

A journey through time

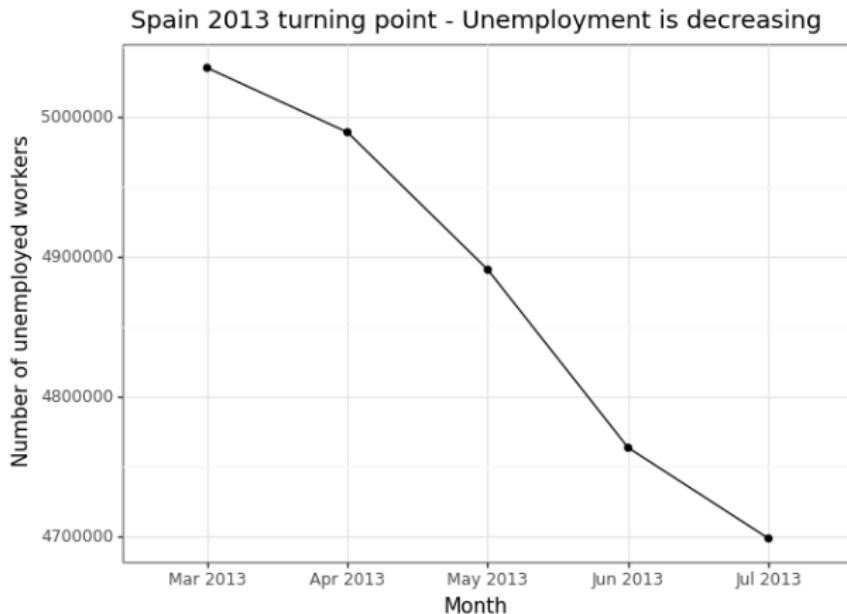
MVE080/MMG640 Lecture 4

Sebastian Persson
sebpe@chalmers.se
November 9, 2022

Time-series

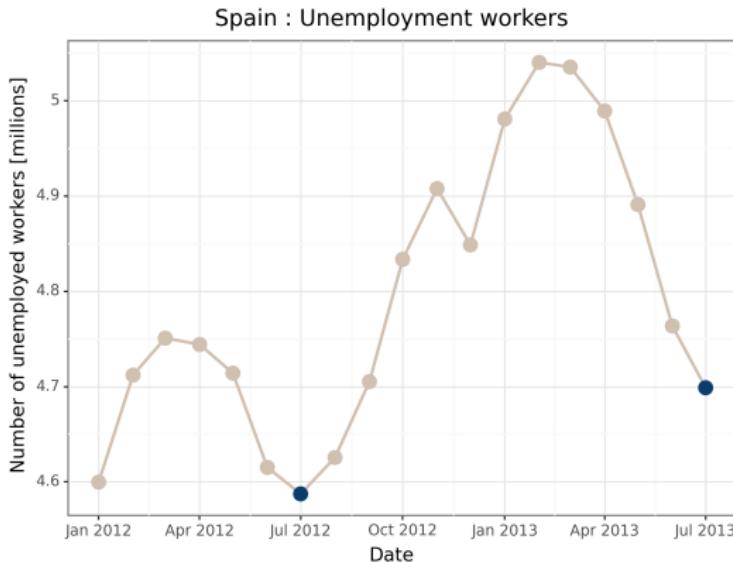
From last lecture ...

- ▶ Why is this a bad visual?



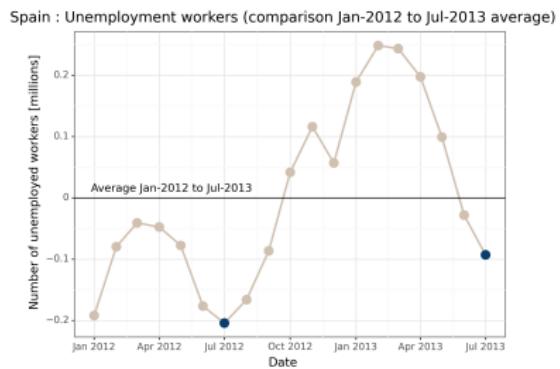
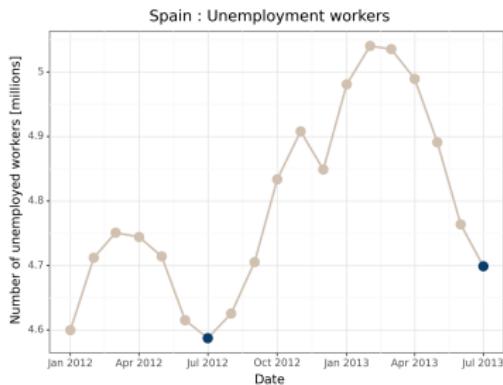
The context is important

- ▶ If you torture the data your data long enough, nature will always confess - R. Coase
- ▶ Spain is a tourist country



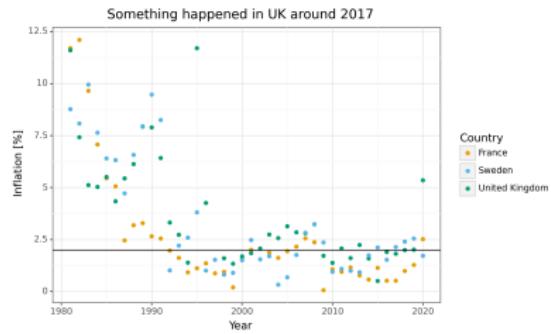
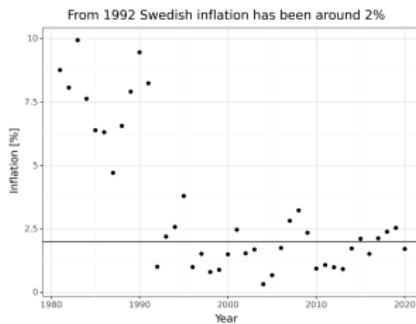
The context is important

- ▶ Always include sufficiently long timespan to be truthful
- ▶ Transforming the y-axis can ease interpretation
 - ▶ Annotations often helps



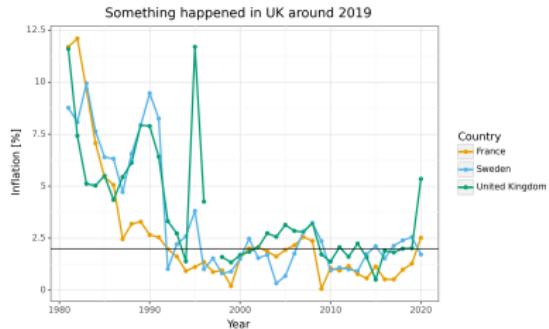
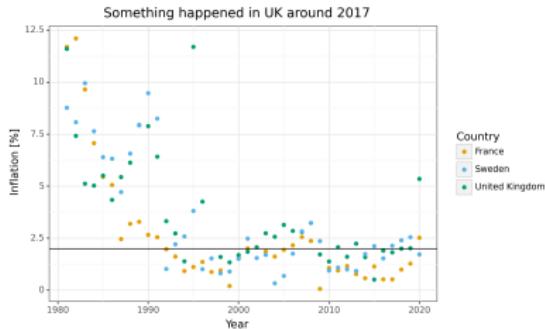
How do we plot time-series data?

- ▶ For a single time-series we can use dots or lines
- ▶ Do dots work for multiple time-series?



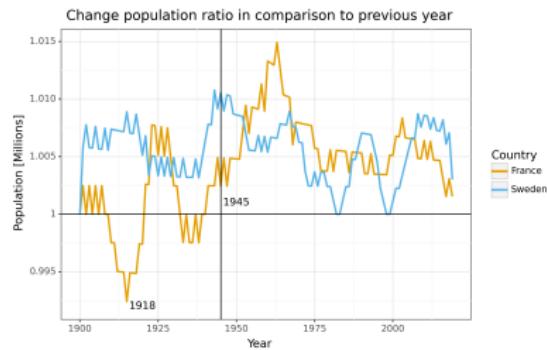
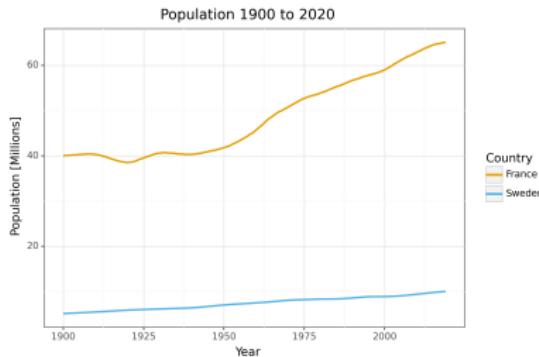
How do we plot time-series data?

- ▶ For multiple time-series use lines
 - ▶ If we remove the points emphasise is on the trend
- ▶ Use empty space for missing values (default plotnine)



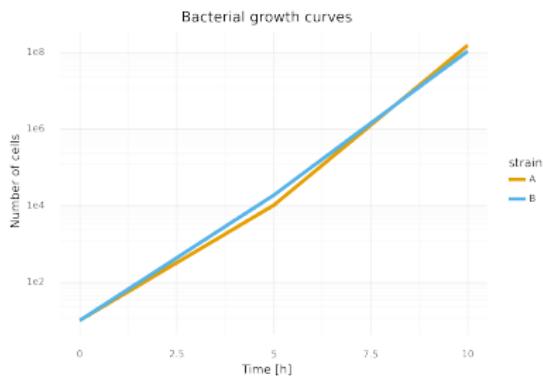
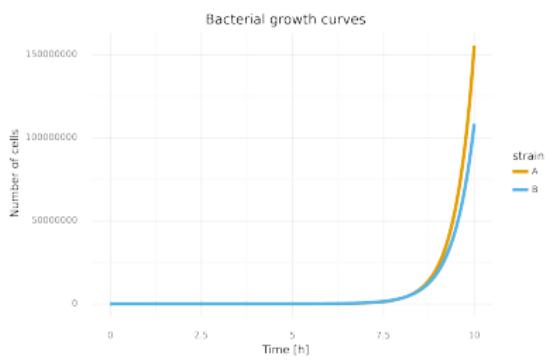
Rate of Ch-ch-ch-ch-Changes

- ▶ Which years did France's population decrease?
- ▶ Which years did Sweden's population decrease?
- ▶ Ratio of change $R_i = N_i/N_{-1}$



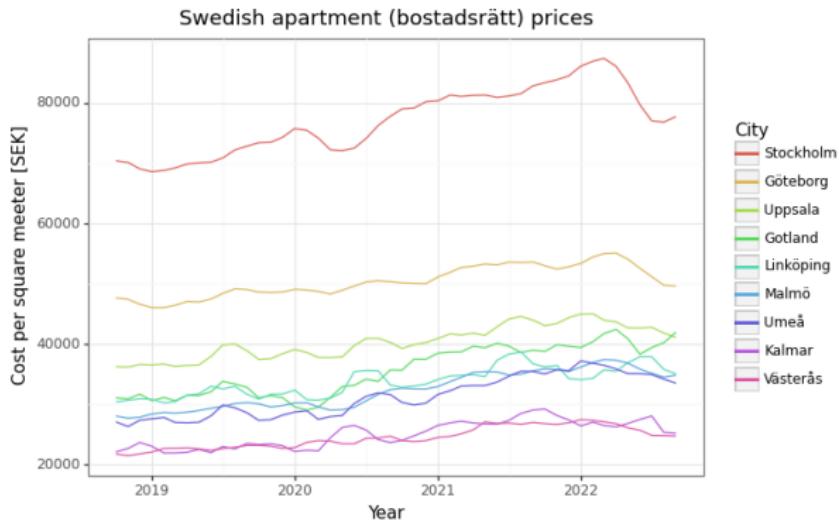
How do we handle exponential growth?

- ▶ Which bacterial strain has the fastest growth rate?
- ▶ Use log-scale for exponential growth
- ▶ For time-series often needed to compute index or transform the y-axis to discern trends



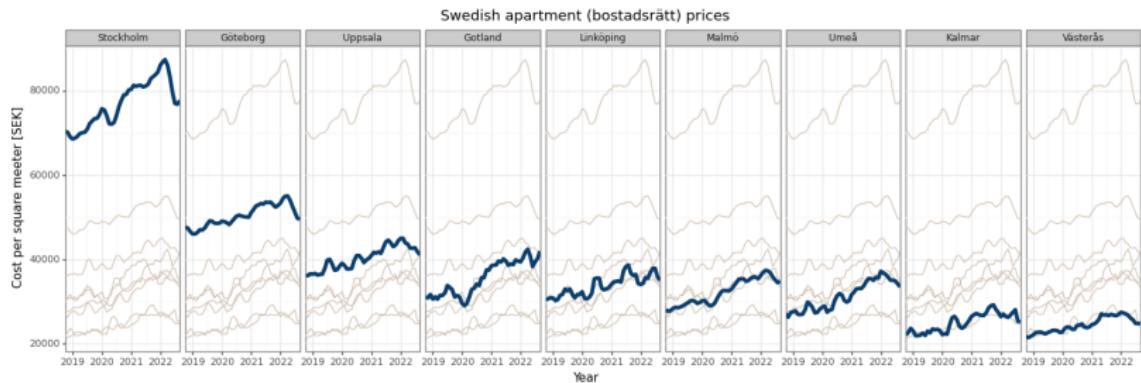
How do we handle many categories?

- ▶ Do you find this plot easy to read?
- ▶ Typically avoid more than 8 colours



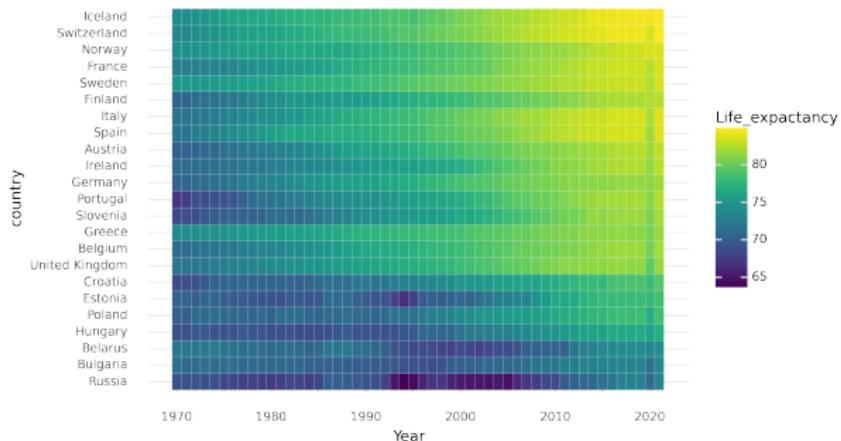
How do we handle many categories?

- ▶ Facets useful for handling many categories
- ▶ Adding background data → context
 - ▶ You are here plot



How do we handle MANY categories?

- ▶ For MANY¹ heatmaps are usefull
- ▶ Do you think there is any better way we can visualise the data?

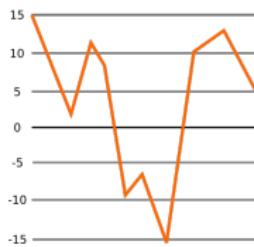


¹MANY = many²

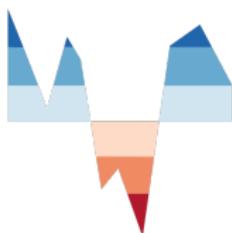
The horizon plot for large scale data

- ▶ Combines both colour and magnitude

Original



1



2

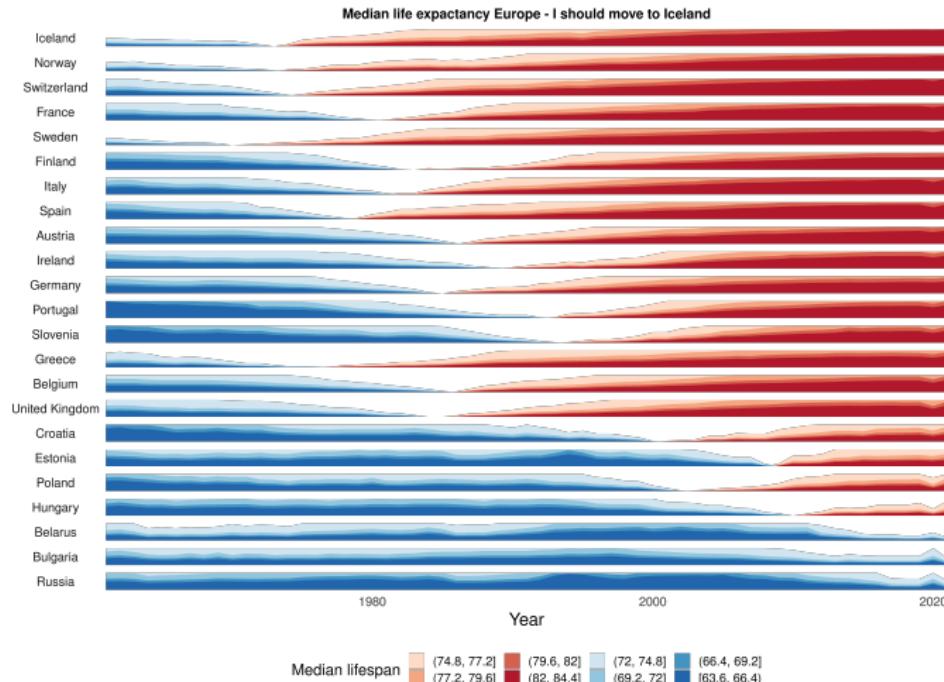


3



The horizon plot for large scale data

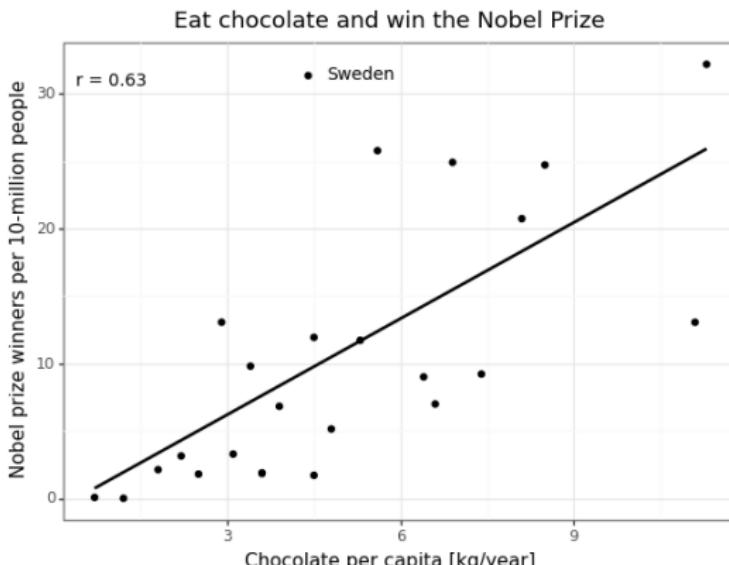
- ▶ Suitable for time-series data with MANY categories
- ▶ Drawback - Training required to read



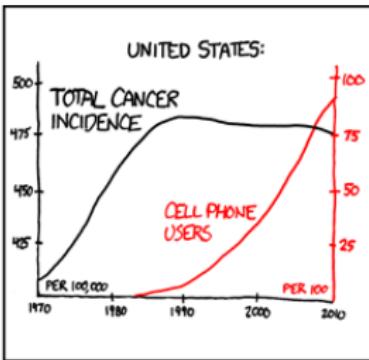
Associations

Breaking News : We should buy Tony's

- ▶ What is the problem with this visual?
- ▶ Correlation does not imply causation
 - ▶ tylervigen.com/spurious-correlations

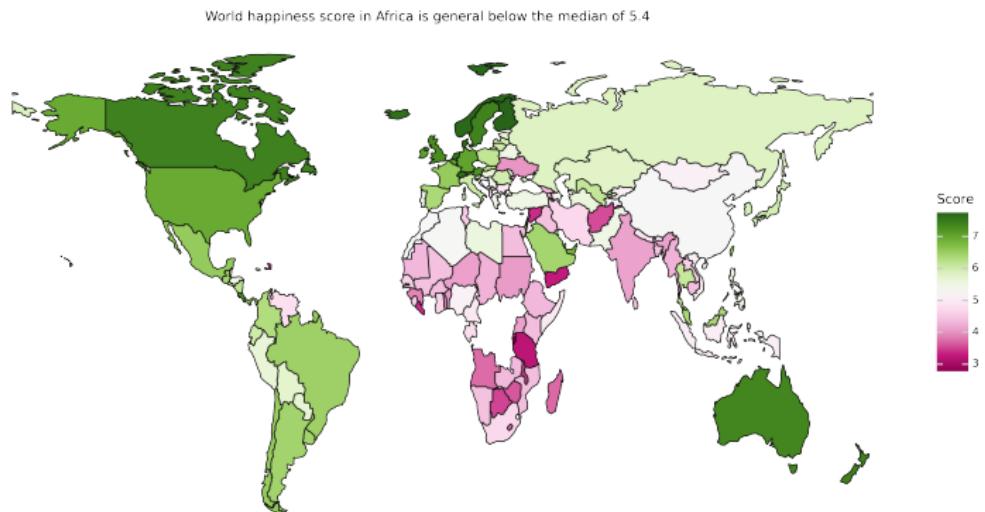


Proving causation is hard!



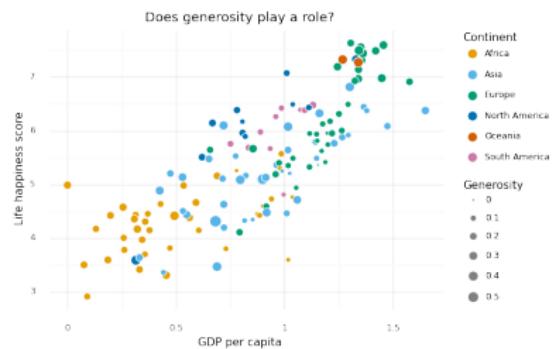
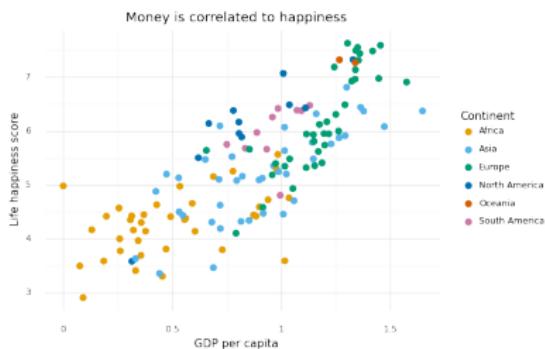
Why is Finland so happy?

- ▶ Let us see if we can find any relationships



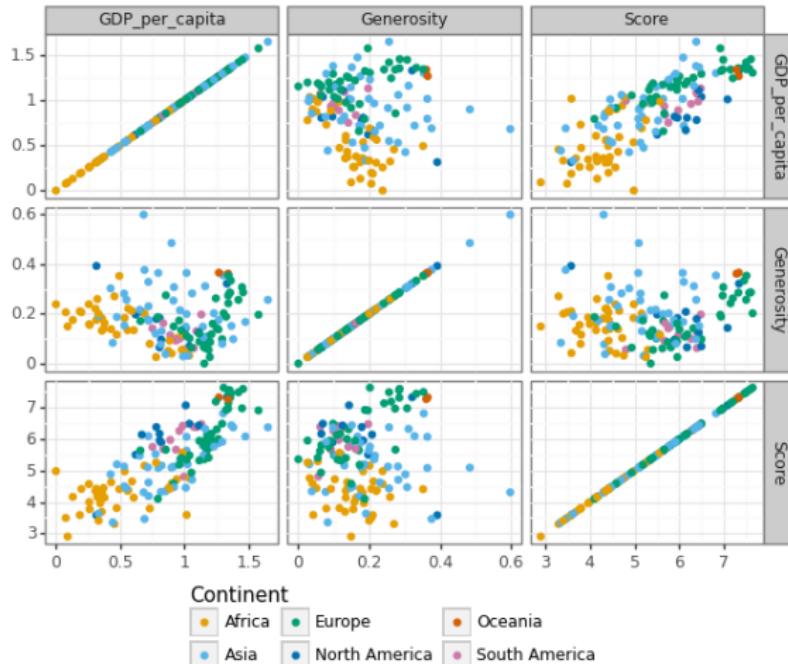
Scatter plot - The working horse of associations

- ▶ Great tool for comparing two continuous variables against each other
- ▶ How do we compare three variables?



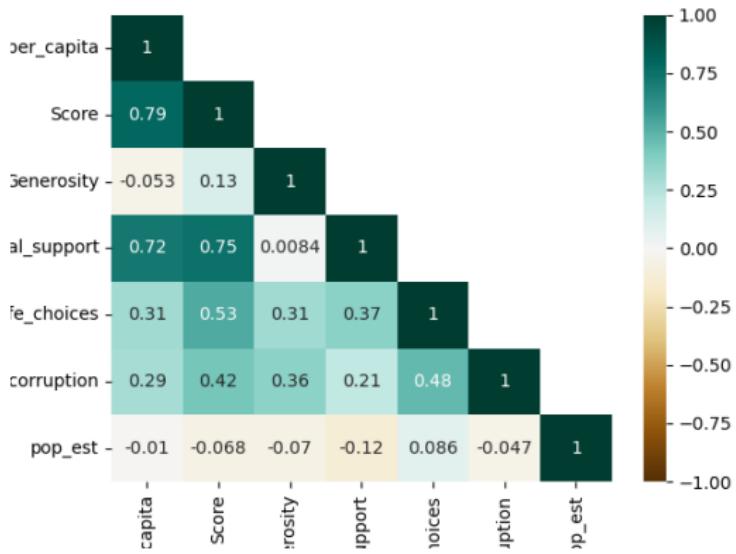
The scatter matrix

- ▶ Great for wide overviews
 - ▶ Packages available in R and Python



Correlation heatmap

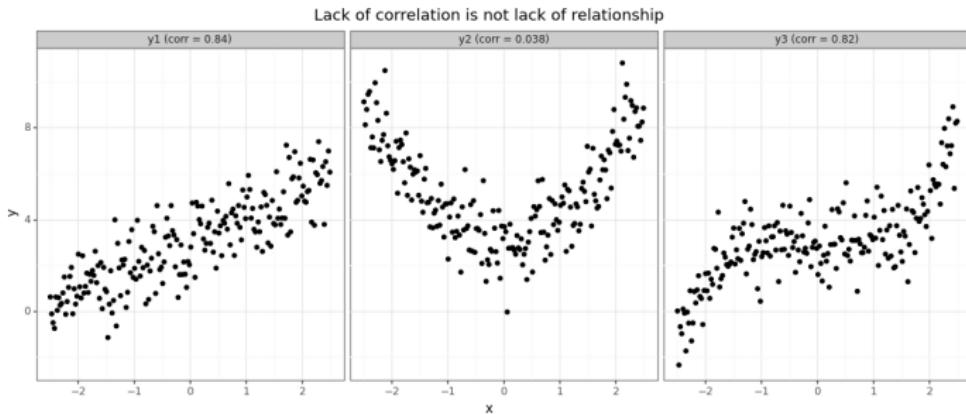
- ▶ Overview for large datasets
- ▶ Do you see any problems?



Correlation is not a perfect association measure

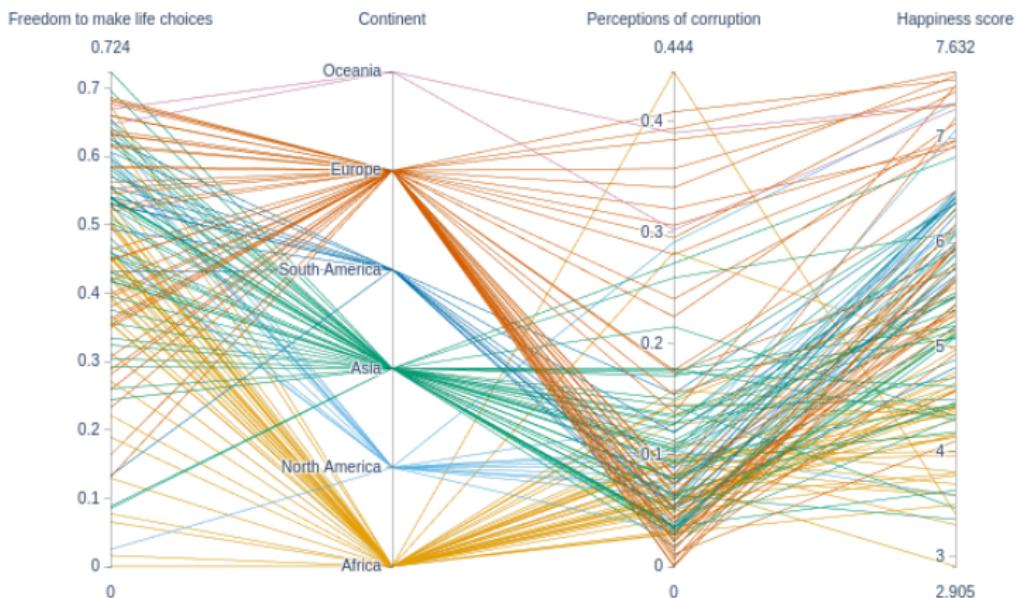
- Correlation ρ measures co-linearity;

$$\rho = \text{corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



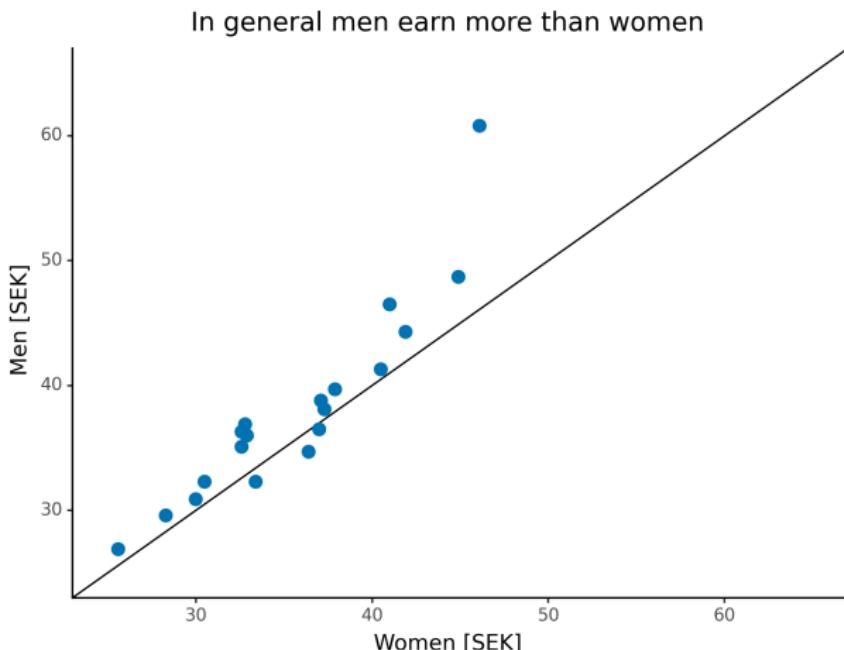
A newer approach - Parallel coordinates

- ▶ Can handle both categorical and continuous data
- ▶ Works best with few datapoints



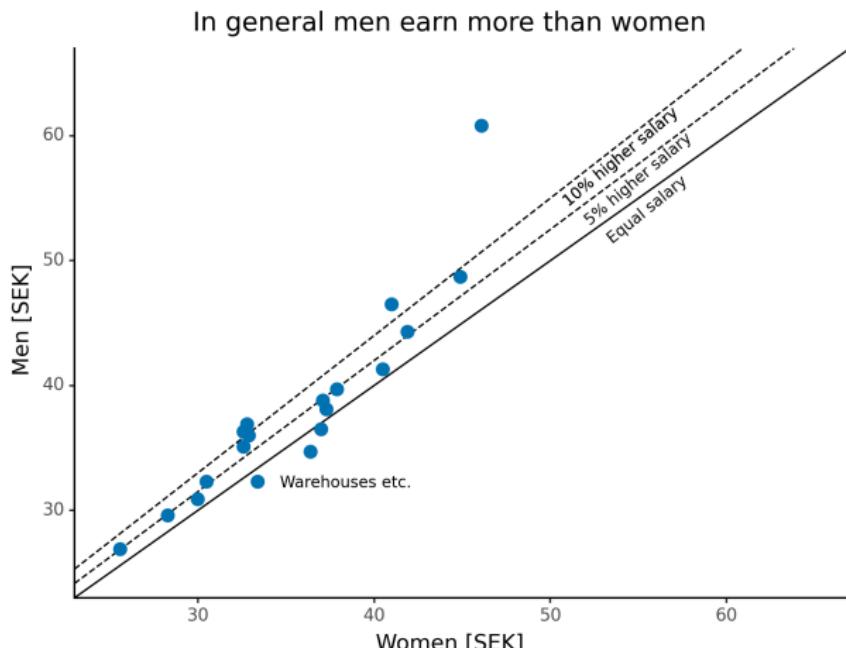
Special case - paired data

- When observing the same variable under slightly different conditions



Special case - paired data

- ▶ Adding more than one line can be extremely helpful
 - ▶ Do not forget the annotation



Take home messages

- ▶ For time-series
 - ▶ Do not forget to account time-dependent effects such as seasonality
 - ▶ Transforming the data and/or computing an index often eases interpretation
 - ▶ Horizon plots work well for large-scale data
- ▶ Correlation does not mean causation, and correlation is neither a perfect measure of association (if possible use scatter matrix)
- ▶ For paired data using several trend-lines (slide 22) can drastically ease interpretation