

Lecture 3a

Overview of Plot Types

[Data visualization · 1-DAV-105](#)

Lecture by Broňa Brejová

Types of variables (columns)

Categorical / qualitative

- **Nominal:** values have no fixed ordering (for example, gender, country, color)
- **Ordinal:** values are ordered (for example education level primary / secondary / university; star ranking 0-5)

Numerical / quantitative

- **Discrete:** typically counts
- **Continuous:** typically measurements

Types of variables (columns)

Numerical / quantitative

- **Discrete:** typically counts
- **Continuous:** typically measurements

Numerical variables also categorized as follows:

- **Ratio (poměrová):** if zero means "none", and it is meaningful to compute ratios / percentages (mass, length, duration, cost, ...)
- **Interval:** does not have "true zero", we can subtract but not make ratios (temperature in degrees C, date)

Data for today

- Various country indicators downloaded from the World Bank for years 2000, 2010, 2018
- Population, area, GDP per capita, life expectancy, fertility (number of children per woman)
- Also classification into regions and income groups
- Which are categorical / numerical?

We will also use Gapminder life expectancy 1990-2020 from IO1

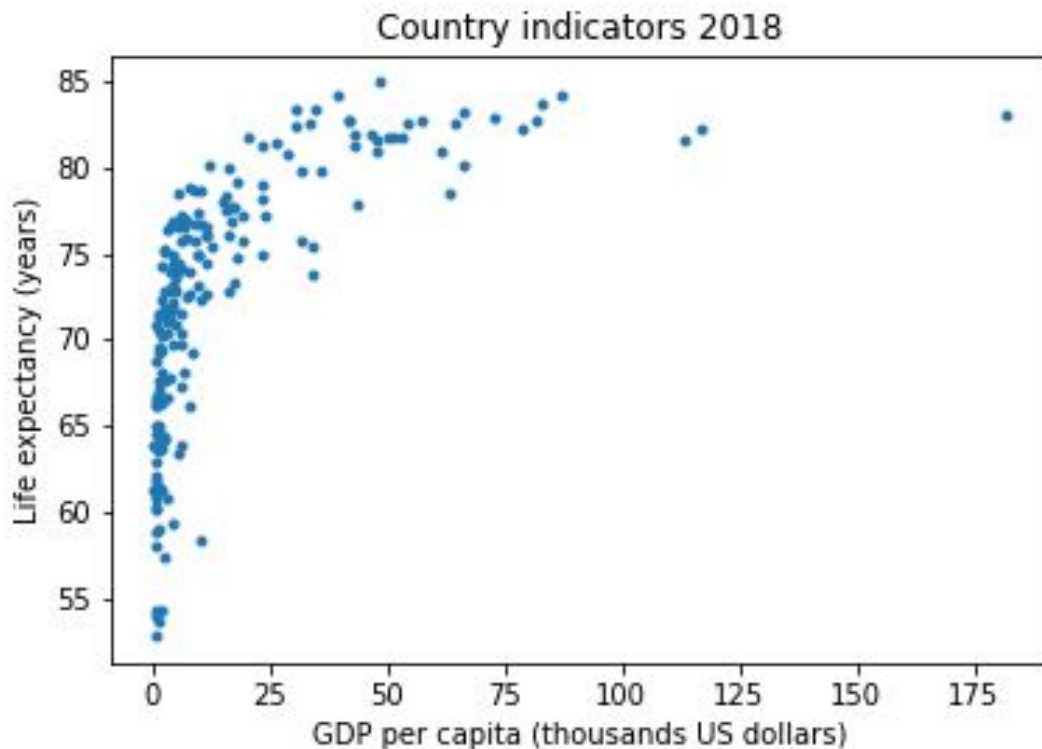
Scatter plot (bodový graf)

Good for two numerical variables (x and y).

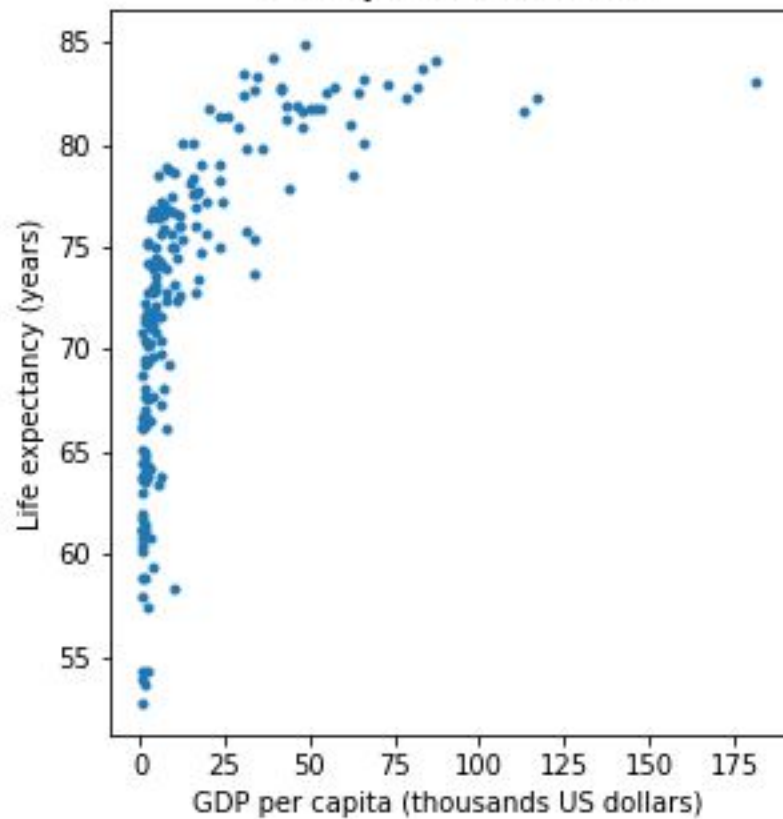
In this plot, many points near left boundary, most space empty.

Solution 1: combine
overall view and detail

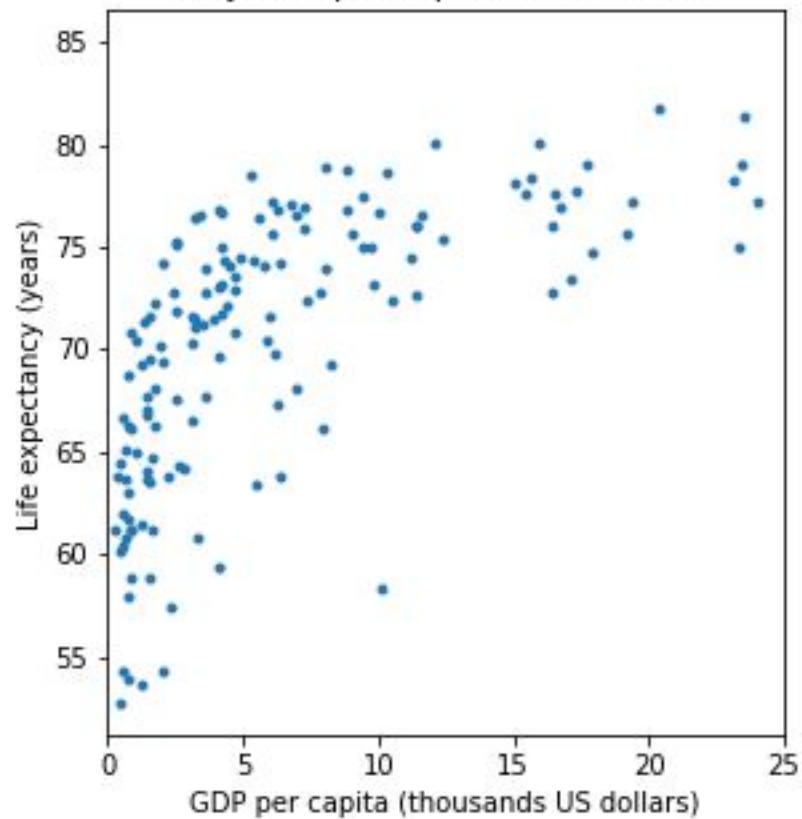
Solution 2: log scale

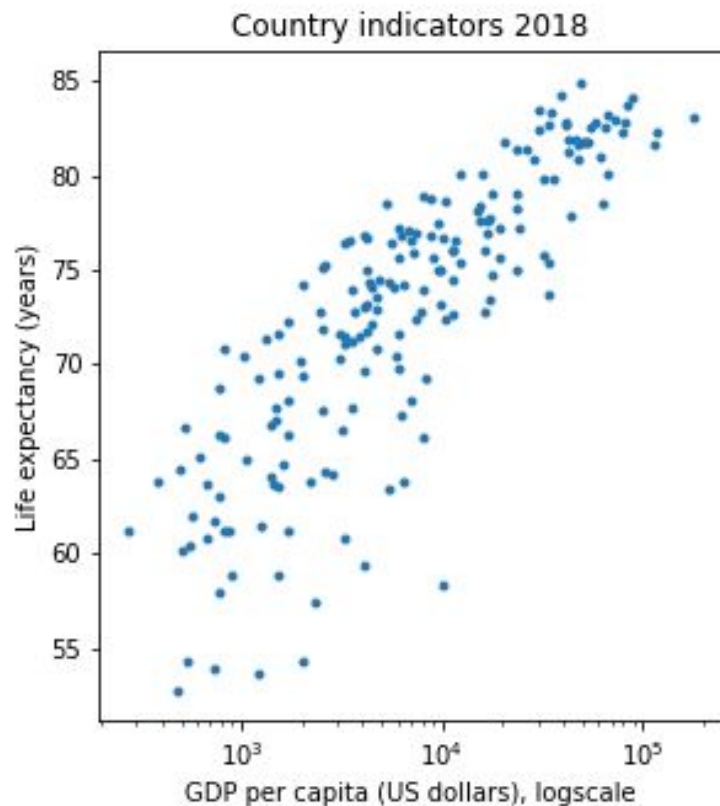
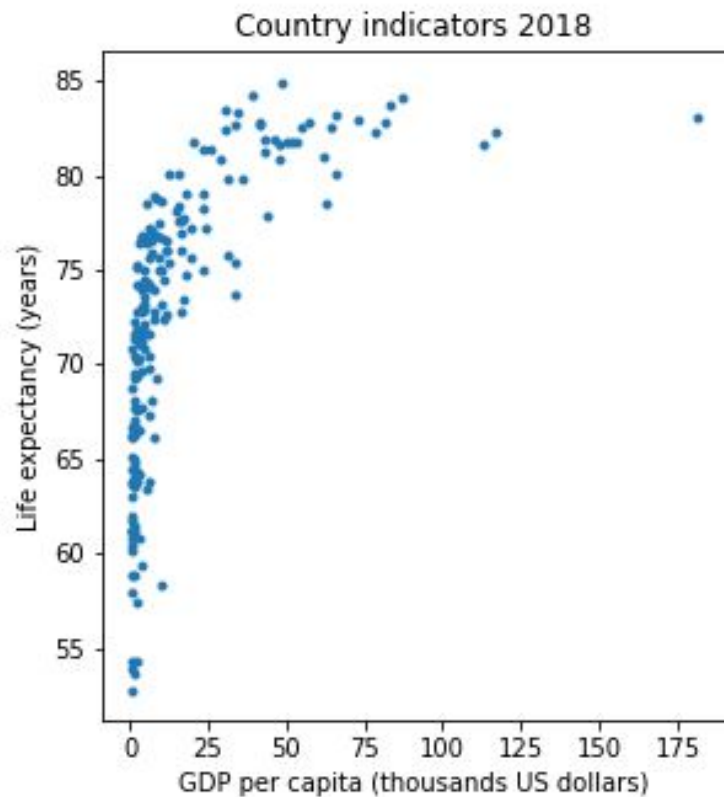


Country indicators 2018



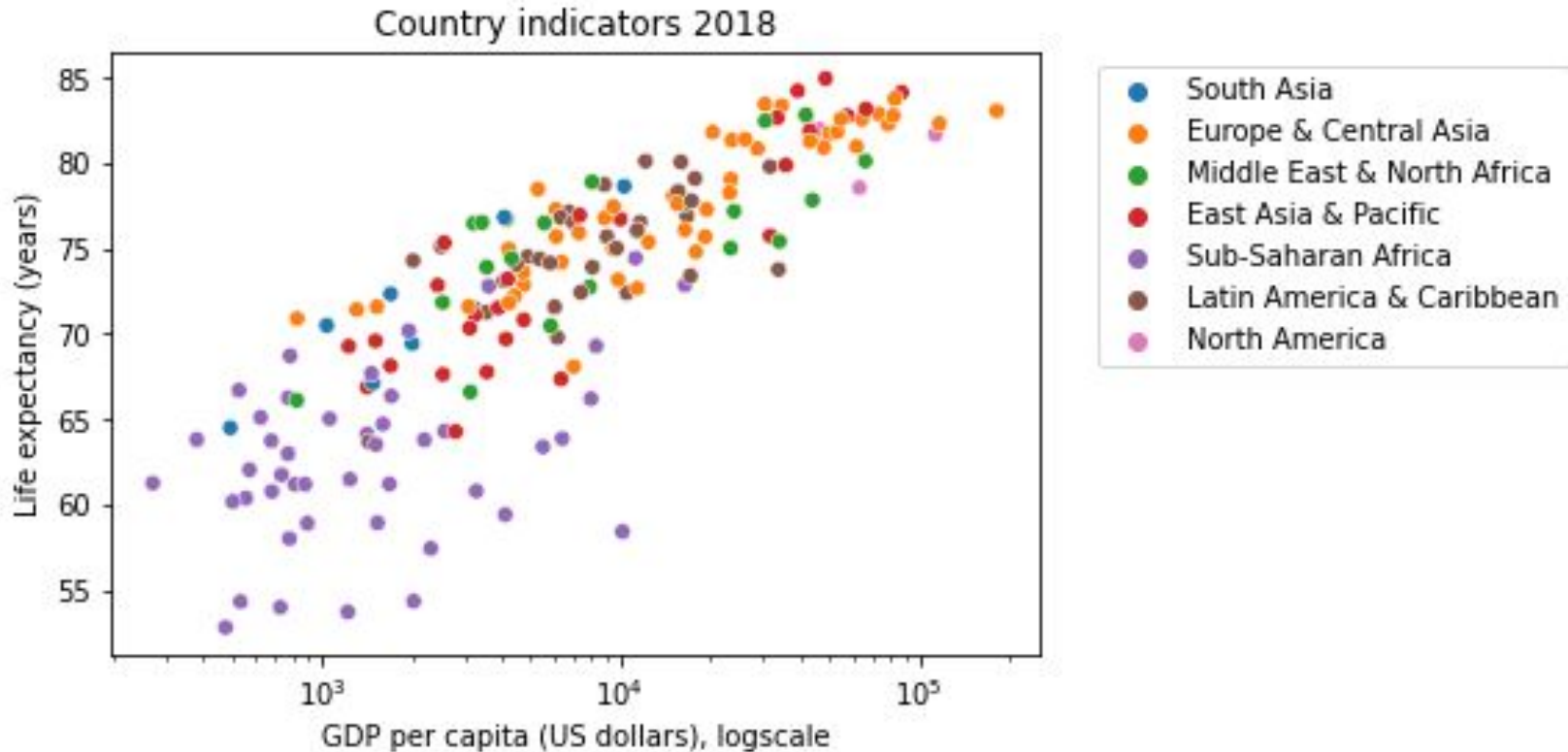
Only GDP per capita < 25000 USD



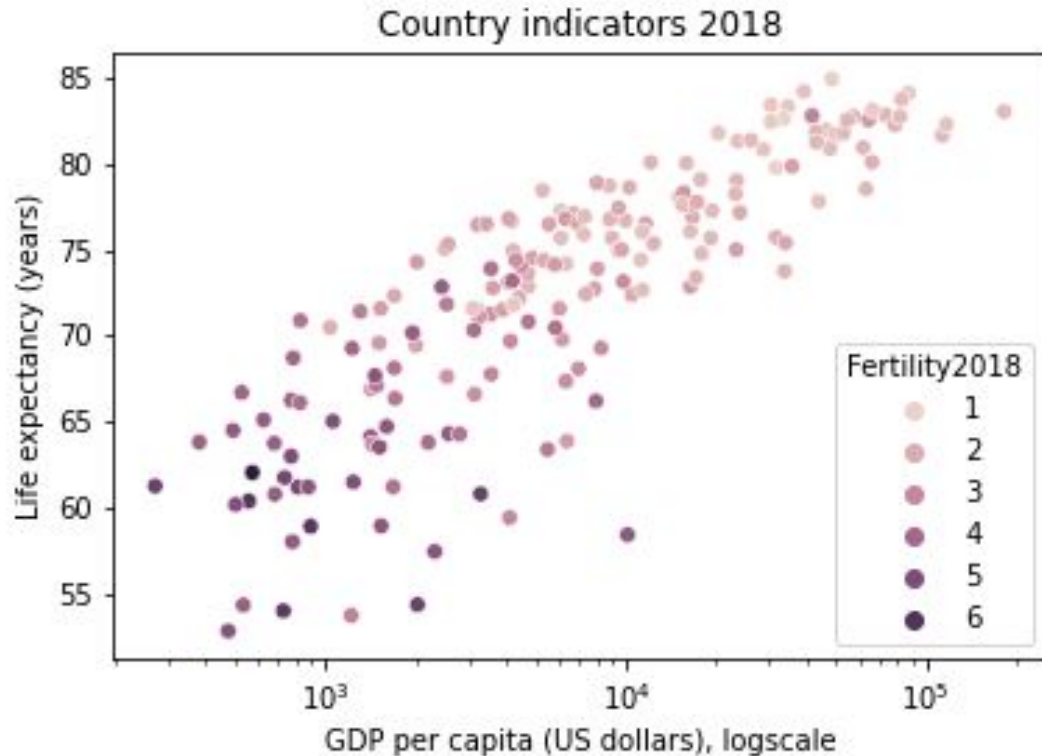


Log-scale x-axis: draw at $\log(x)$ instead of x , but axis ticks show values of x

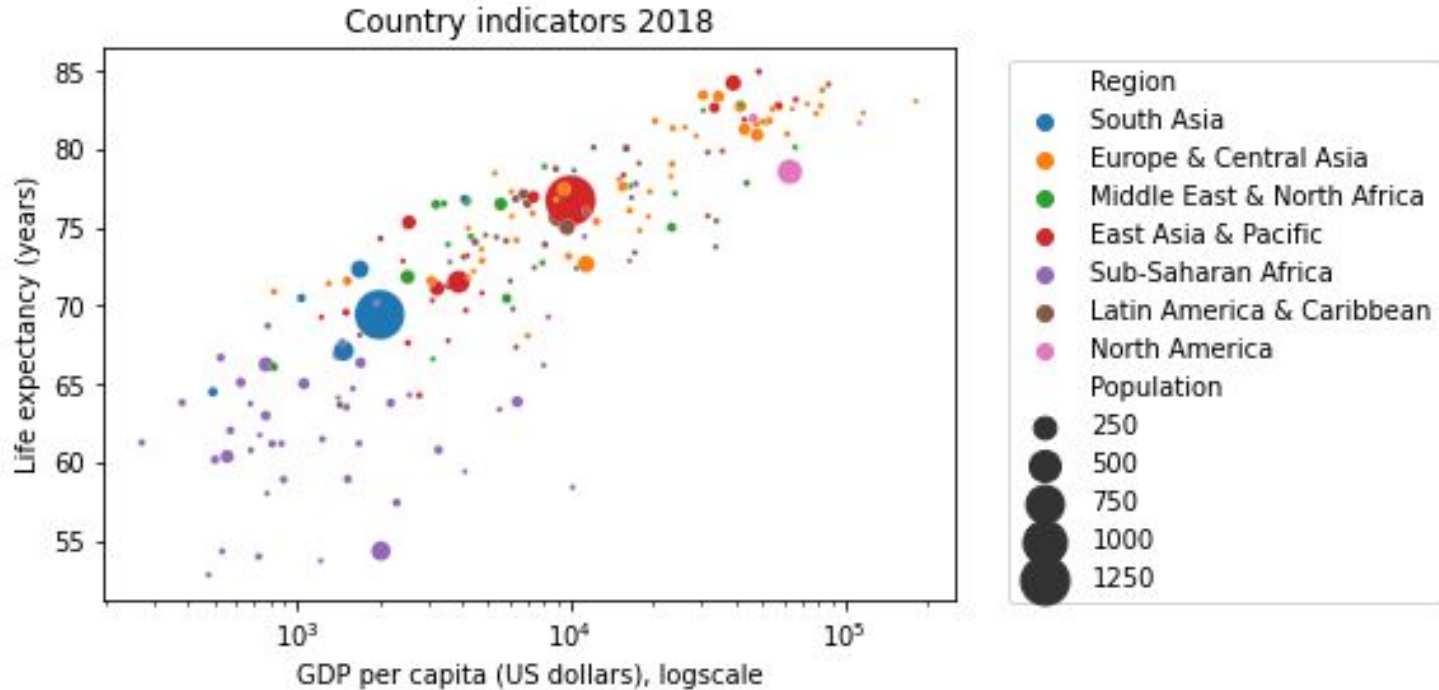
Adding categorical variable with color



Adding numerical variable with color scale

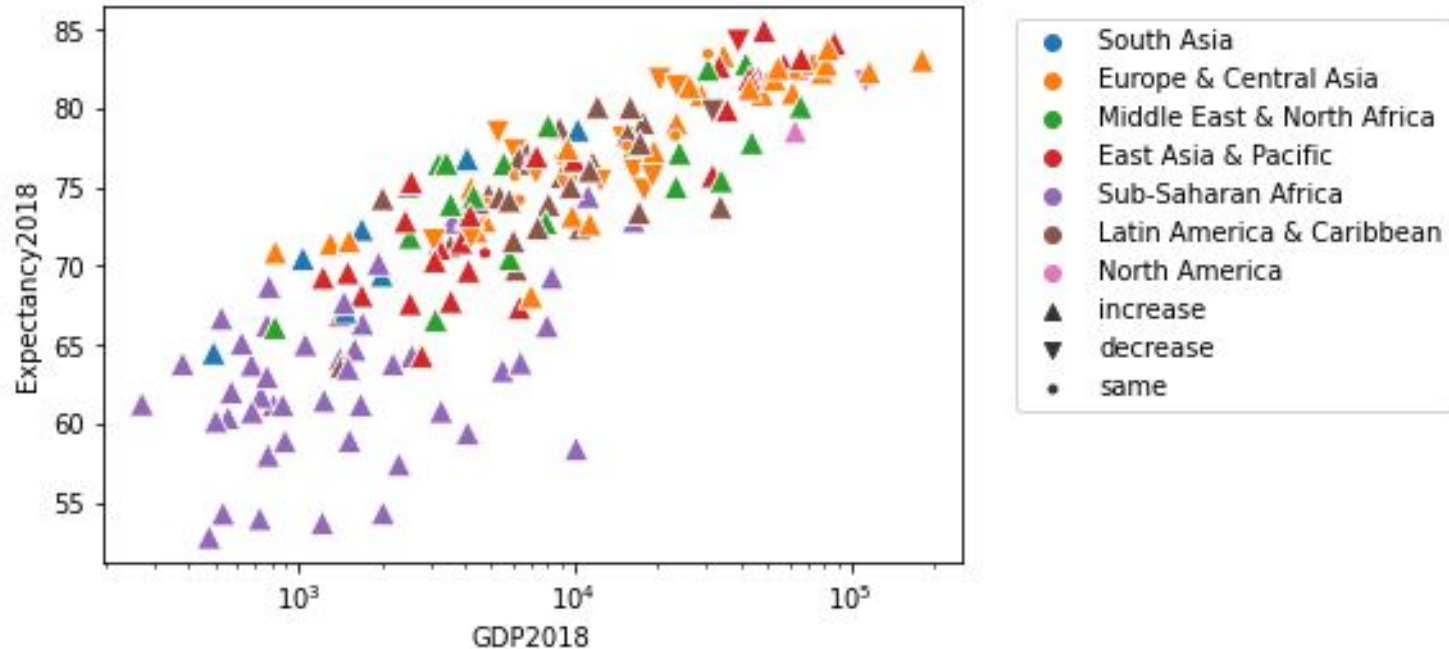


Adding numerical variable with marker size



Variable value should be proportional to circle area, not diameter!

Adding categorical variable with marker shape

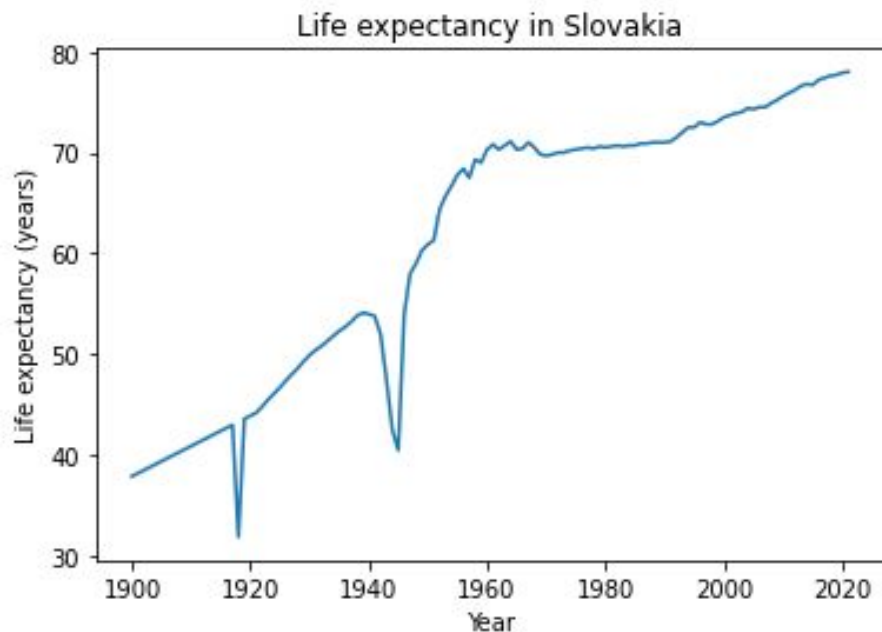


Hard to read, particularly for many data points

Showing population change between 2000 and 2018

If less than 1% change, marked as equal

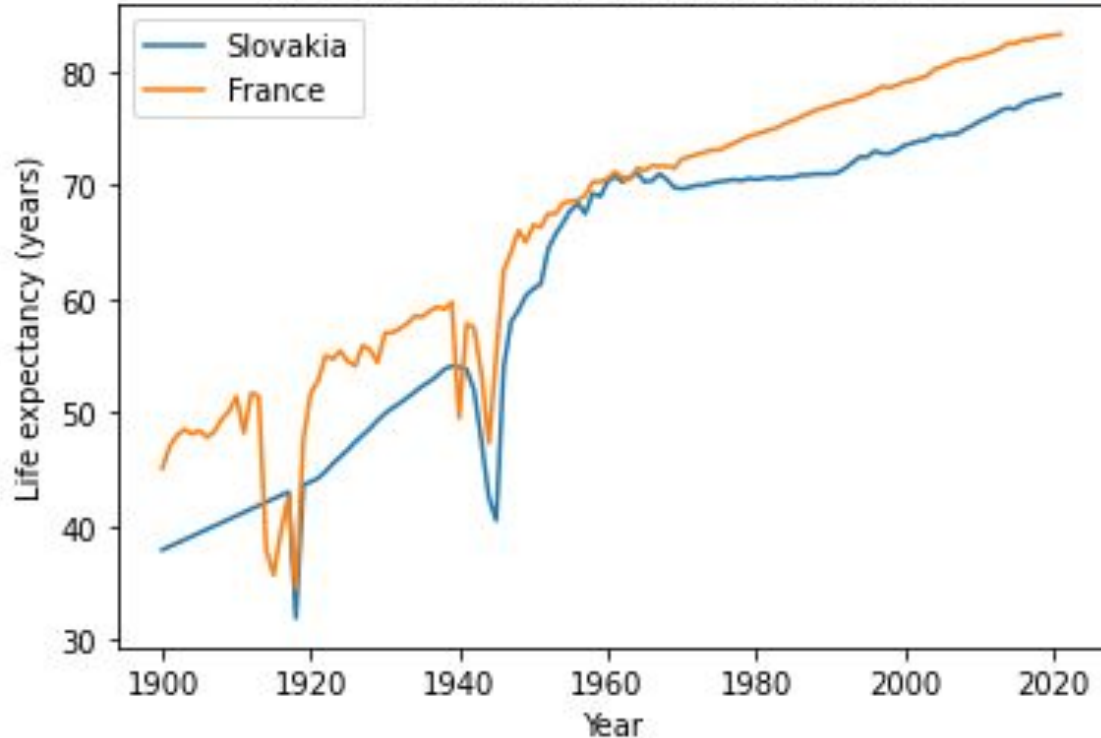
Line graph (čiarový graf)



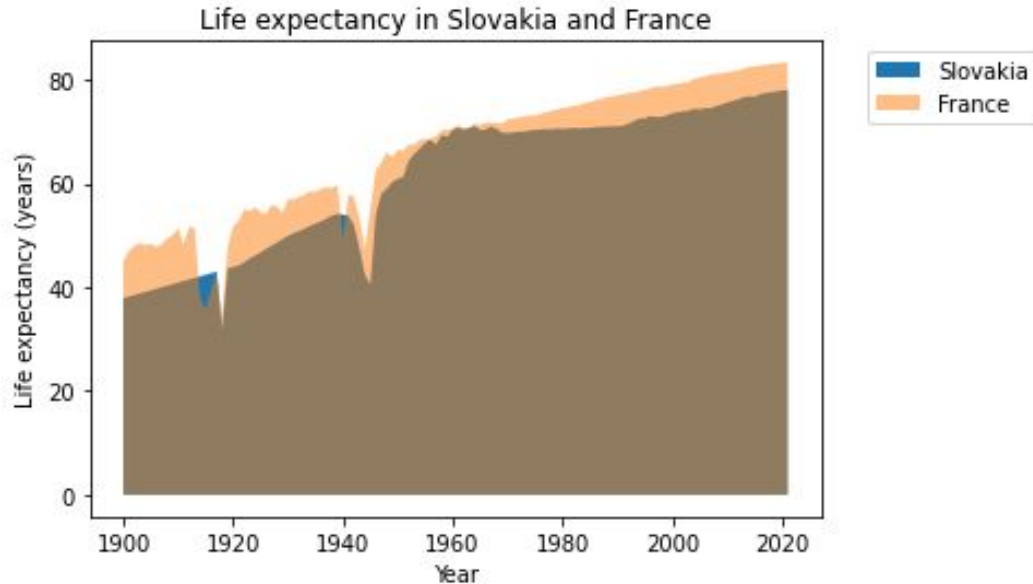
Emphasizing continuity between data points
Data points can be also shown as markers

Adding categorical variable with color

Life expectancy in Slovakia and France



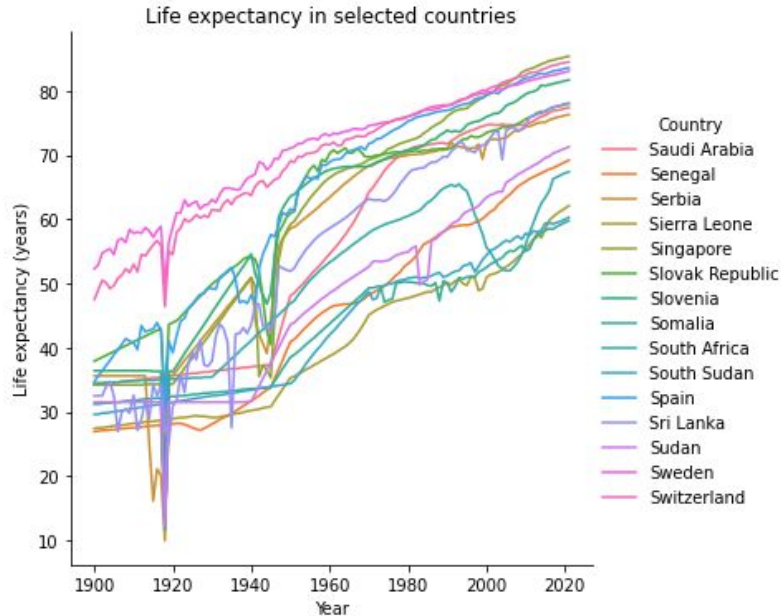
Area graph (plošný graf)



Y-axis must start at 0

Emphasizes differences more than line graph, but also more cluttered

Line graph with many lines



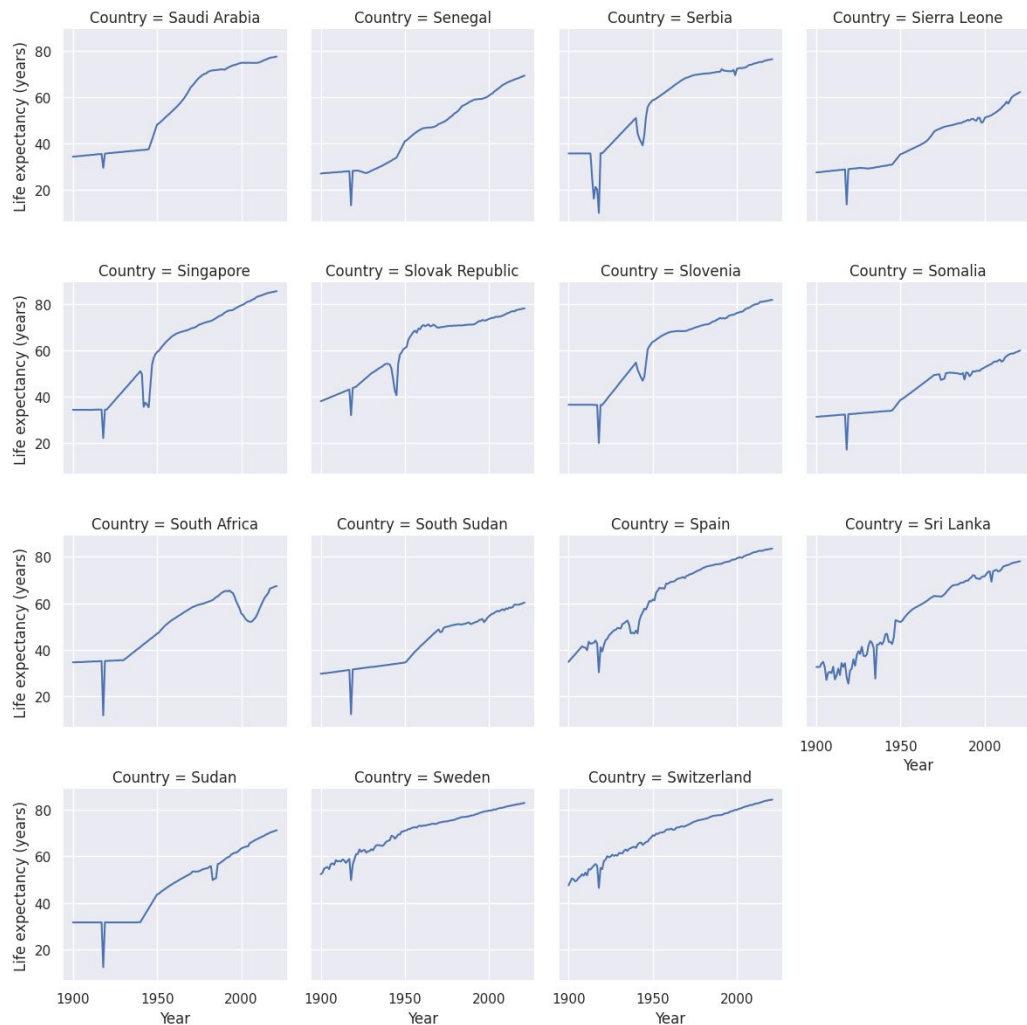
Hard to follow individual lines, but shows general trends and comparisons
Countries starting with S and having population at least 1 million (except Syria)

Small multiples

A small plot for each value of a categorical variable

Must have the same axes!

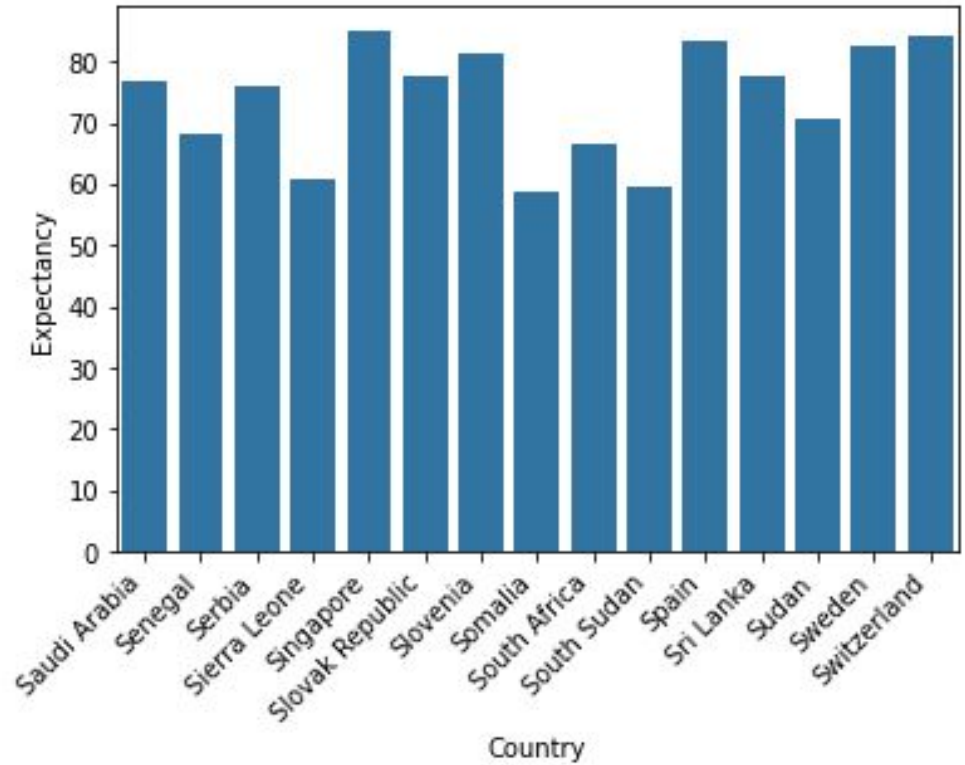
Exact comparison difficult, but it is possible to notice different trends



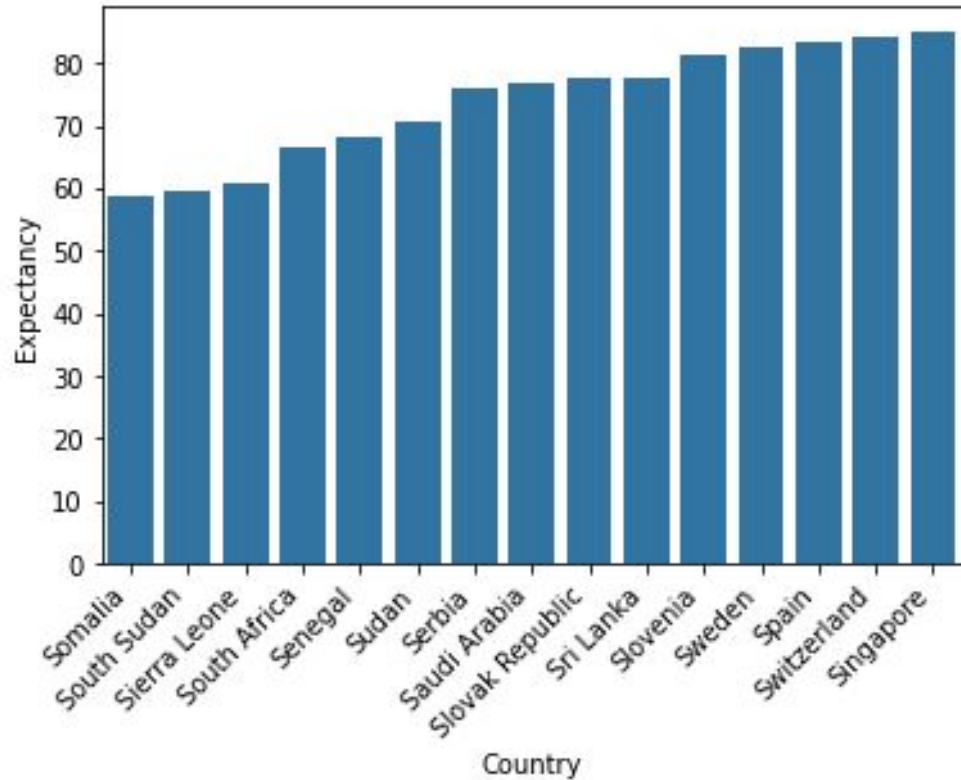
Bar graph (stĺpcový/pruhový graf)

X-axis is categorical

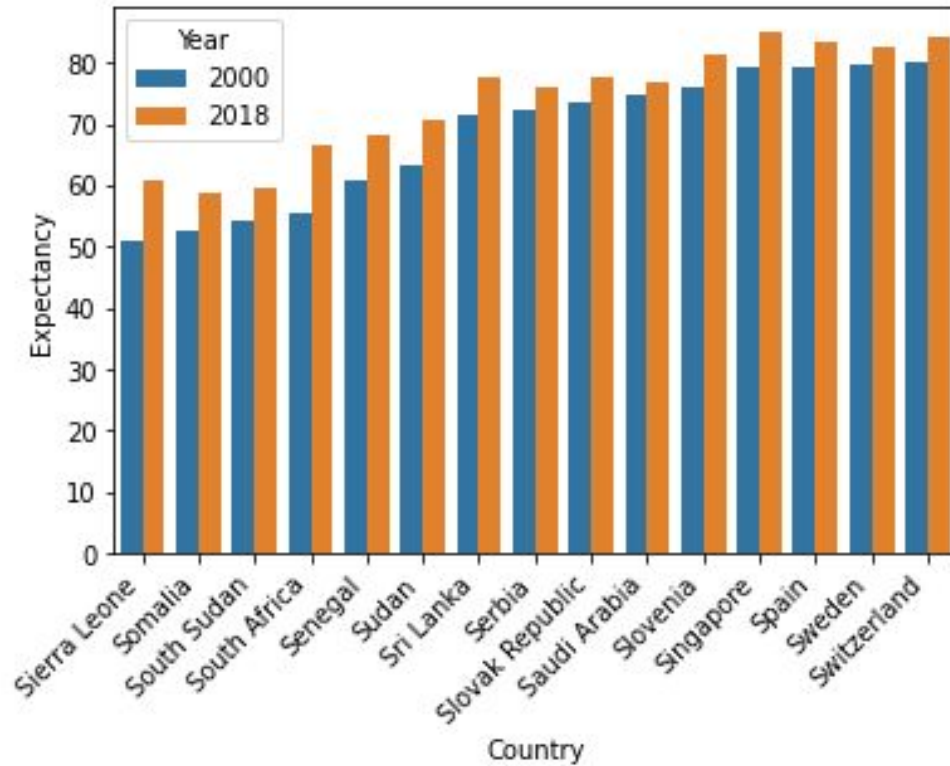
Y-axis must start at 0



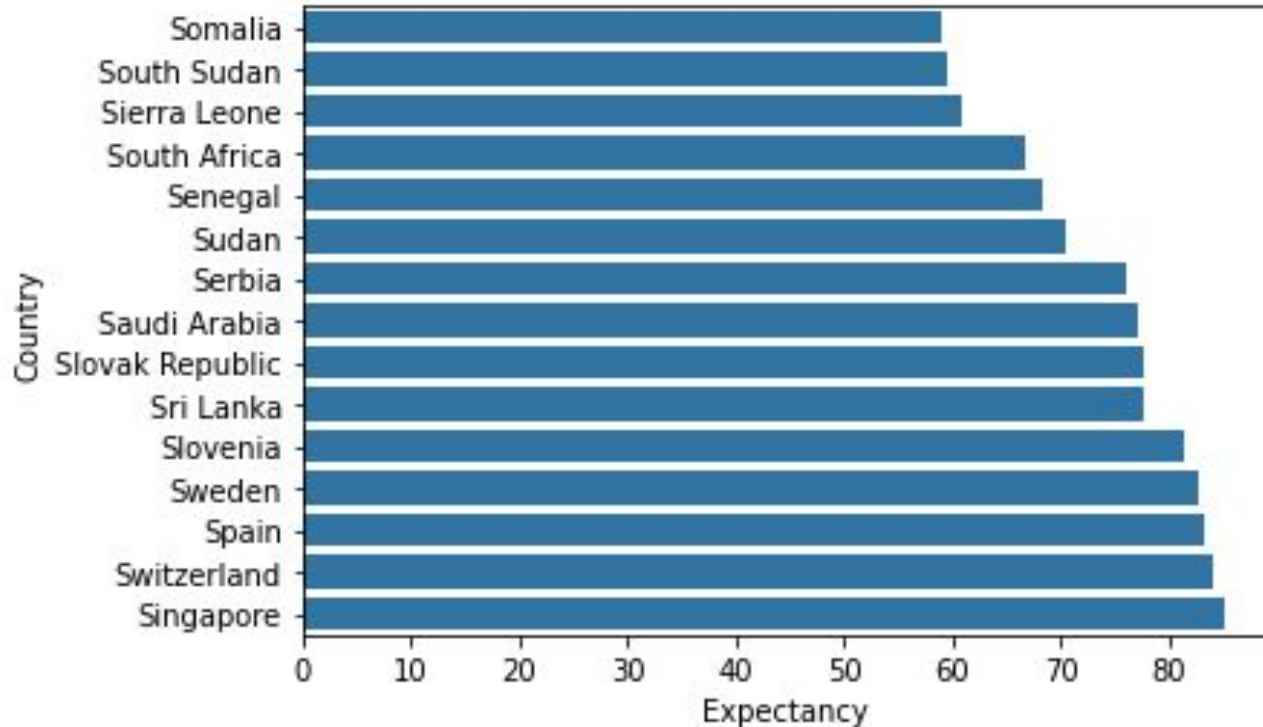
Bar graph with sorted columns



Bar graph with colored columns



Bar graphs can be horizontal



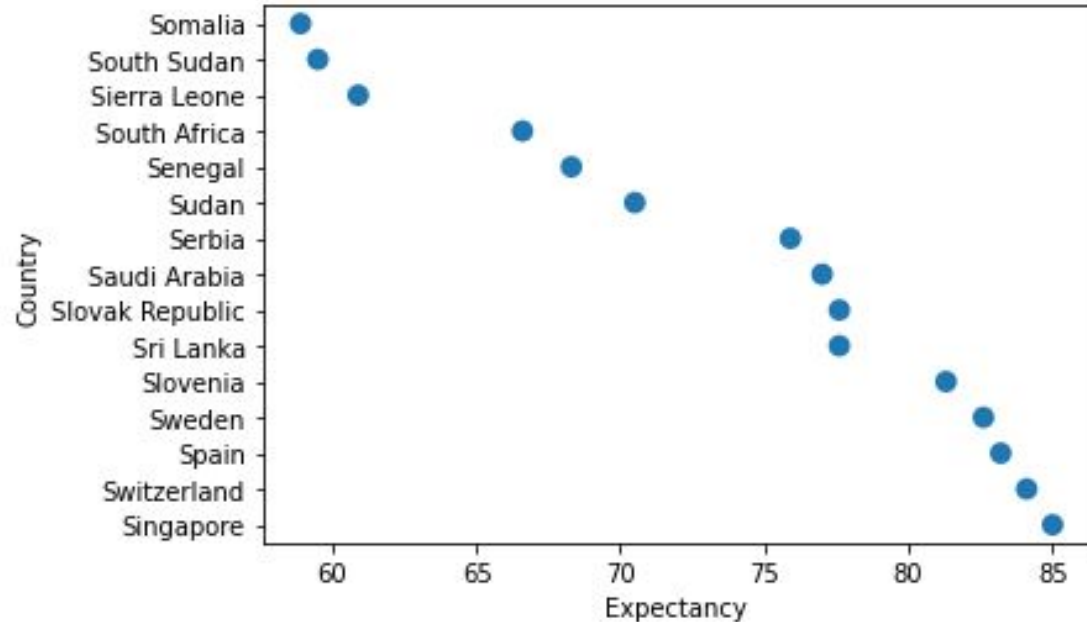
Dot plot

As bar graph but only dots shown at the top of the bar

Less clutter

X-axis does not need to start at 0 - better use of space

Can use multiple colors

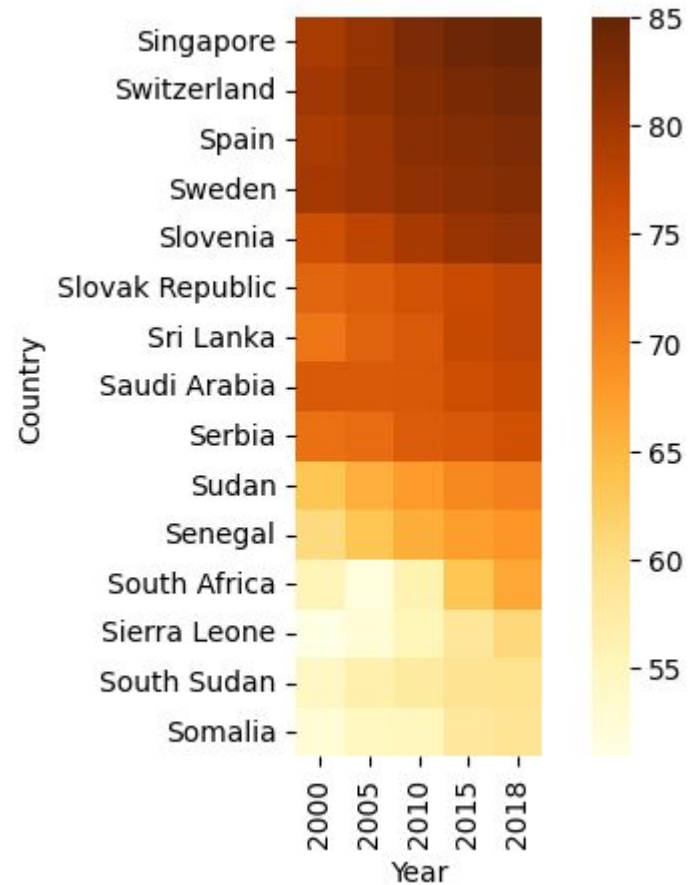


Heatmap

Both axes categorical

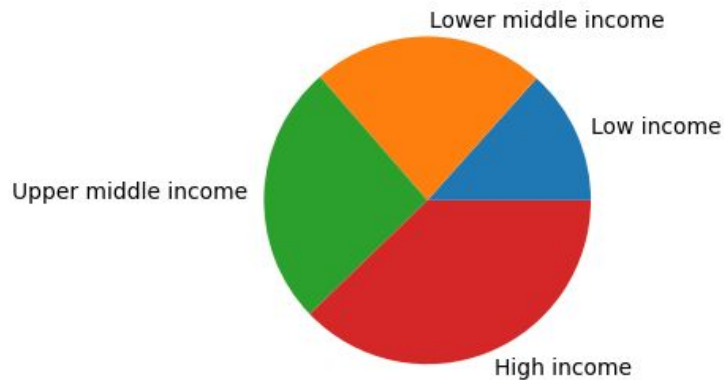
Numerical value shown in a color scale

Compact display, but color scales harder to read

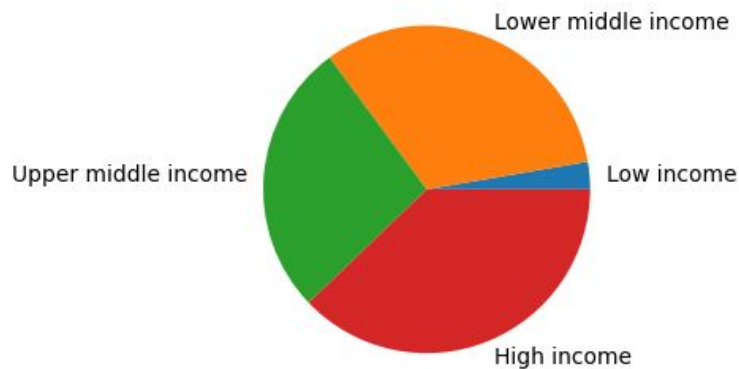


Pie chart (koláčový graf)

Whole world
(the number of countries)



East Asia & Pacific
(the number of countries)



Obvious that percentages displayed

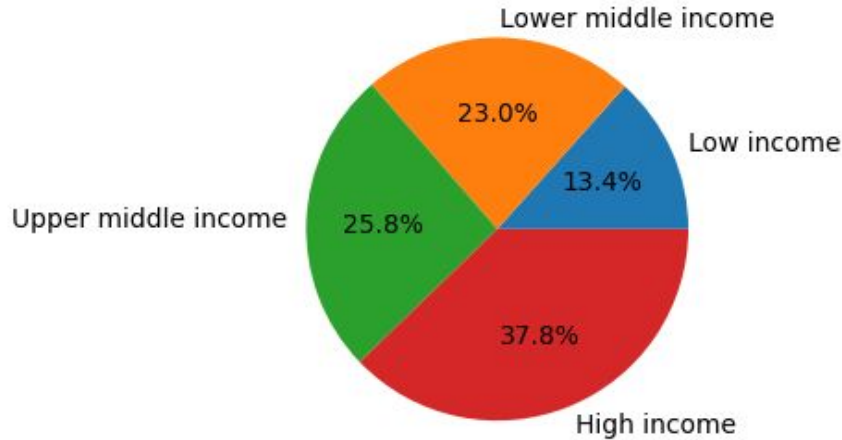
Very large values are easy to see (here high income)

Hard to compare similar values to each other

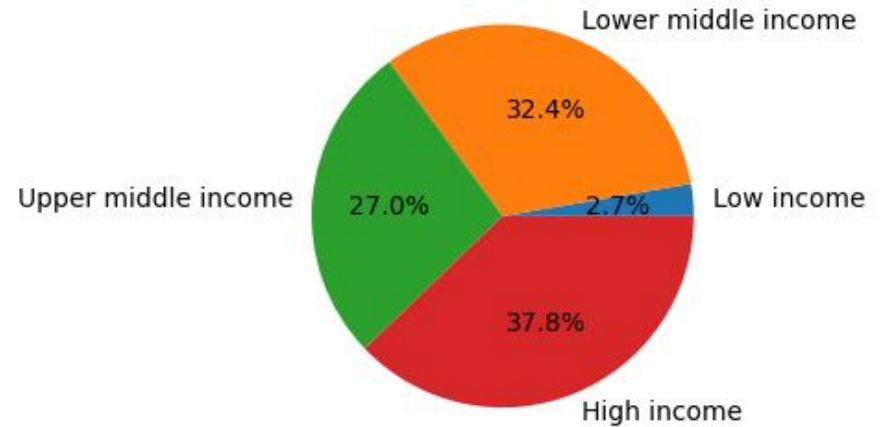
Space use not good

Pie chart with values labeled

Whole world
(number of countries)



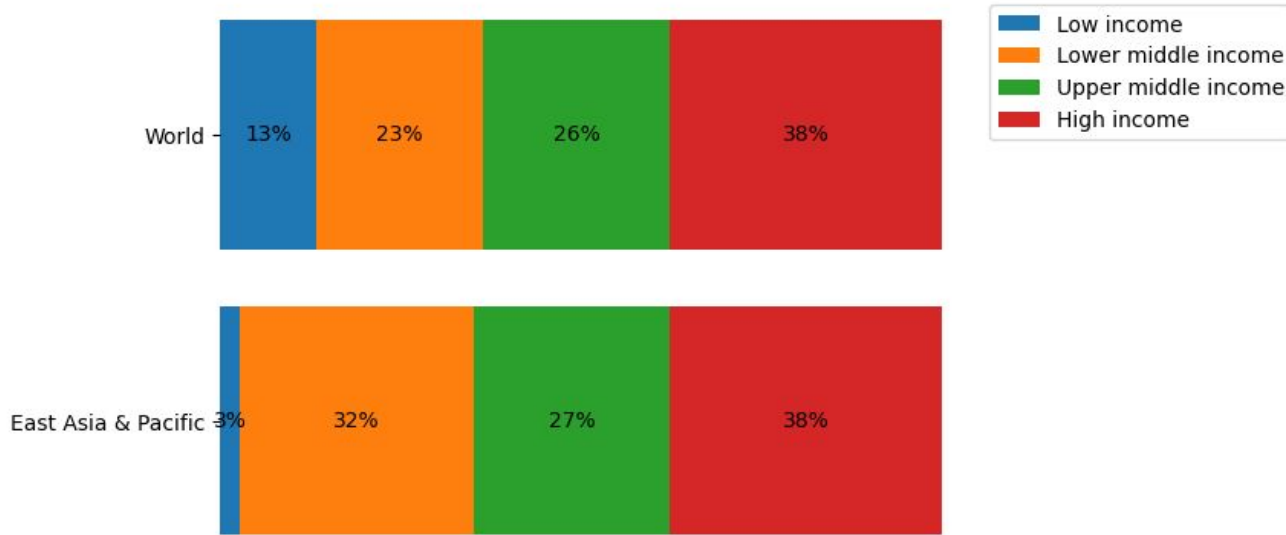
East Asia & Pacific
(number of countries)



Easier to compare but still not ideal

Labeling values also useful in other types of graphs

Stacked (skladaný) bar graph instead of pie chart



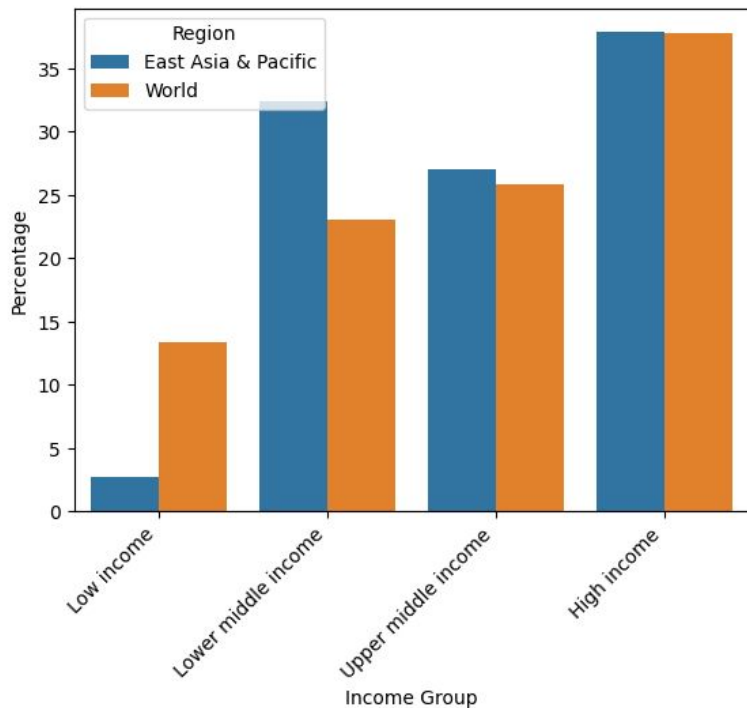
Rectangles easier to compare than wedges

Would benefit from labeled values

Middle colors hard to compare across bars

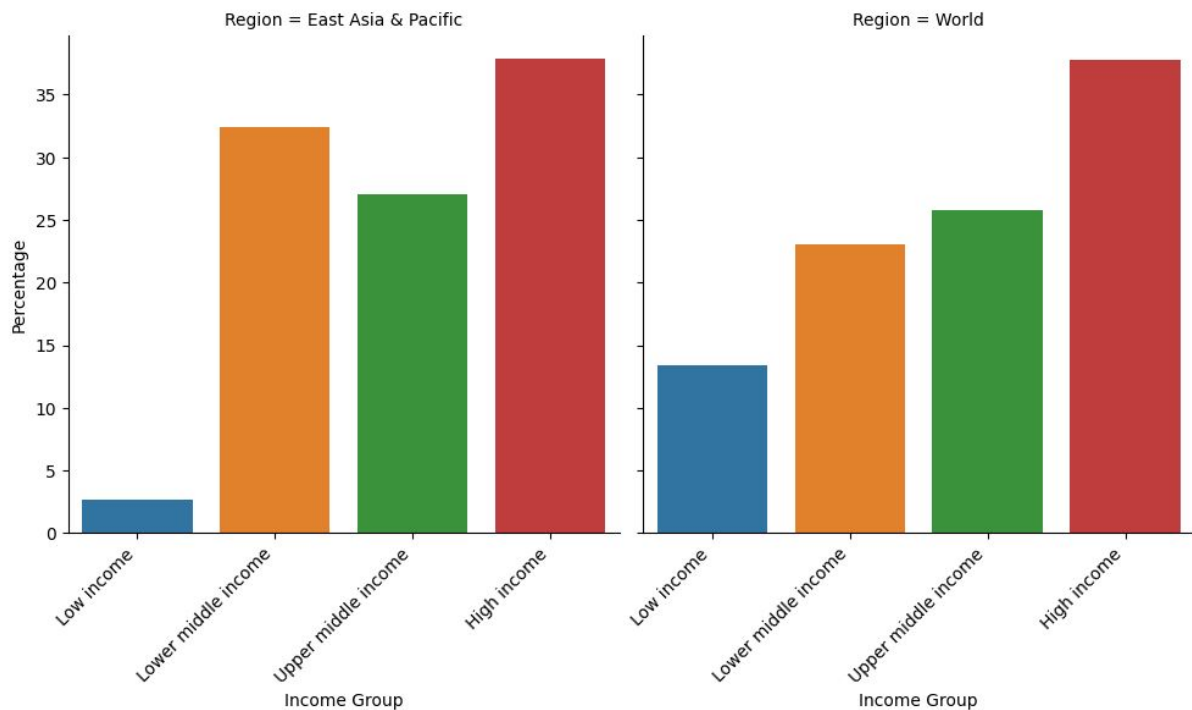
Similar idea: stacked area plot (change in percentages over time)

Colored bar graph instead of pie chart



Easy to compare East Asia vs whole world

Colored bar graph instead of pie chart



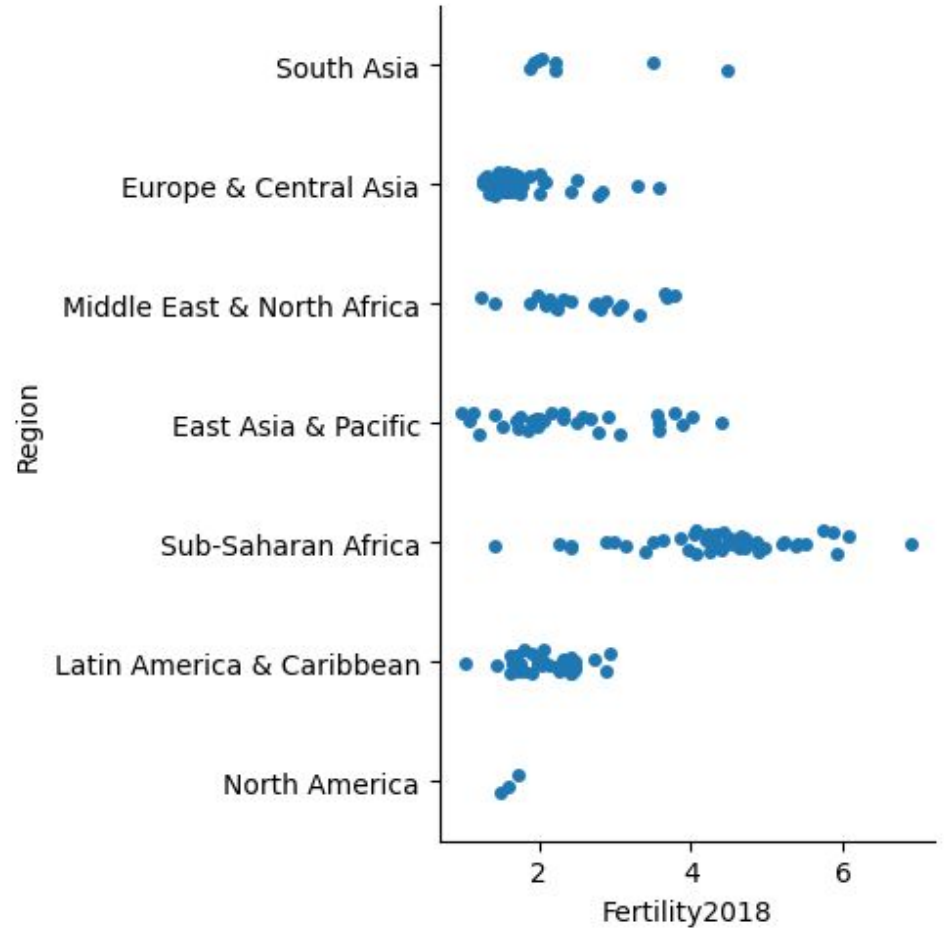
Easy to compare income groups within region

Strip plot

One axis categorical

Other axis shows individual data points

Jitter added in categorical axis to avoid point overlap

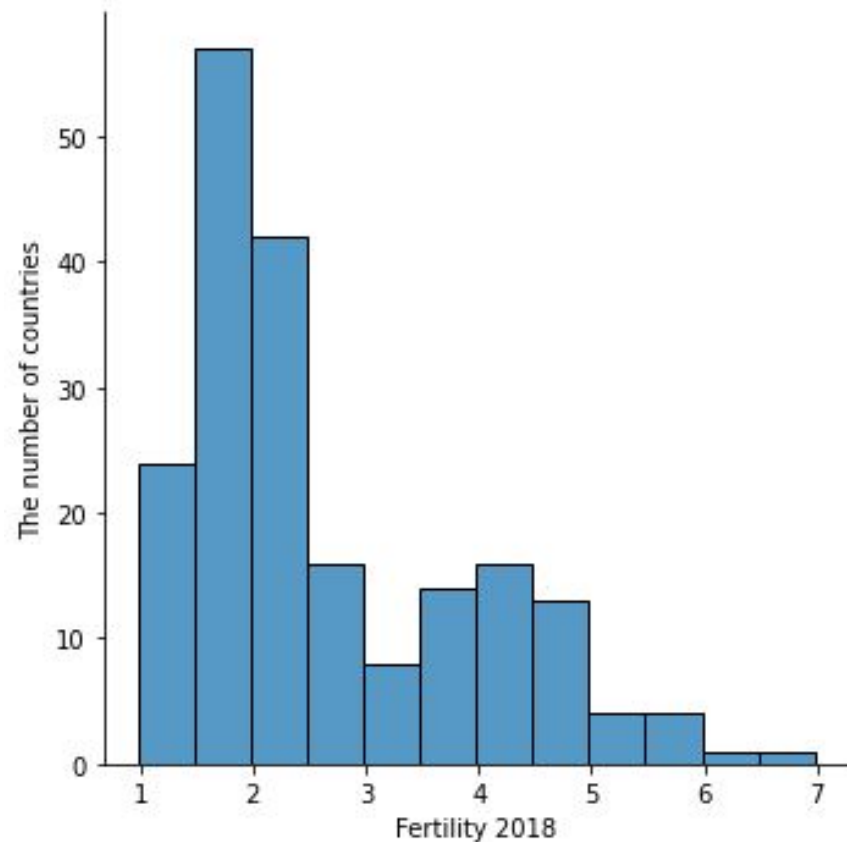


Histogram

For 1D numerical data

Split values into bins, show bin sizes as bar graph

We could use colors to display 2 or more histograms



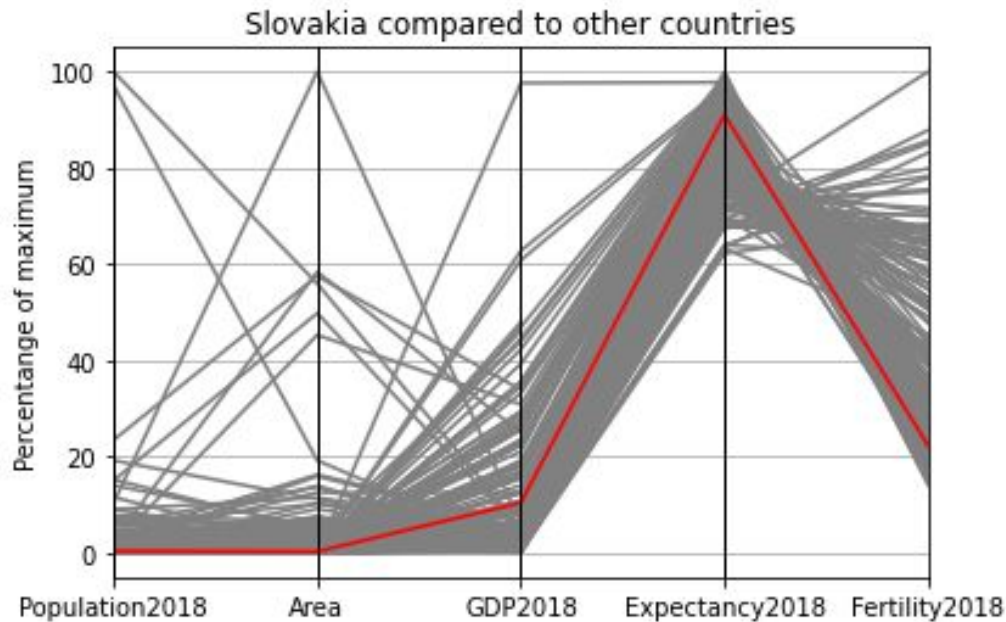
Parallel coordinates

Good for multidimensional numerical data

Each column one dimension

Here scaled as % of maximum value

Hard to see individual lines, but can show trends, compare groups shown in color or selected data point vs others



Parallel categories

Good for multidimensional categorical data

Each column one dimension

The widths of ribbons correspond to the number of countries



Radar chart (radarový graf)

Hard-to-read version of parallel
coordinates

Perhaps some justification in cyclical
domains, such as average
temperature in months of a year

