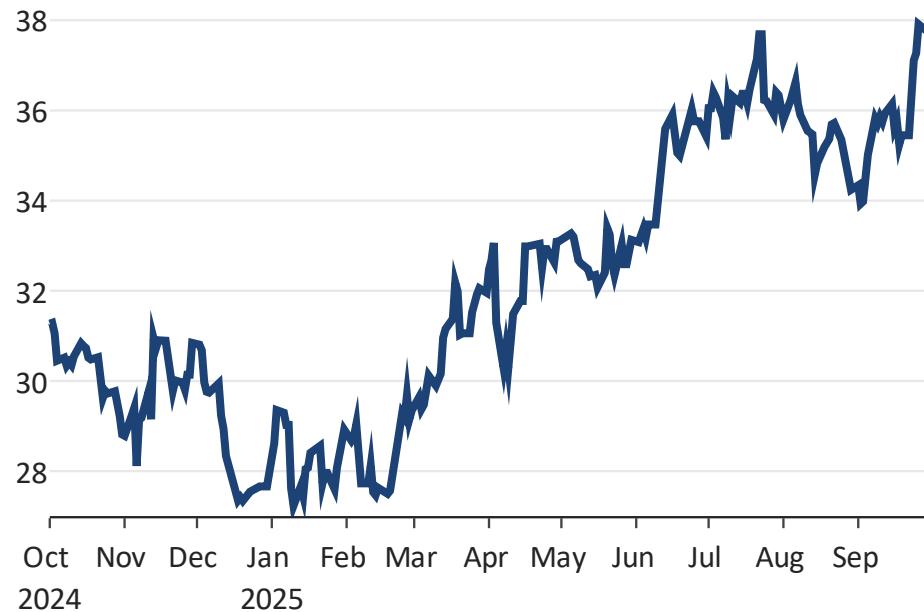
RWE share price in €



Variations are hard to see

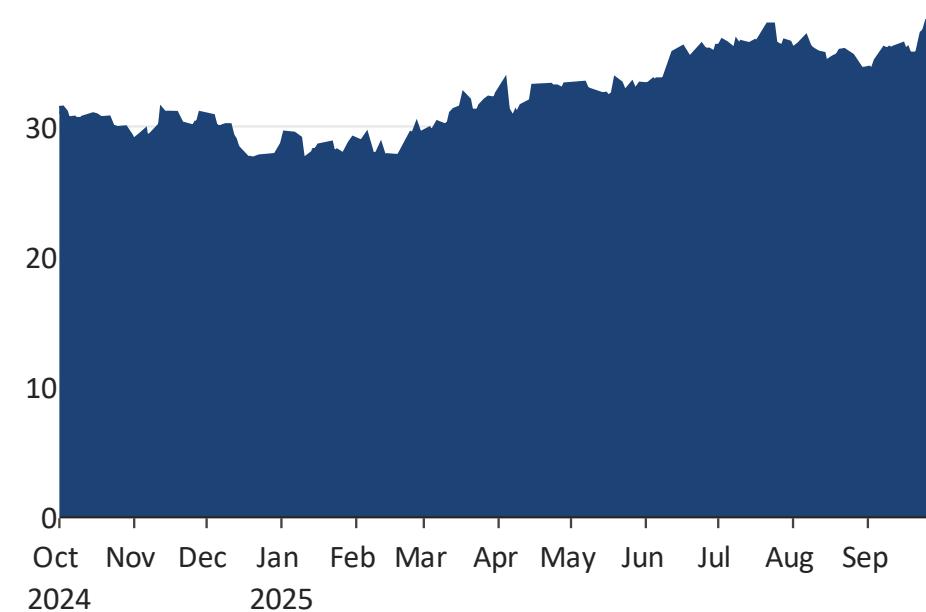
Principle of proportional ink

RWE share price in €



- Does not violate principle of proportional ink!
- Highlights variations over time

RWE share price in €

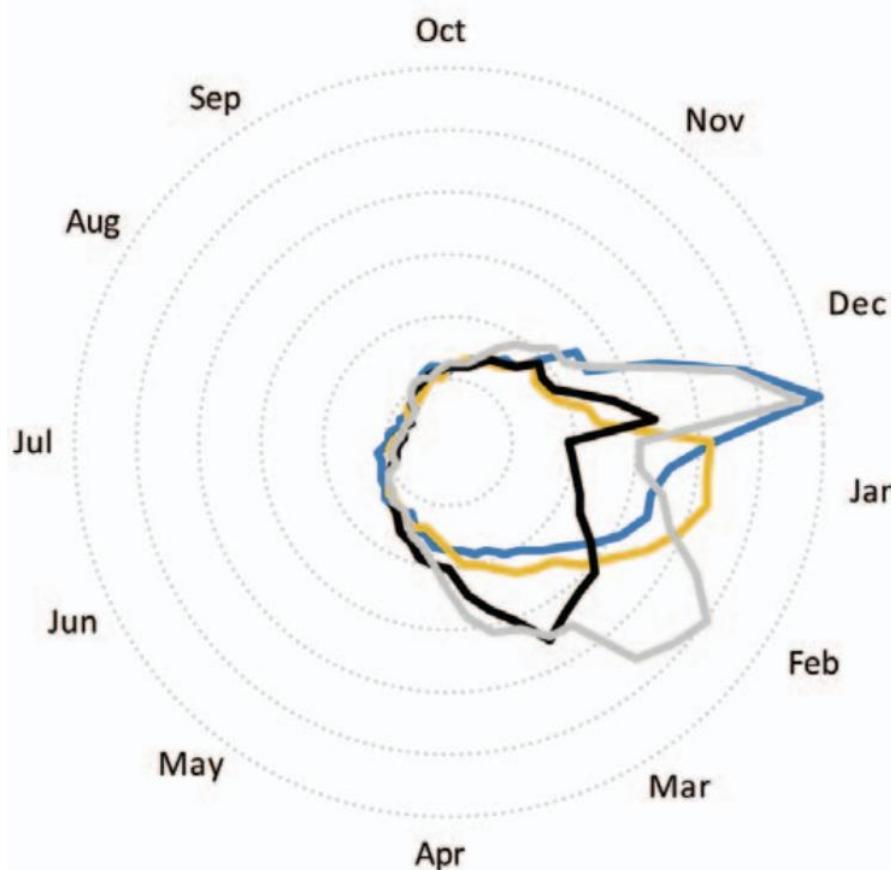


Variations over time are hard to see

Circular charts

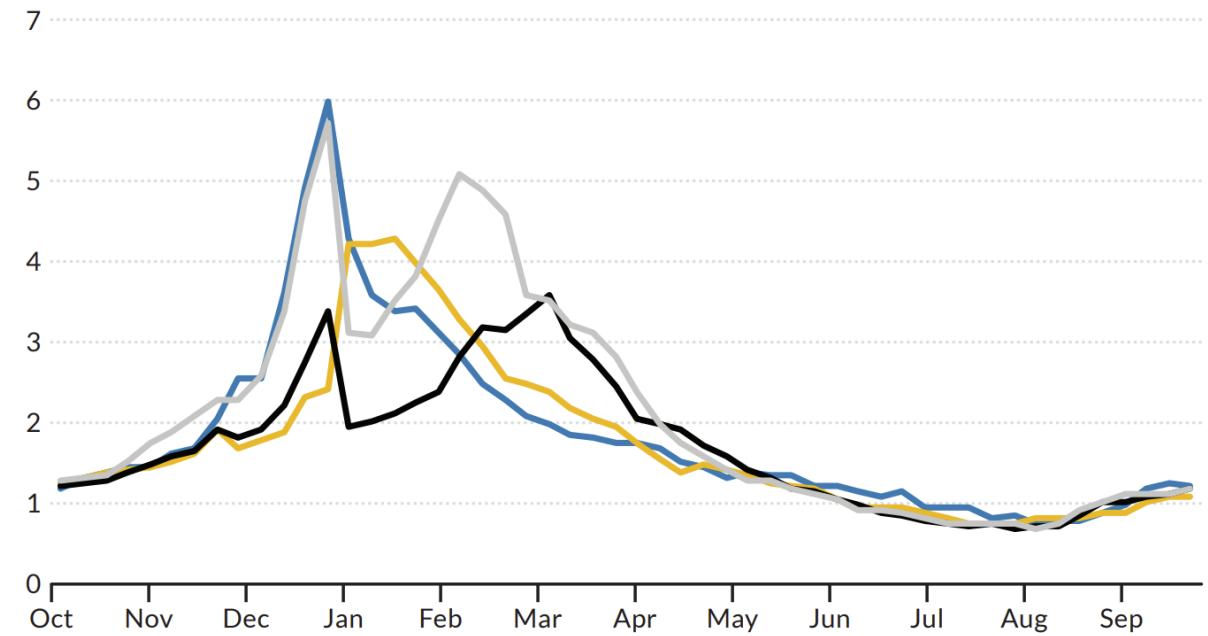
Percent of ER visits for the flu (2014–2017)

— 2014 — 2015 — 2016 — 2017



Source: Centers for Disease Control and Prevention

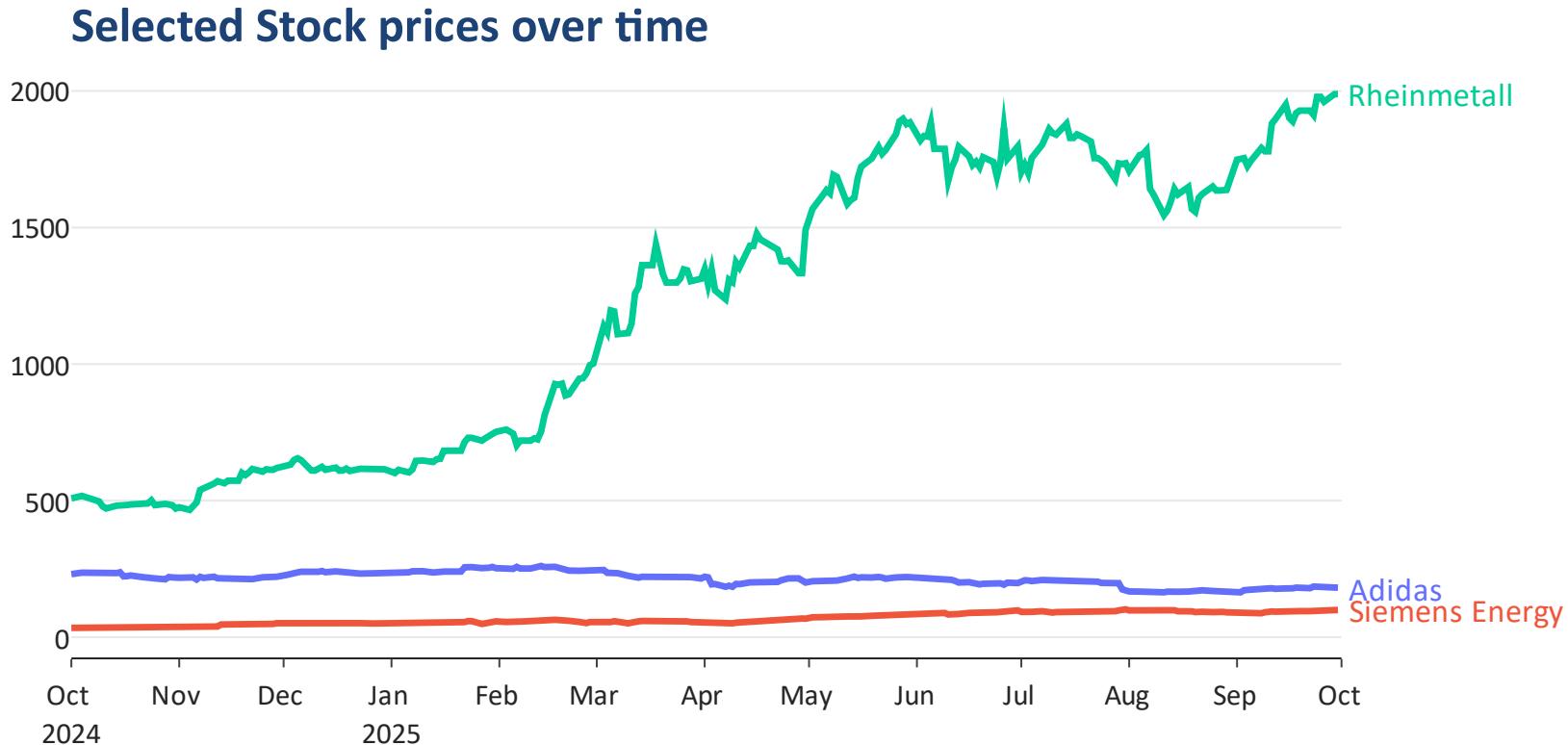
Circular charts can be useful to highlight cyclical patterns in the data, but it is much **harder to accurately perceive the data**





Comparing multiple time series

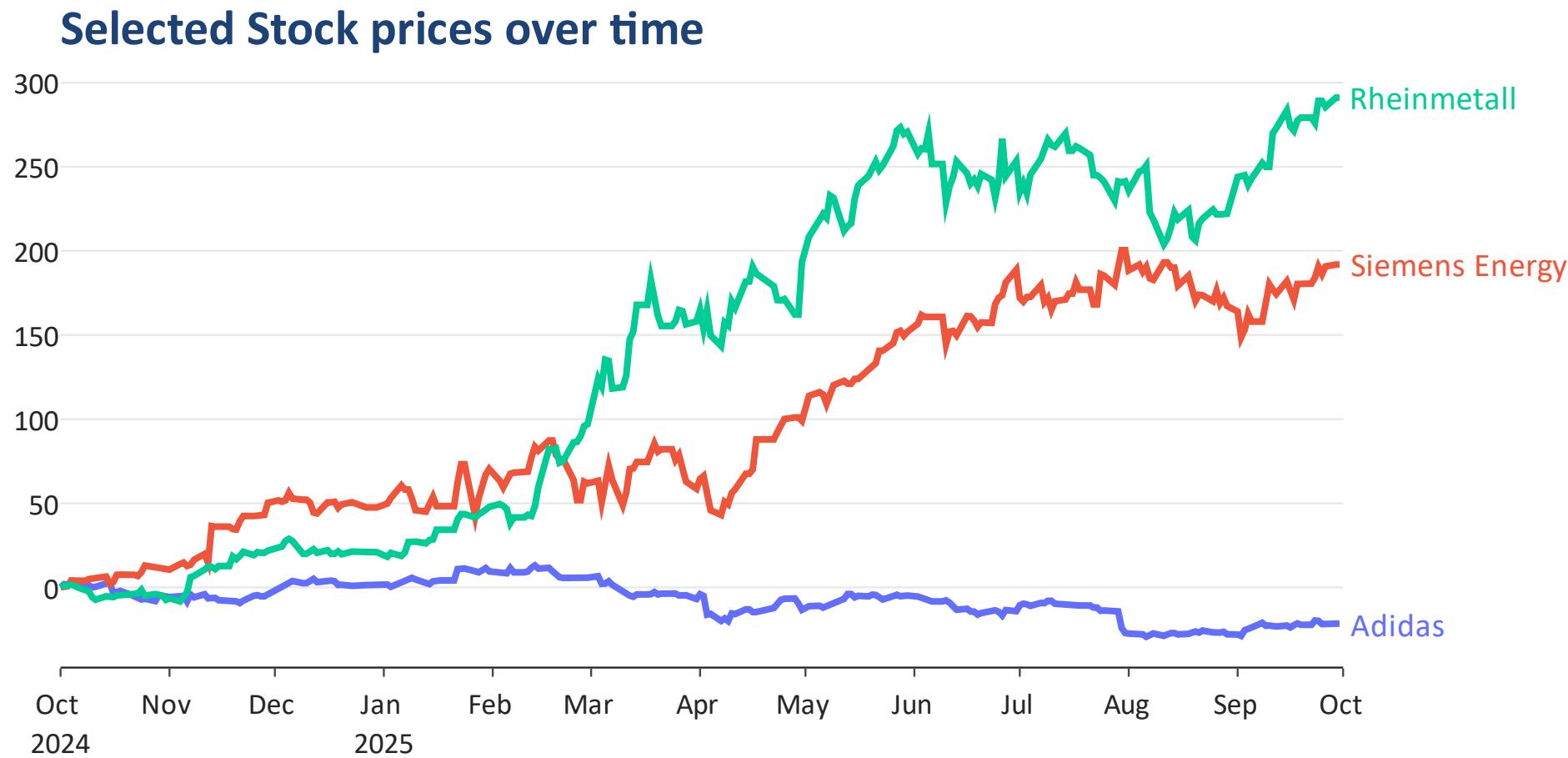
What is the problem here?



The plot draws attention to Rheinmetall due to higher stock price levels, but these are not meaningful. We should care about relative price changes!

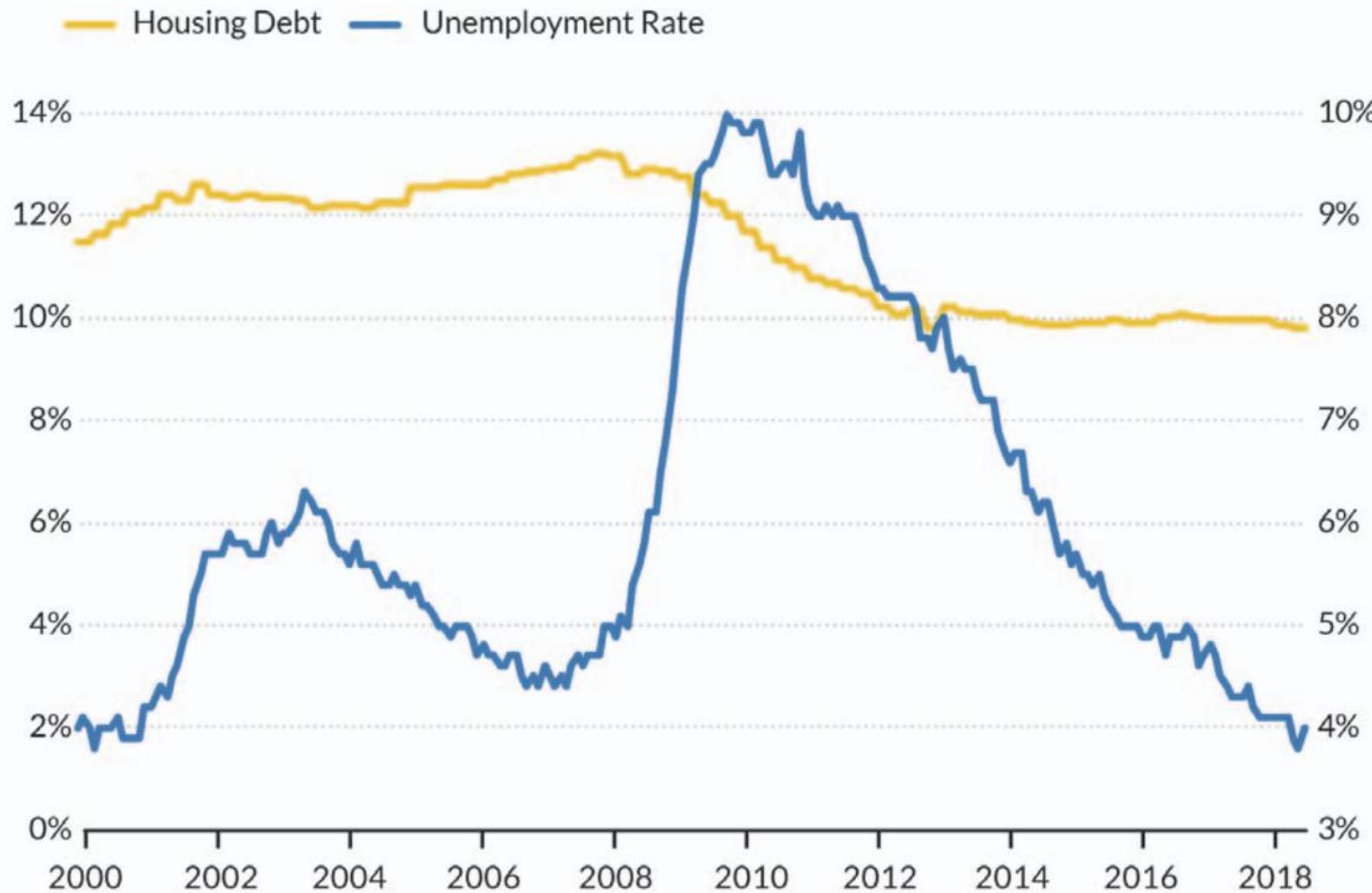
Index Plot

The index plot makes the data comparable: percent change with respect to the first date



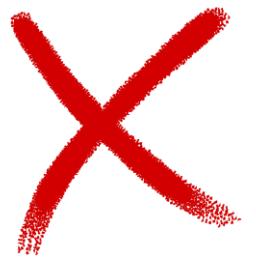
Dual y-axis?

Economic climate for consumers in 2017 and 2018 was quite good



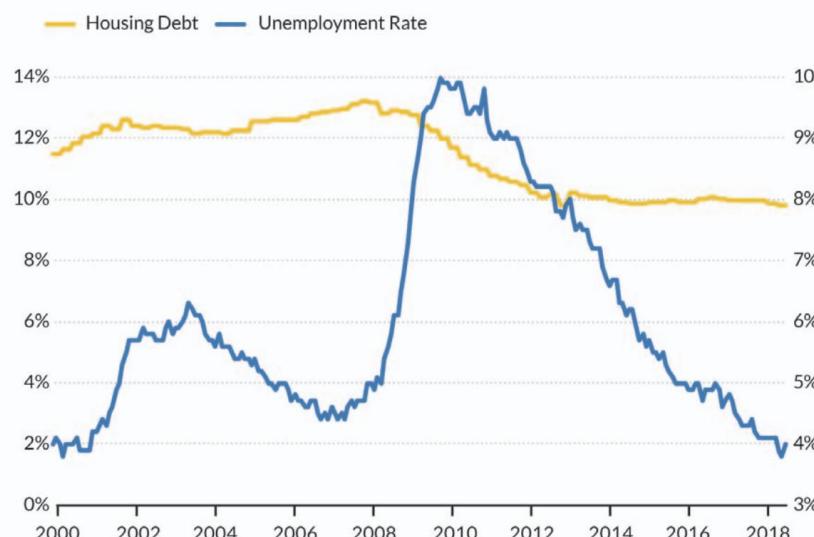
Problems of the dual y-axis

Dual y-axis
must be avoided



- ▶ Hard to read
- ▶ Gridlines may not align well
- ▶ Choice of axes scaling is arbitrary: how we define the left and right y-axis affects the interpretation heavily → Audience will (wrongly) try to compare the levels of the two lines

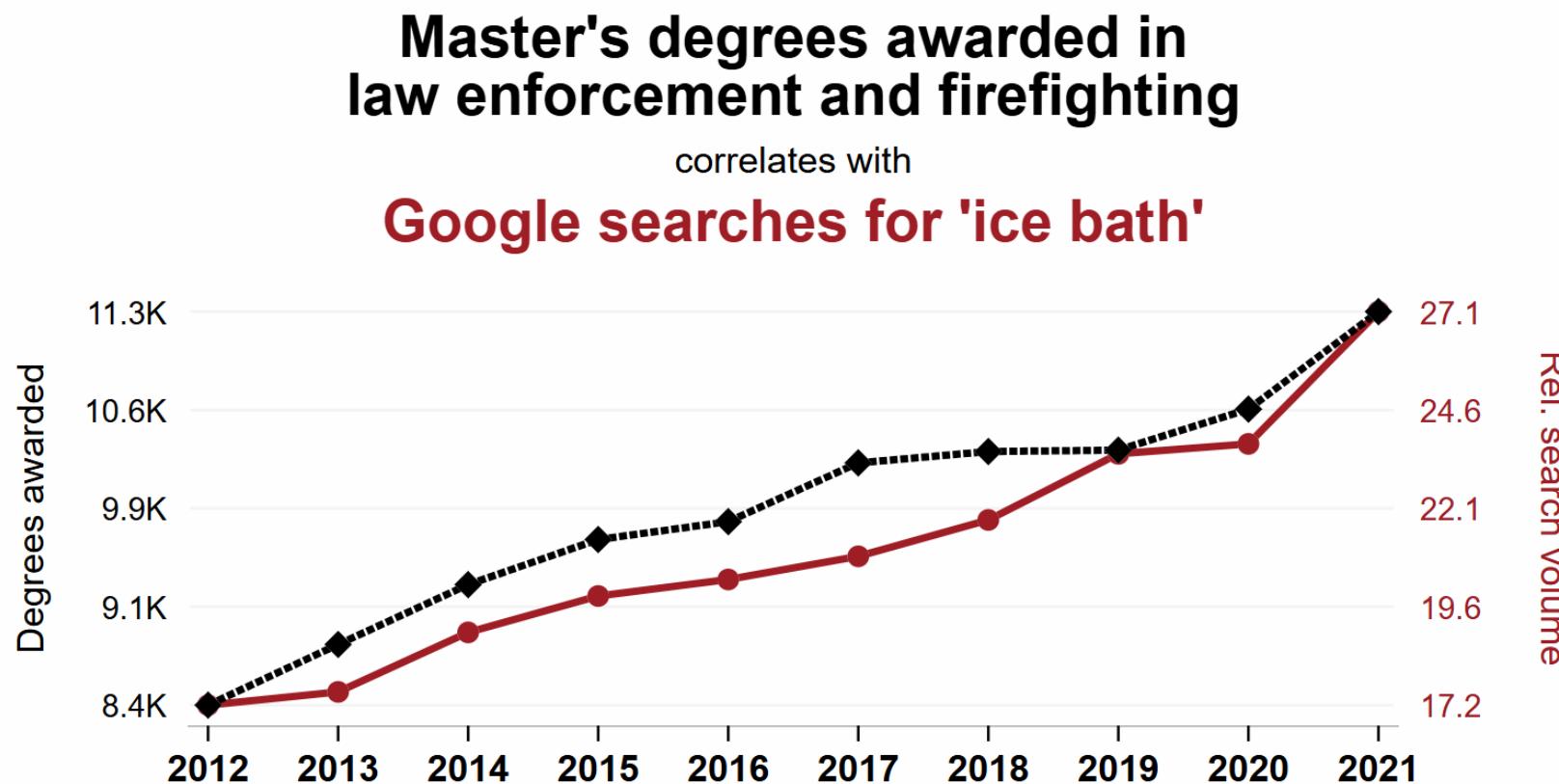
Economic climate for consumers in 2017 and 2018 was quite good



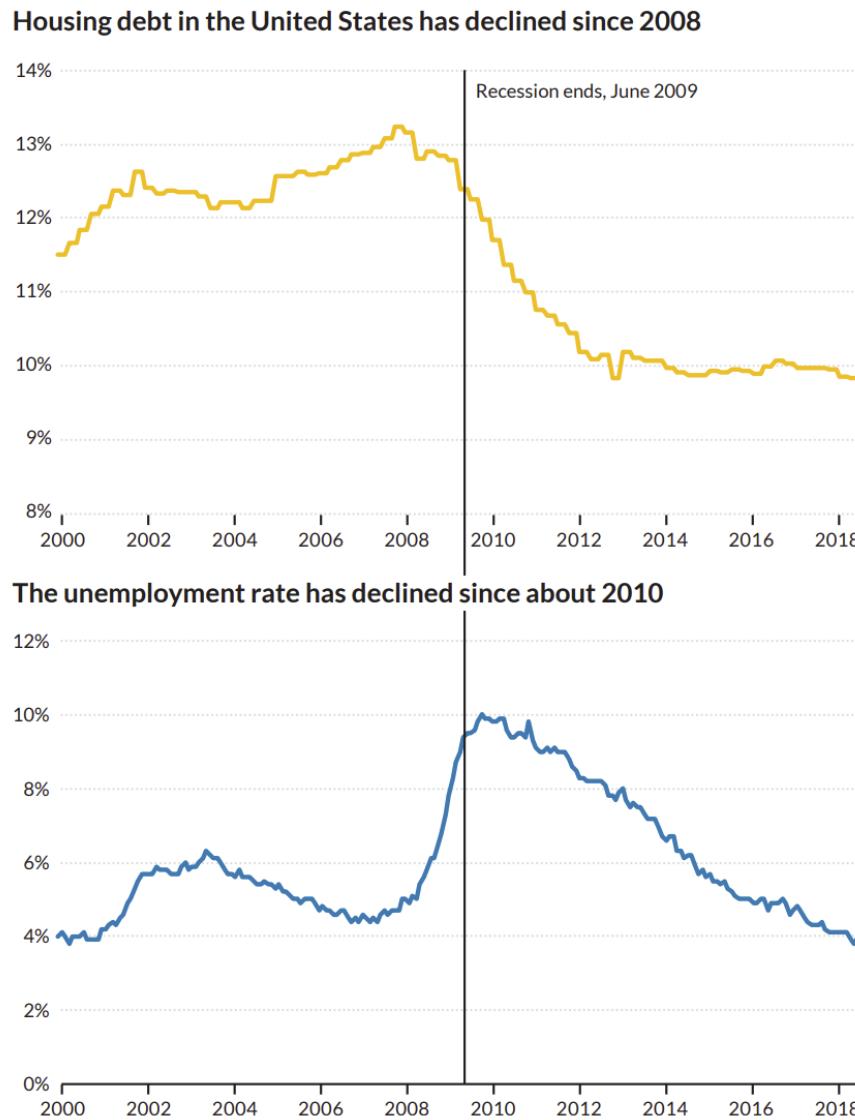
Spurious correlation



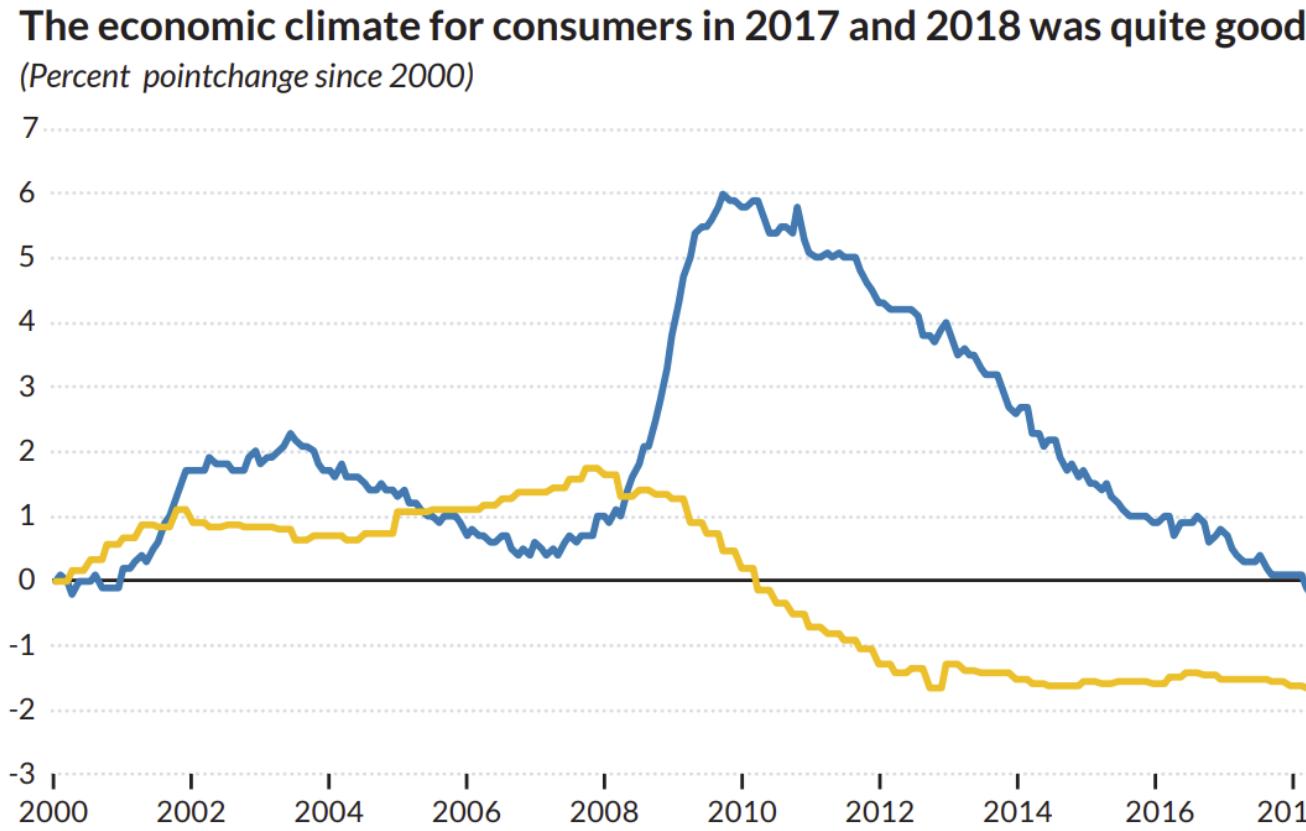
We can create the impression of correlation between seemingly independent data series simply by adjusting the vertical axes.



Alternative 1: Stacked line charts



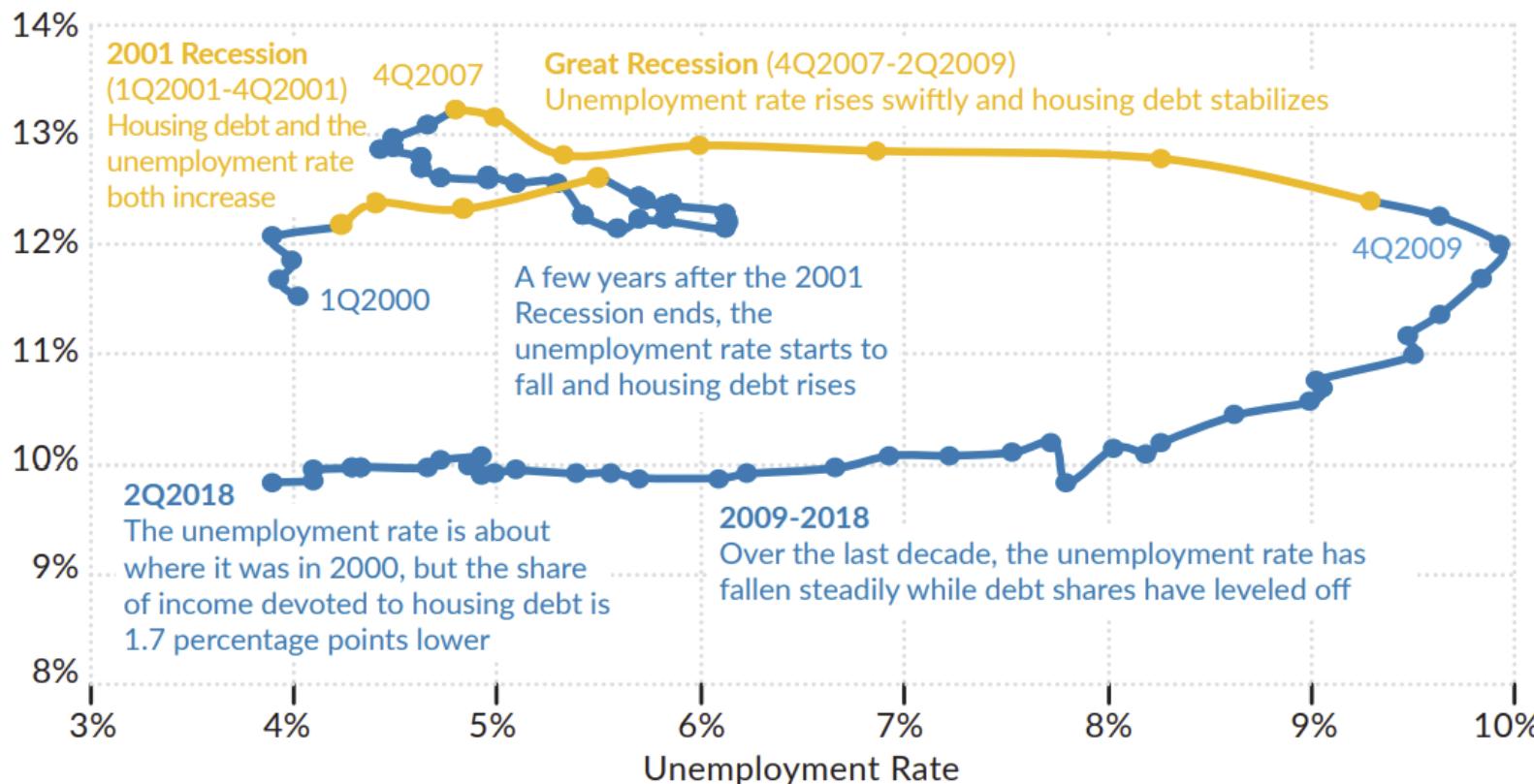
Alternative 2: Index plot



- ▶ Comparison of many time series
 - ◆ in different units
 - ◆ on different absolute levels (Germany vs. Hamburg)
- ▶ Reference point is typically the start or end
- ▶ Works well even with more than 2 time series
- ▶ Focus is on the changes, not the levels → optionally: annotate absolute levels at end point

Alternative 3: Connected Scatterplot

The U.S. economy appears supportive of the consumer with low-unemployment rate and housing debt
(Household debt service ratio)





The implications are clear ...

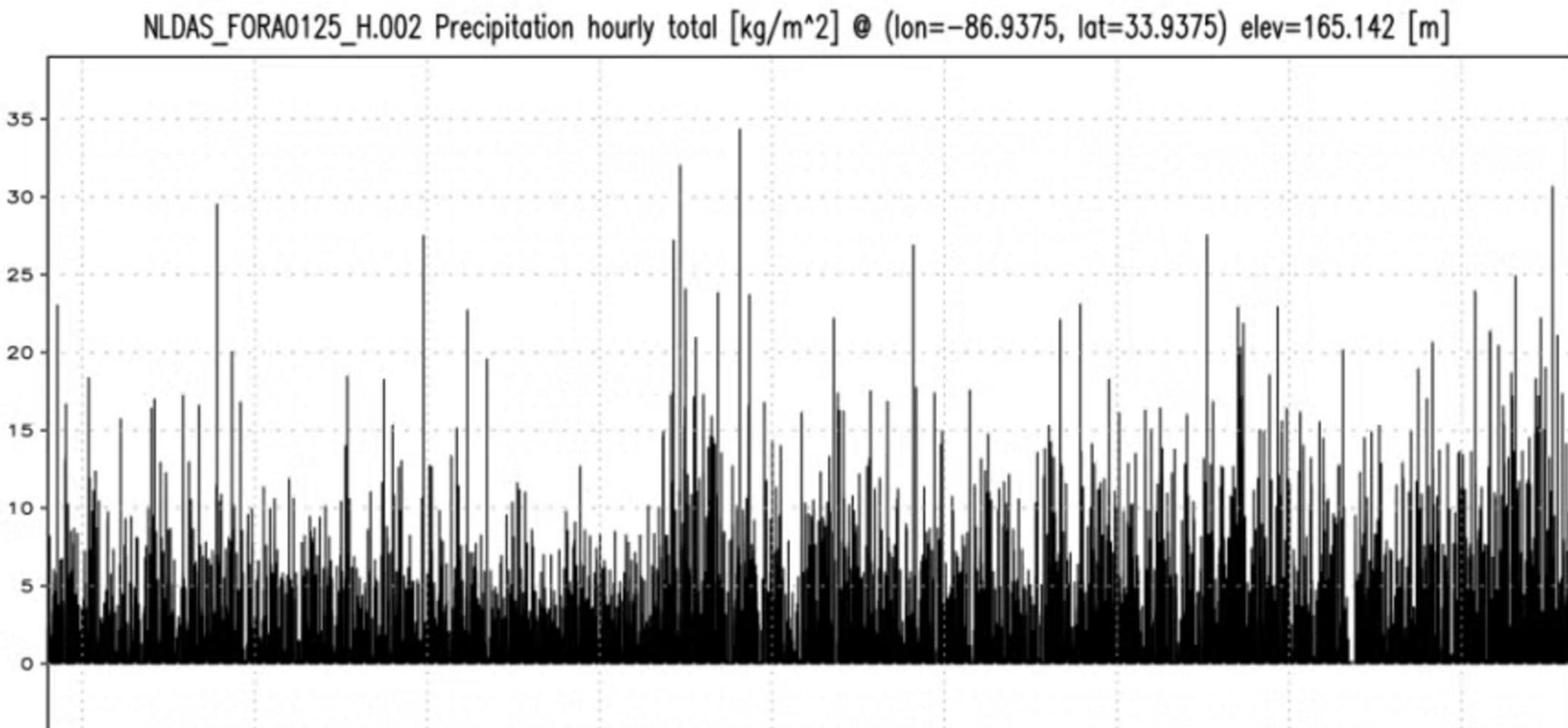
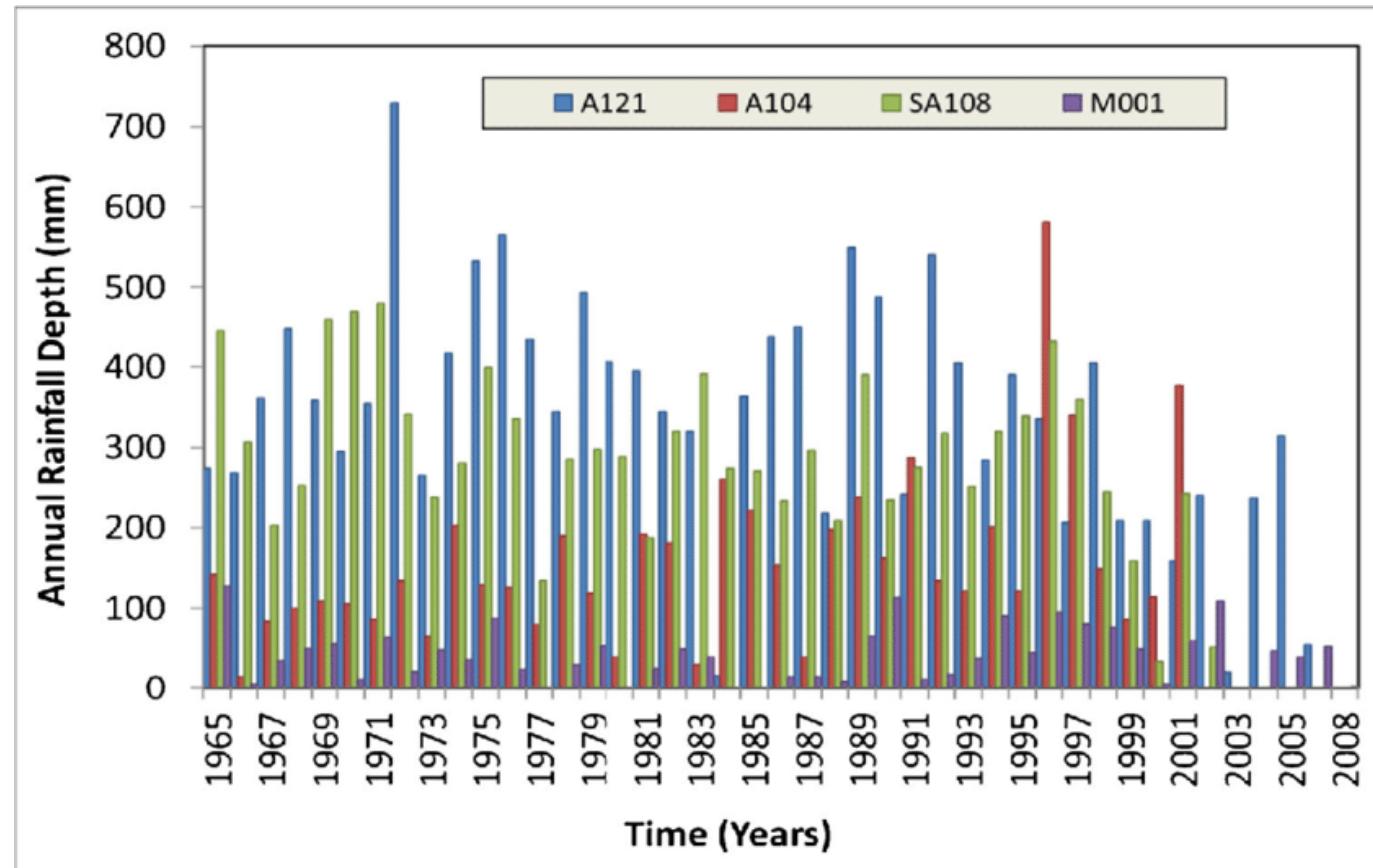


Figure 6. NLDAS-2 Primary Forcing time series of hourly precipitation from 13Z 01 Jan 1979 to 12Z May 2023 at 33.9375 North and -86.9375 West. (credit: David Mocko and Youlong Xia)

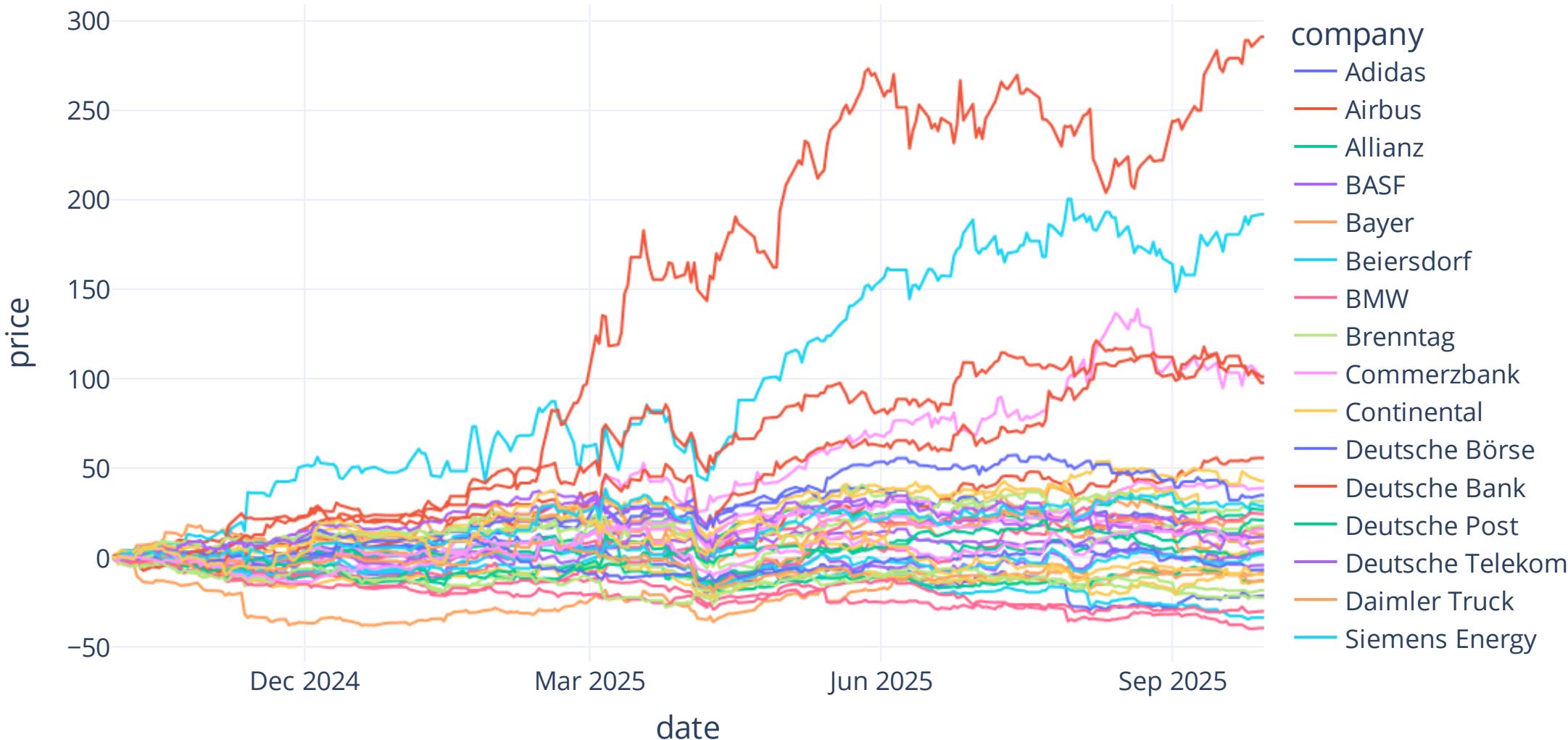
The implications are clear ...



Time series data can become hard to interpret:

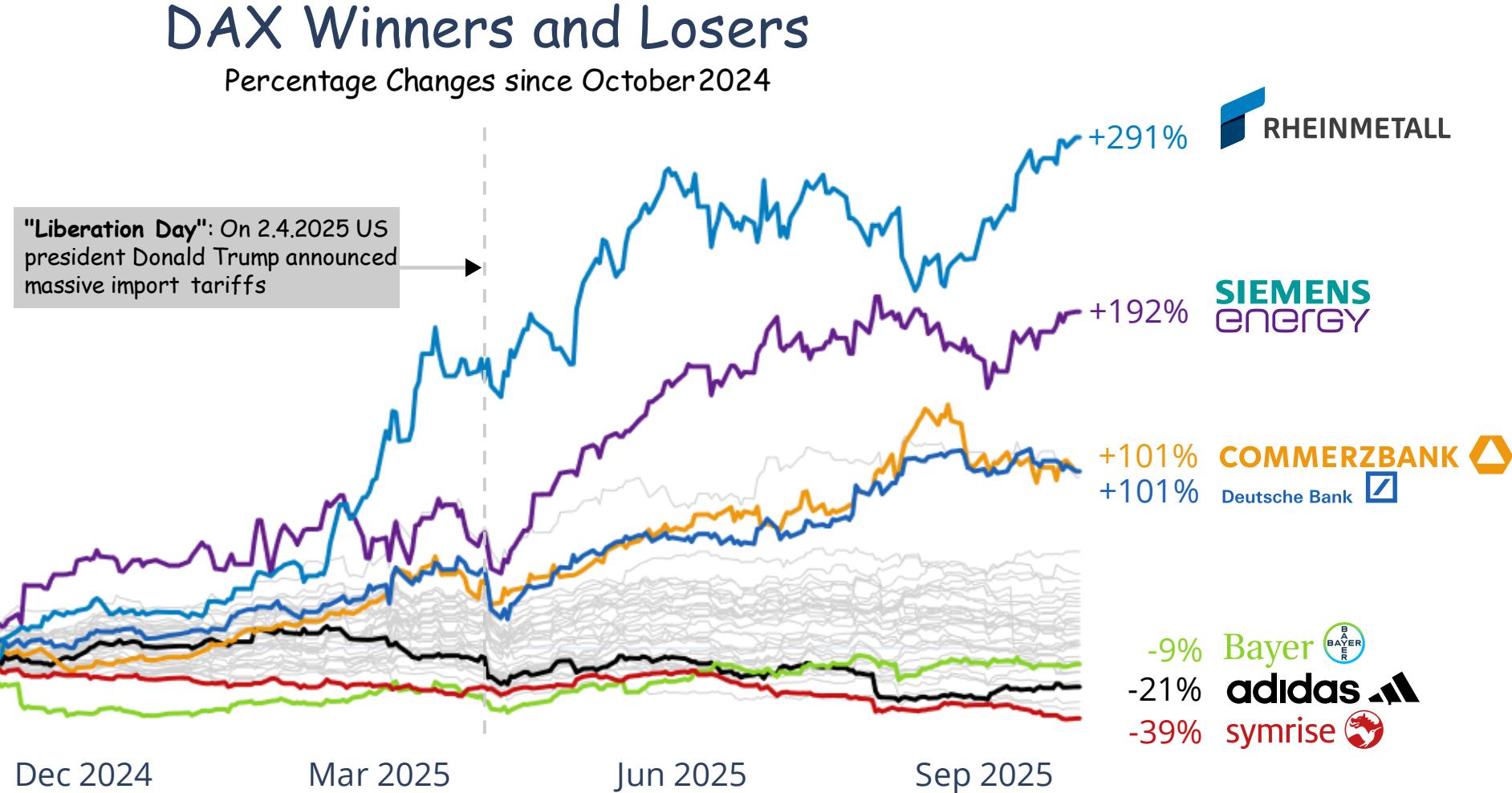
- ▶ Multiple time series
- ▶ Disaggregated data (minutely, ...)
- ▶ High fluctuations

Don't overwhelm the audience

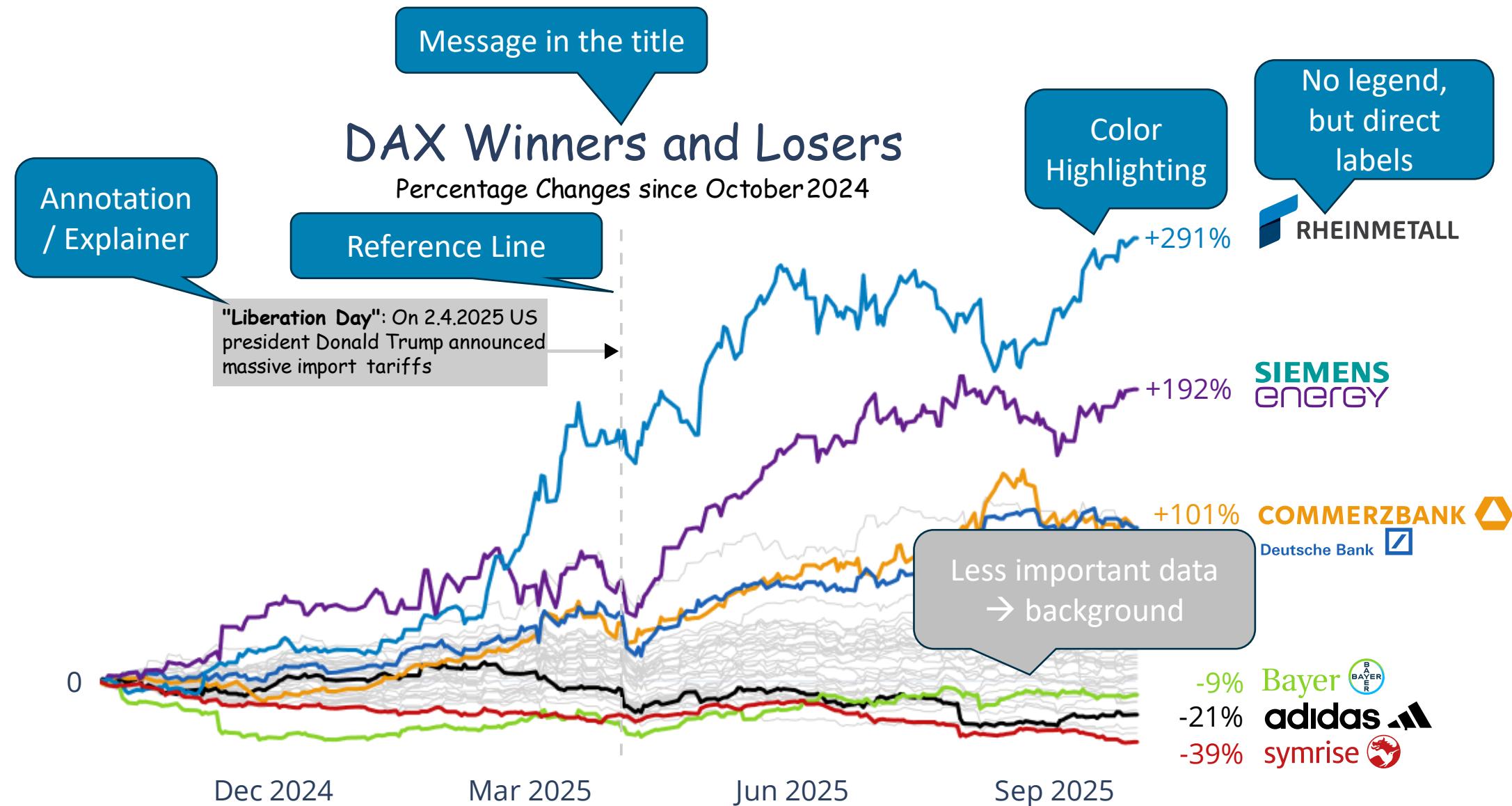


We should
make it as easy as possible for the audience
to see relevant patterns

Highlight patterns of interest



Highlight patterns of interest



Smooth lines

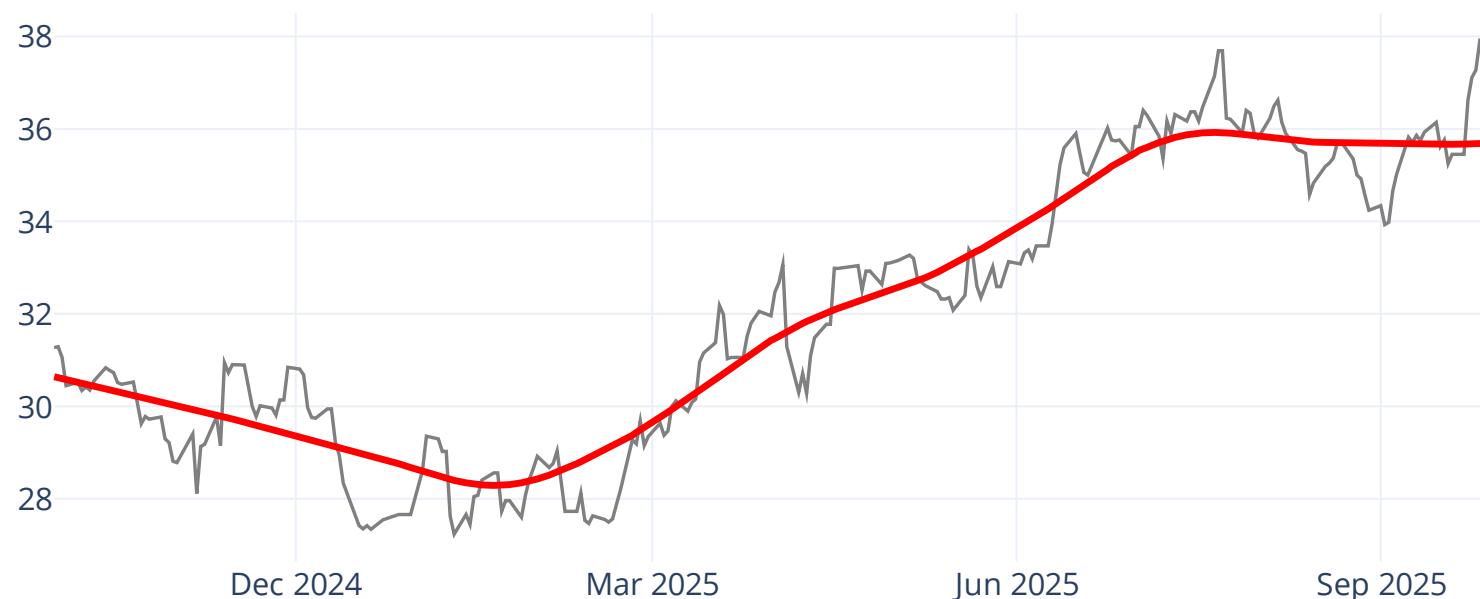
- ▶ Calculate the **moving average** over 7, 14, 30 days to abstract away extreme fluctuations



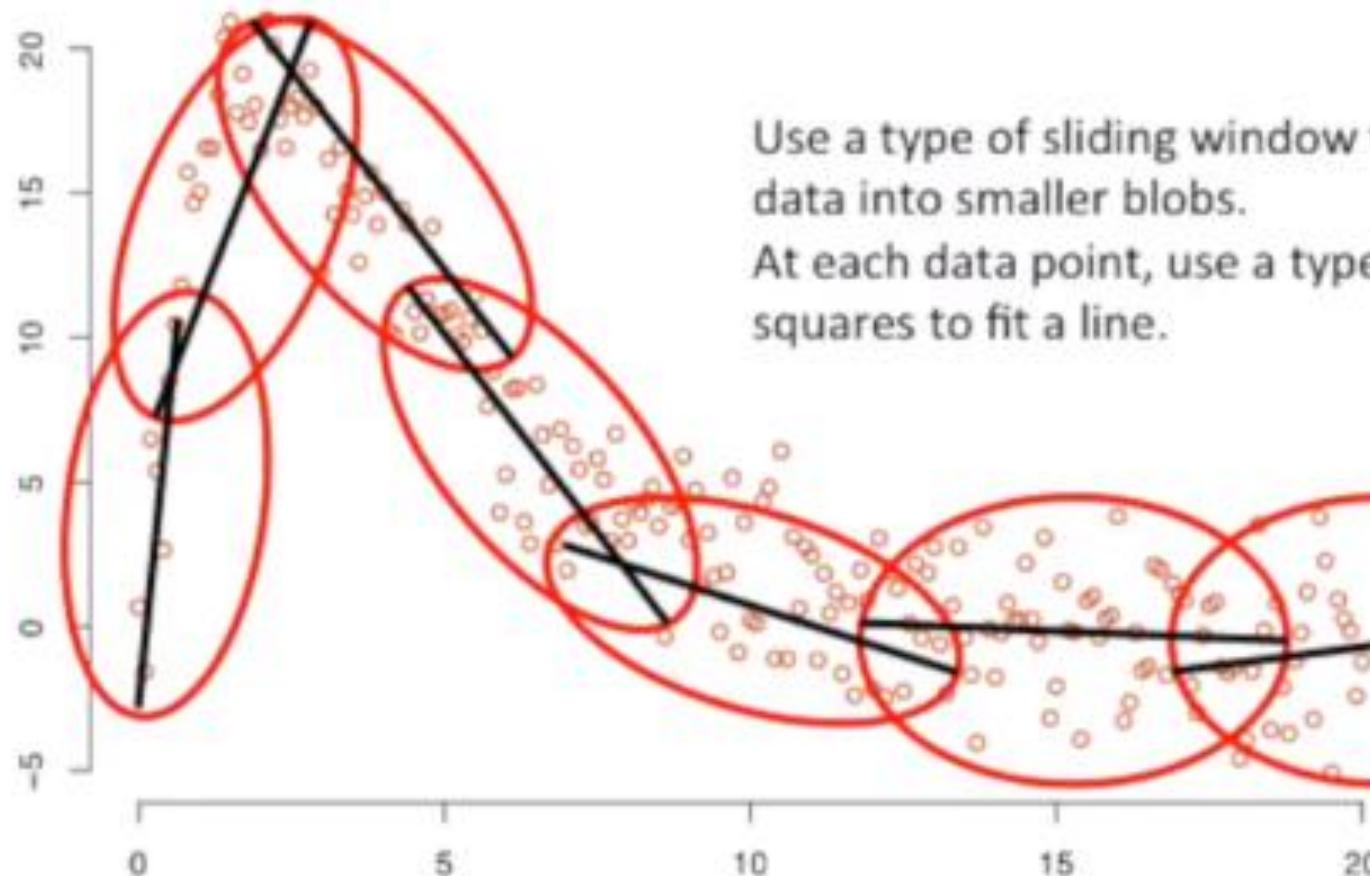
Smooth lines

- ▶ Locally estimated scatterplot smoothing (Lowess) is an alternative smoothing form
- ▶ You can tune the fraction parameter to achieve the desired degree of smoothing
- ▶ Interpretation less intuitive than moving averages

Lowess smoothing with a fraction of 0.3



Lowess



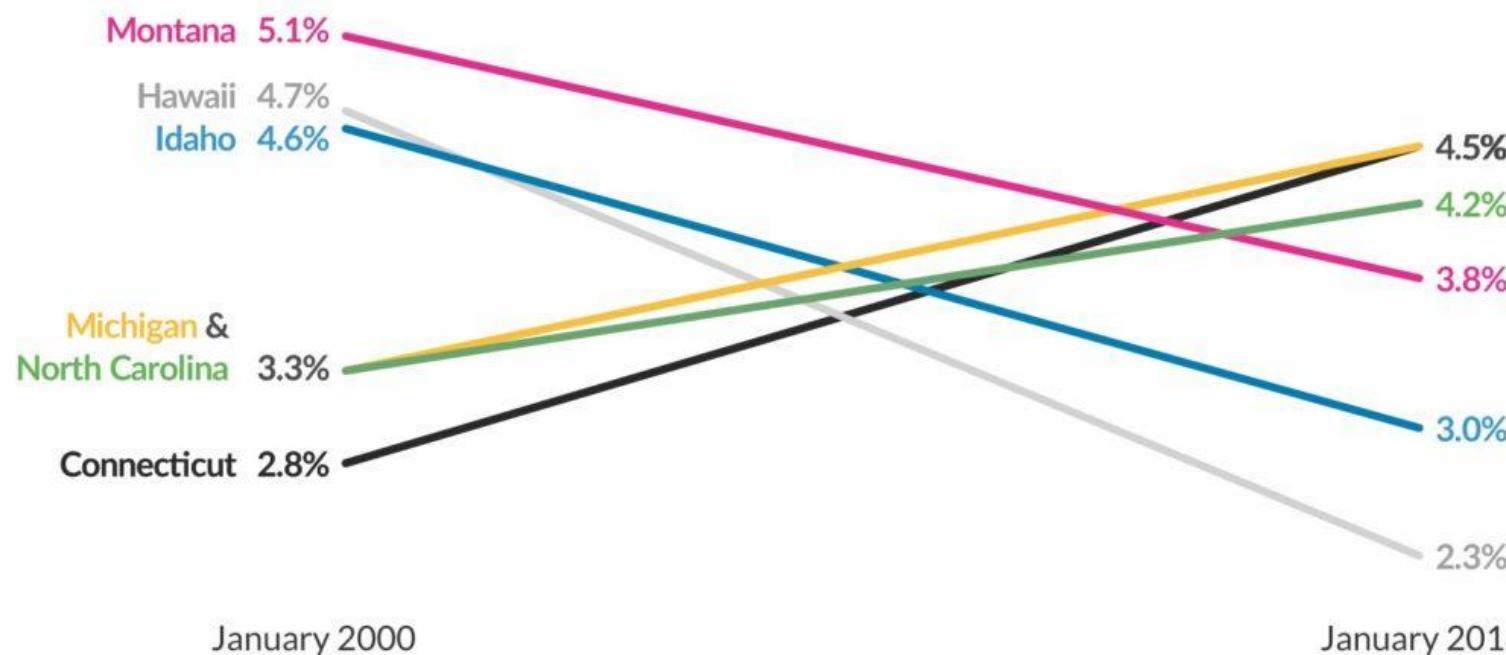
Use a type of sliding window to divide the data into smaller blobs.
At each data point, use a type of least squares to fit a line.

Source: <https://www.andreaperlato.com/theorypost/fitting-a-curve-to-data-lowess-and-loess/>

Slopechart: show only start and end

If the story is about the **change between** some **start and end point**, then we may ignore the fluctuations in between

Biggest changes in the unemployment rate between, January 2000 and January 2018



Analyse Seasonalities

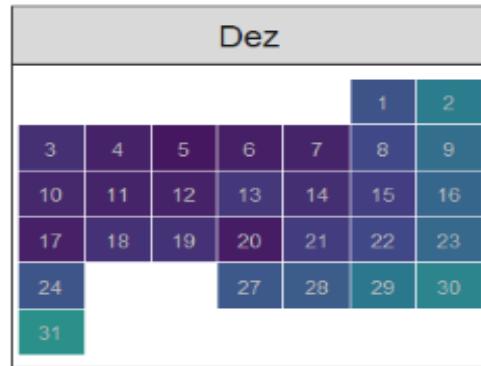
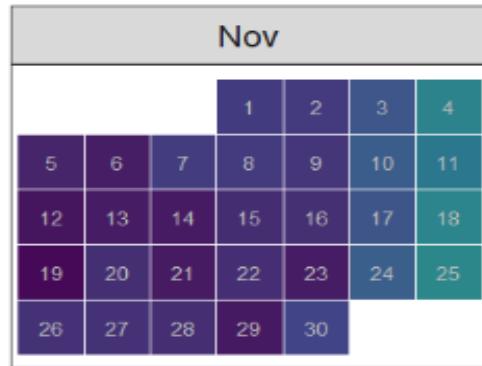
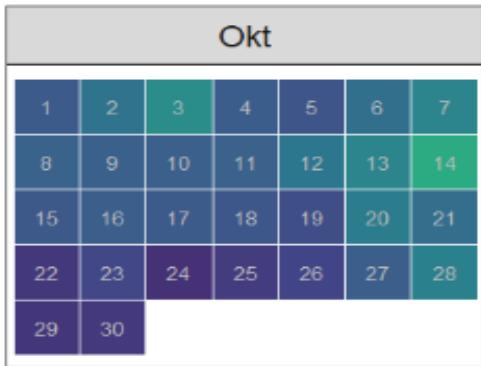
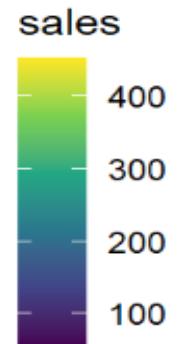
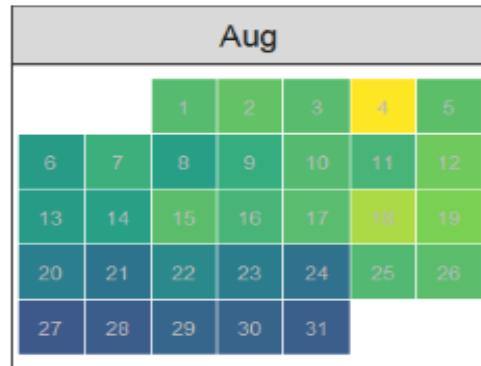
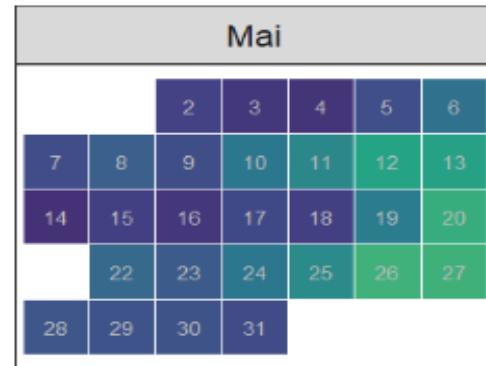
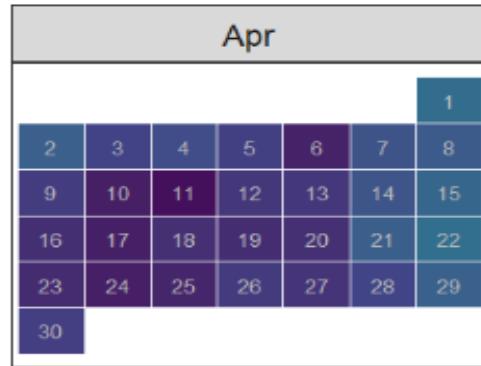
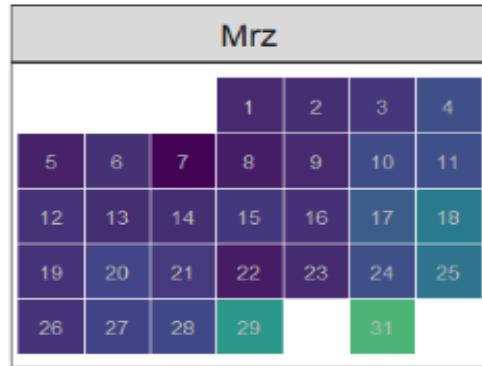
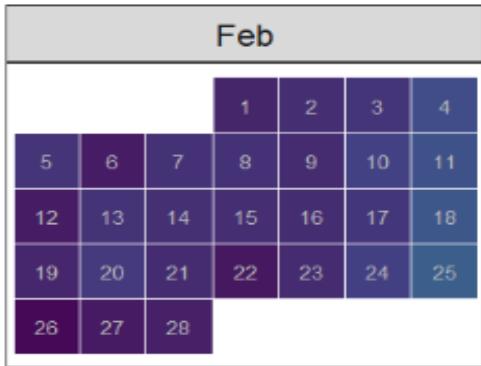
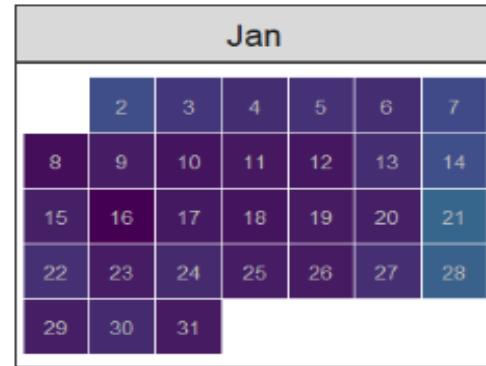
Time Series data is often noisy due to the presence of multiple forms of seasonalities and trends. How can interesting patterns be identified?

- ▶ Focus on a single time series
- ▶ Zoom into short time ranges
- ▶ Aggregate the data to some higher level: hourly, daily, weekly, monthly, etc.

Use domain knowledge and common sense!

There won't be day-of-week patterns on the stock market, but maybe in the buying behaviour of customers

Seasonal, monthly, weekly patterns



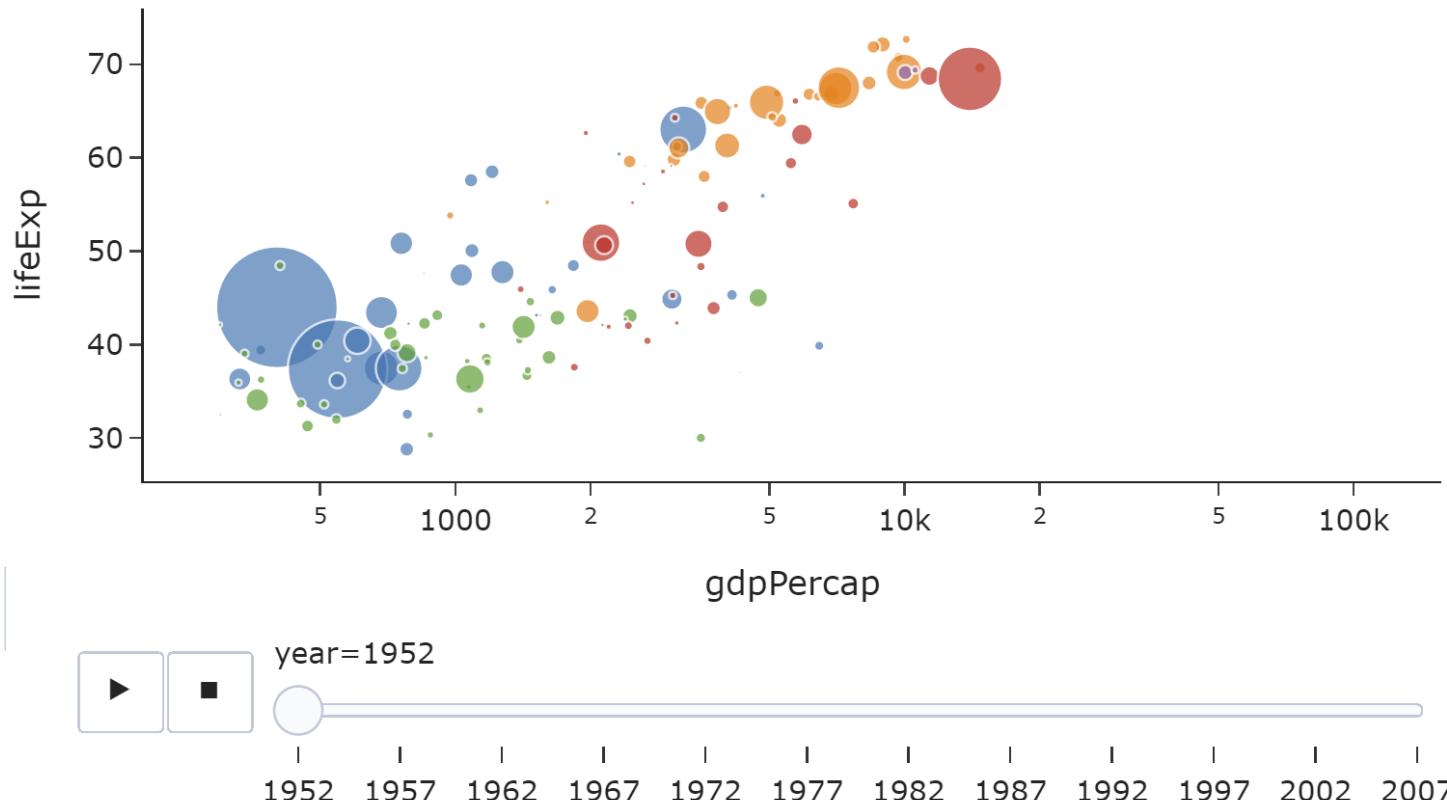
Mo Di Mi Do Fr Sa So



Adding a time dimension
onto an existing visualization

Animations

GDP per Capita vs Life Expectancy Over Time (Animated)



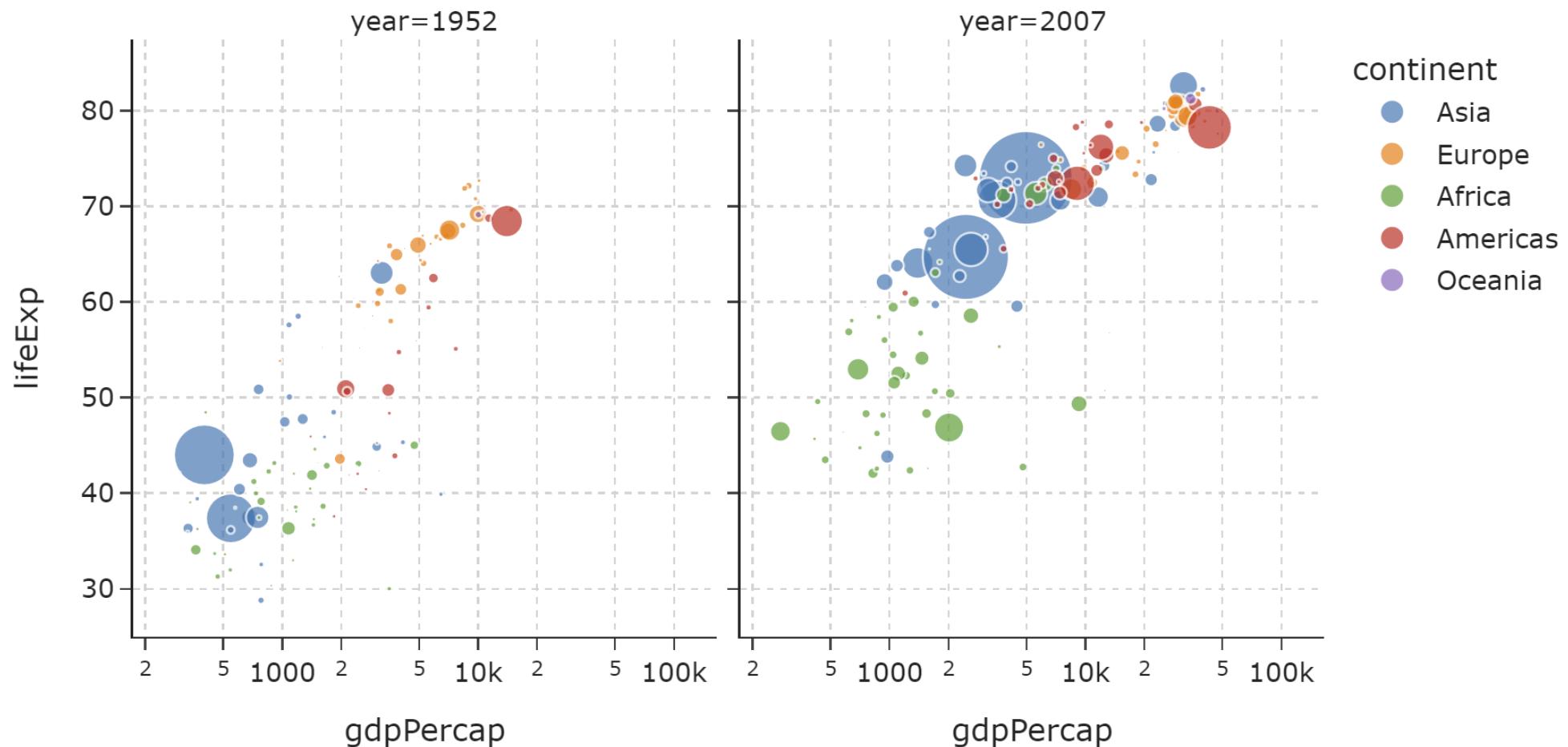
- continent
- Asia
 - Europe
 - Africa
 - Americas
 - Oceania

Animations often carry too much information, such that people feel overwhelmed.

They need to be **extremely well designed**

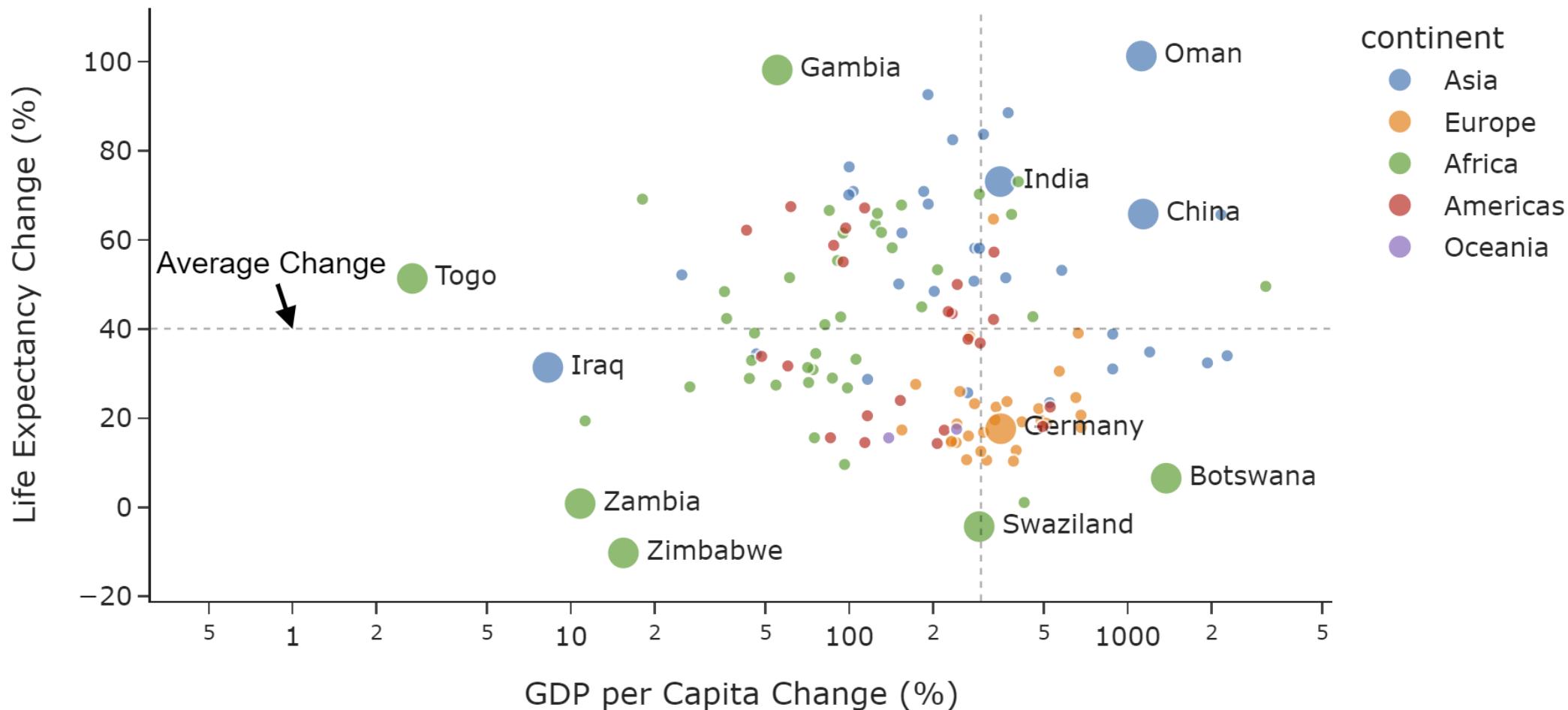
See: <https://plotly.com/python/animations/>

Facets: showing changes

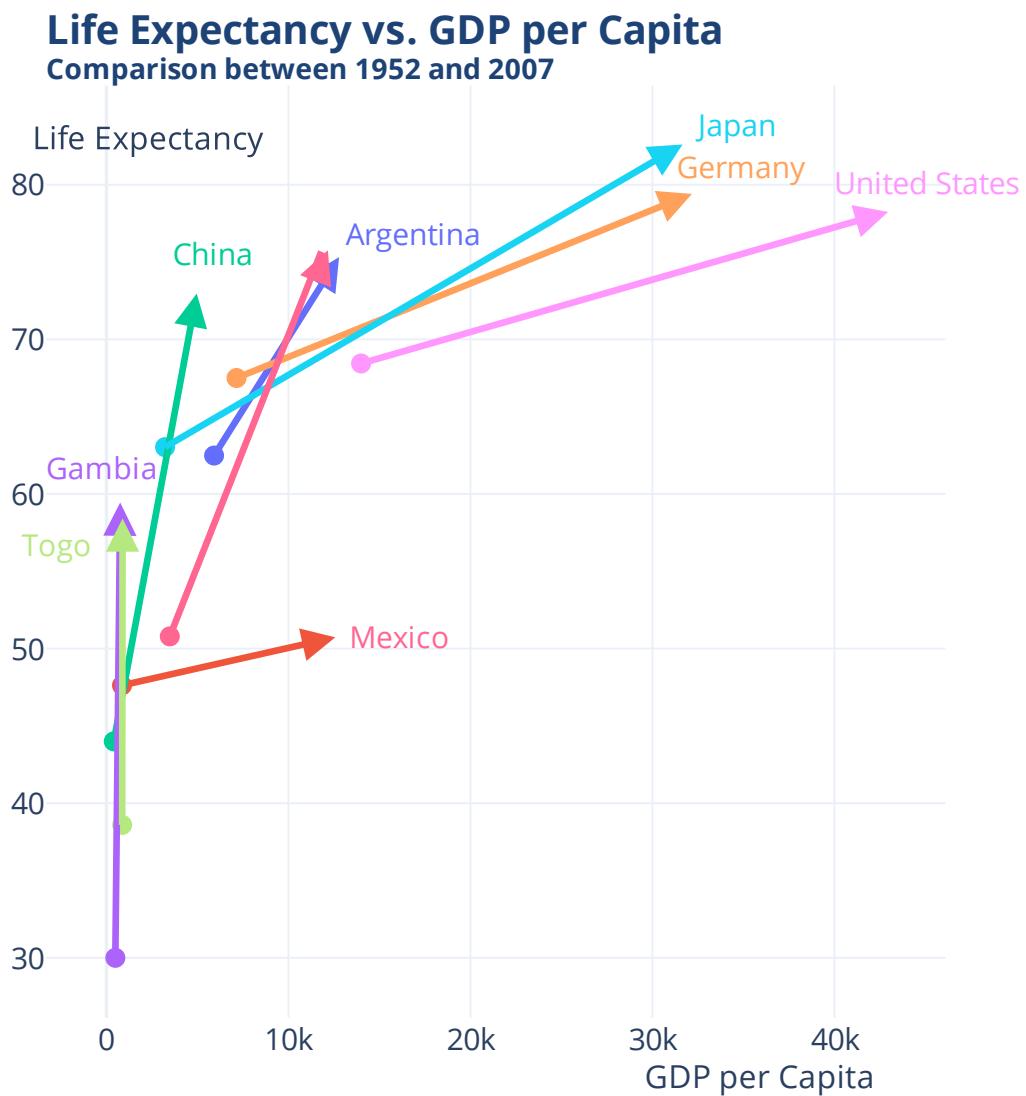


Scatterplot of changes (%)

How GDP per Capita and Life Expectancy Changed Between 1952 and 2007



Connected Scatterplot



Choropleth map of changes

