Fachhochschule Kiel
University of Applied Sciences

Data Visualization and Visual Analytics
Comparing Categories

Study Program Data Science
Prof. Dr. Tillmann Schwörer

Lecture Roadmap

Data Domains

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

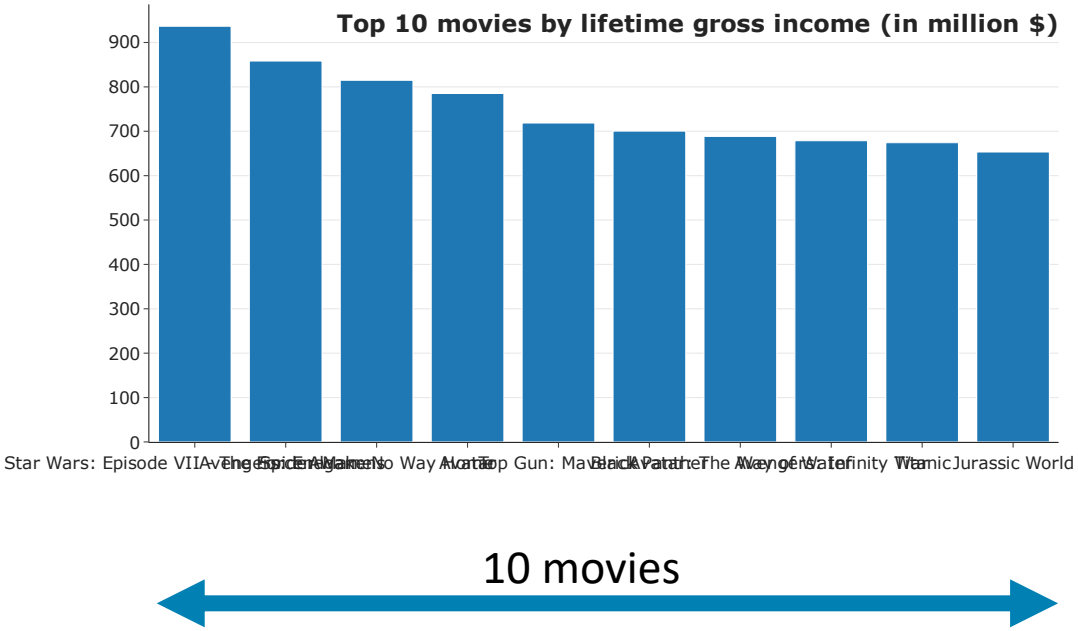
Storytelling

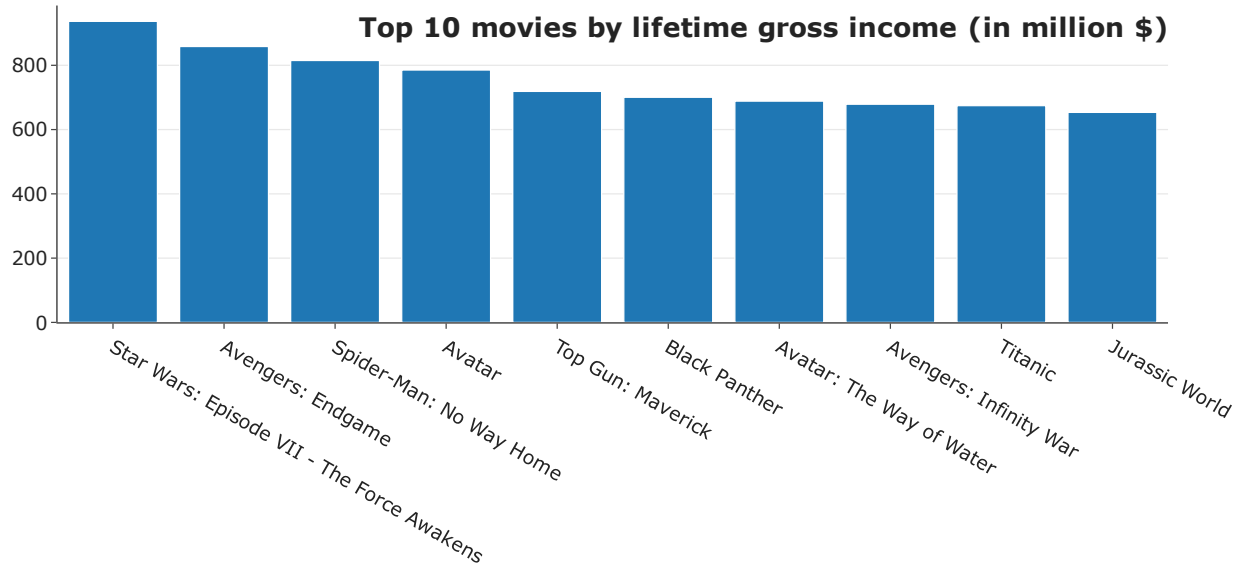
Perception + Visualization Design

Python + Tools

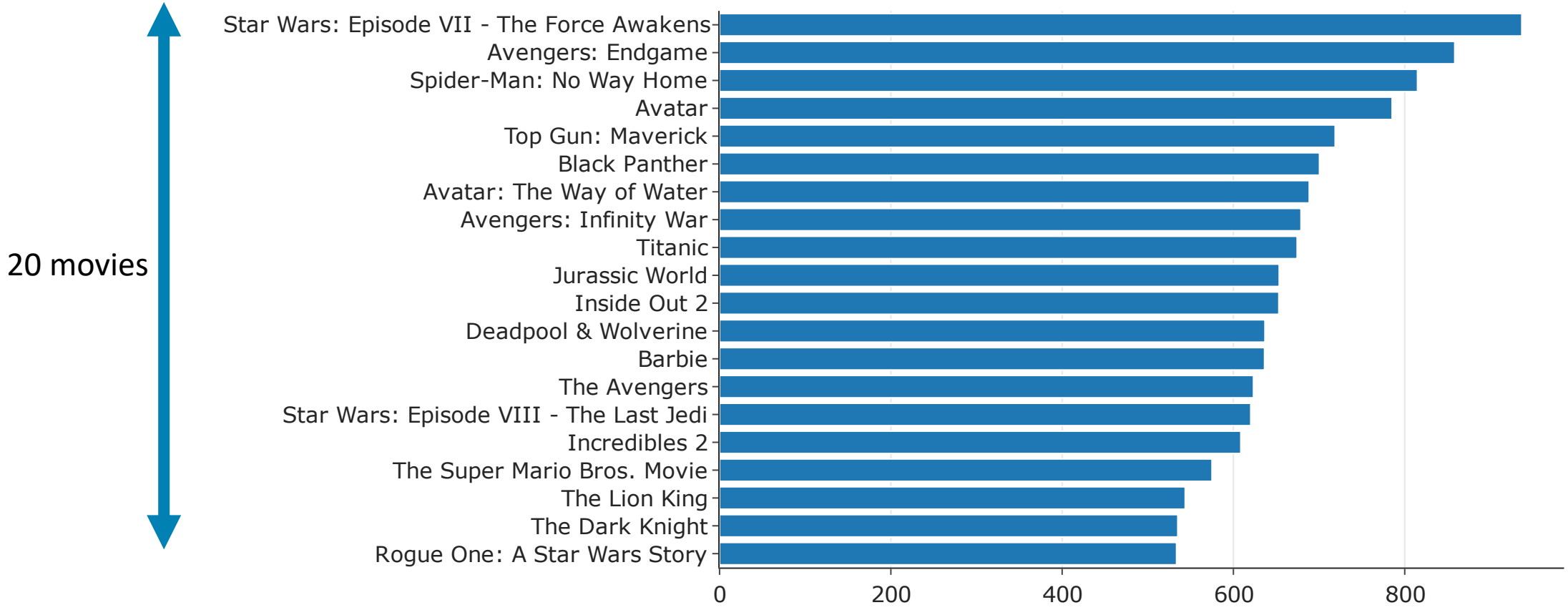
Interactive Visualization

Long Label Problem





Horizontal bars

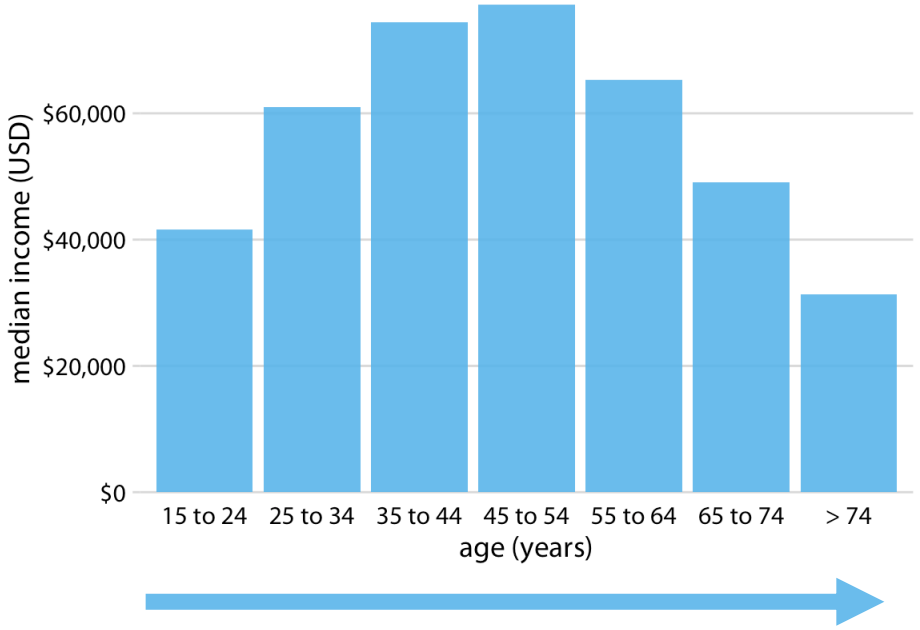


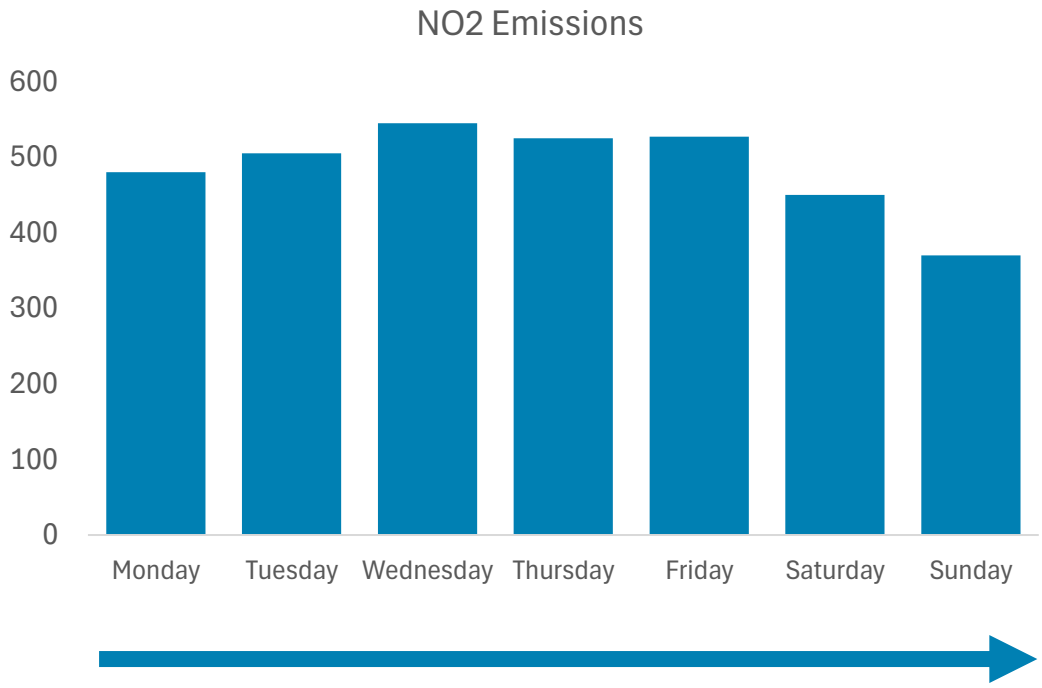
Horizontal bars

- Work well with **long category names** and with **many categories**
- No need to rotate or abbreviate names
- Supports left-to-right **reading flow**: label → bar → value

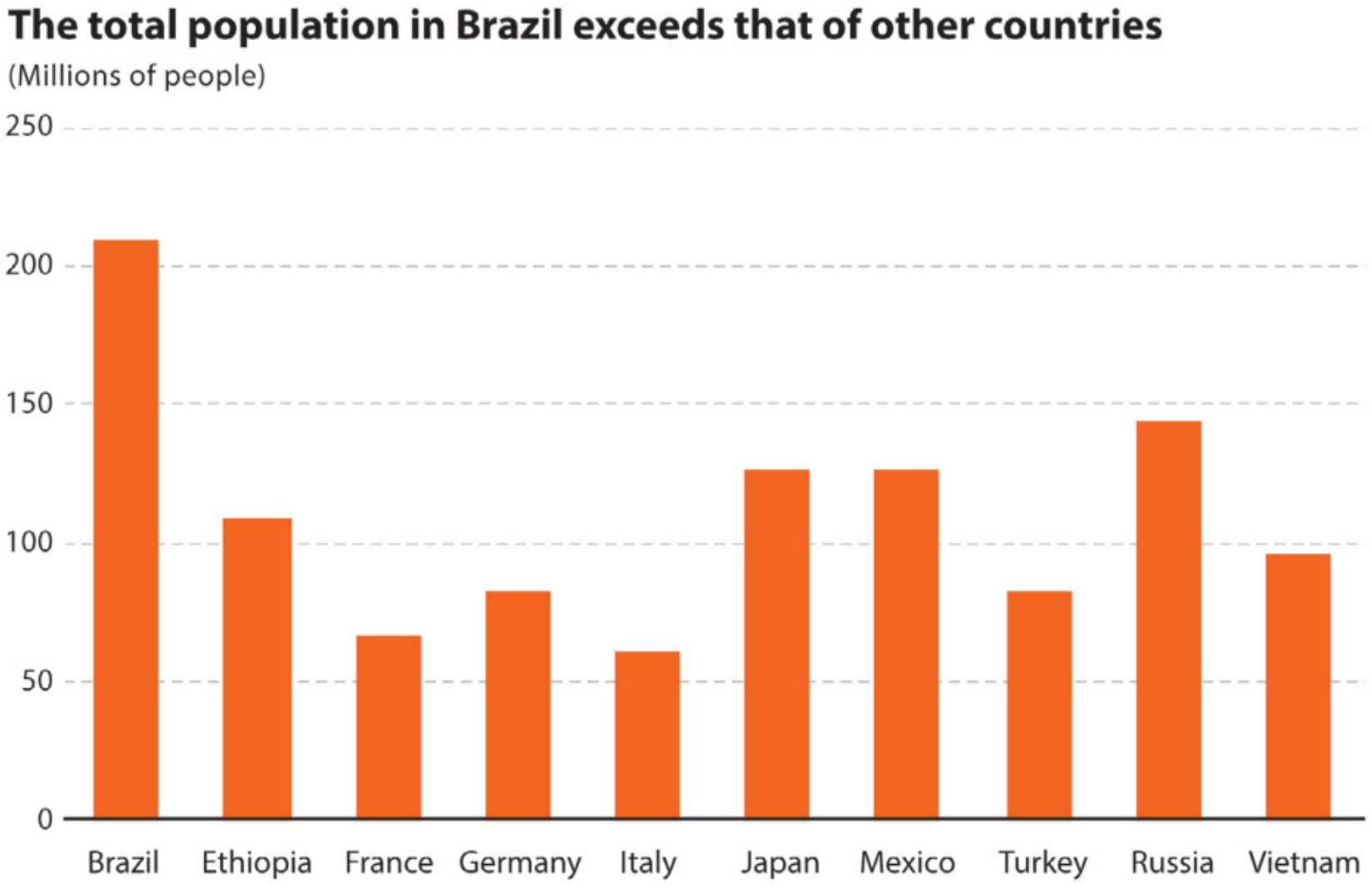
Always horizontal bars?

- **No** – there is not the one chart that fits all cases
- Vertical bars are great if
  - category labels don‘t overlap
  - categories have a natural order (supports left-to-right reading flow)

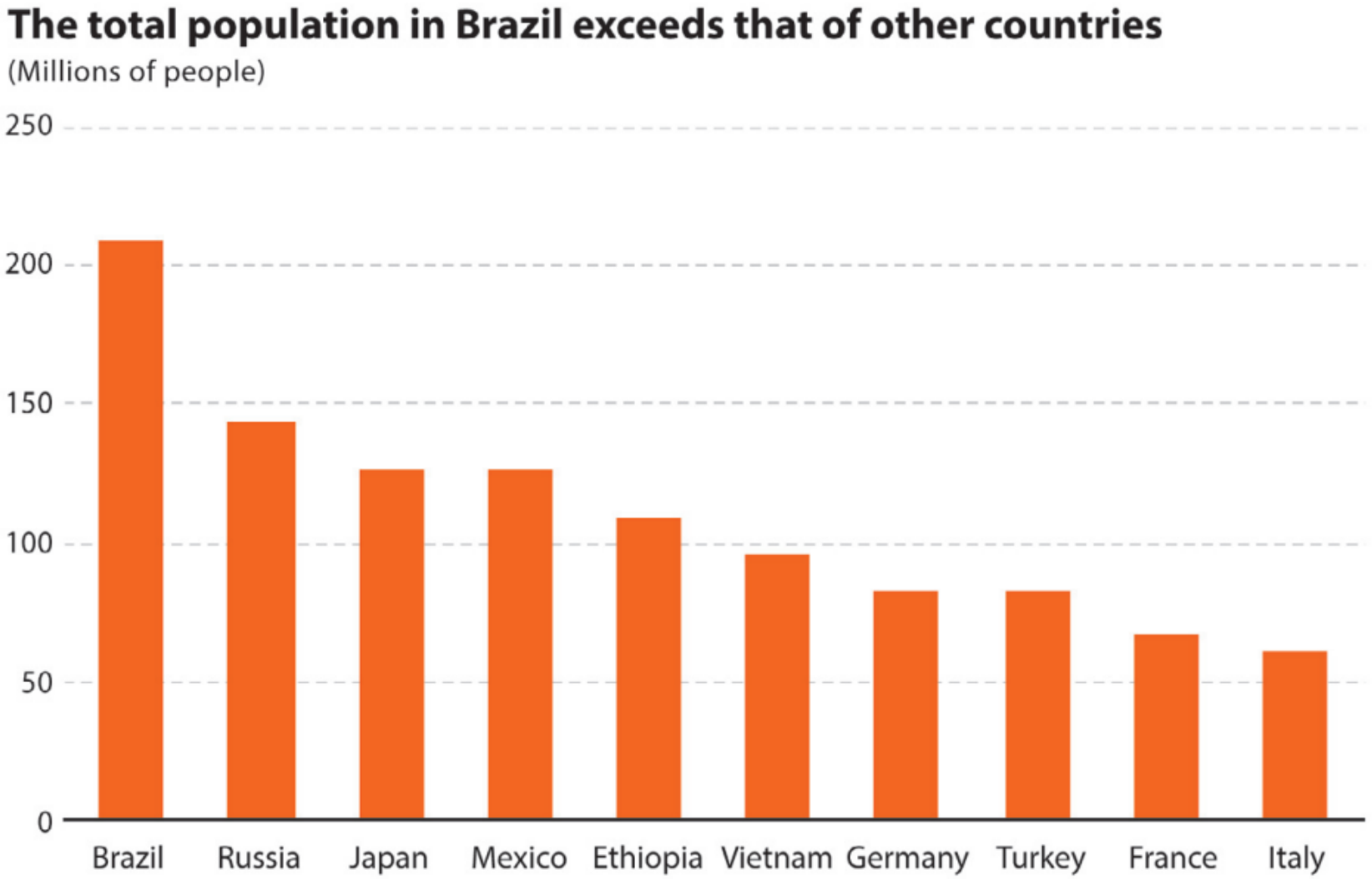




Alphabetical sorting?



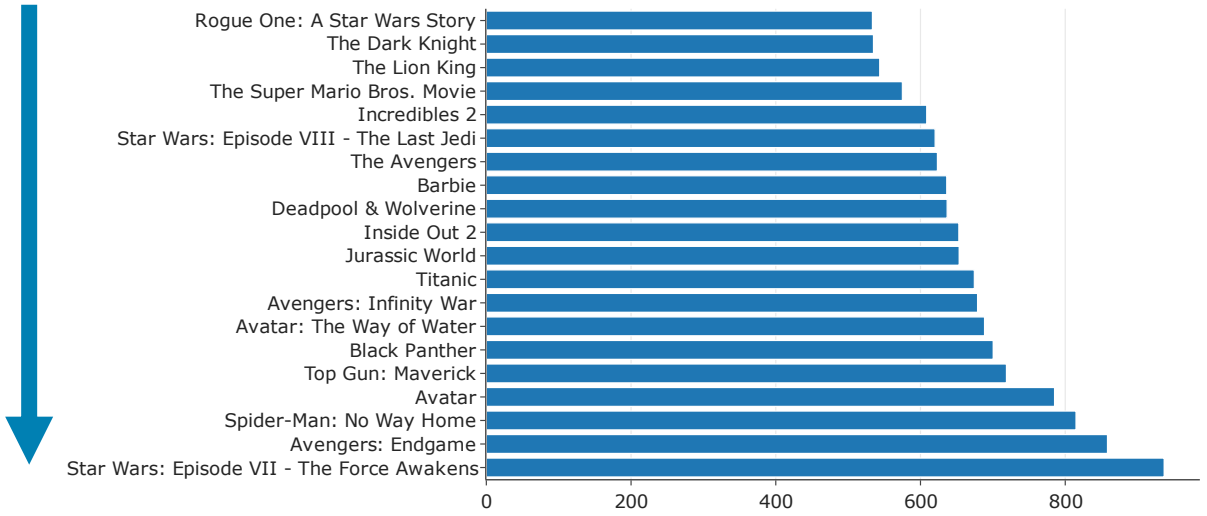
Length sorting!

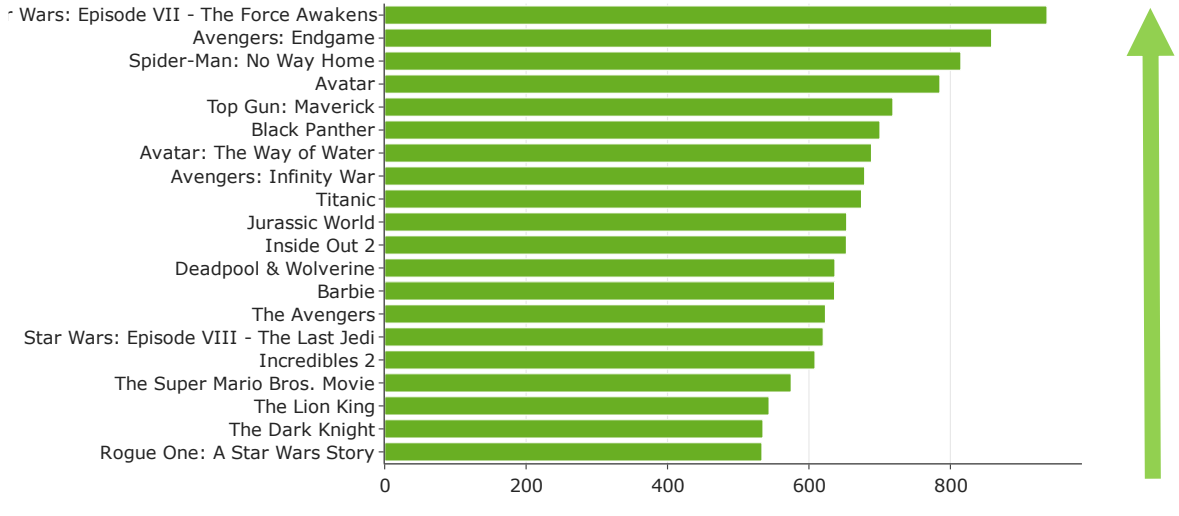


Usually, we sort bars according to the bar length

→ Pattern perception!

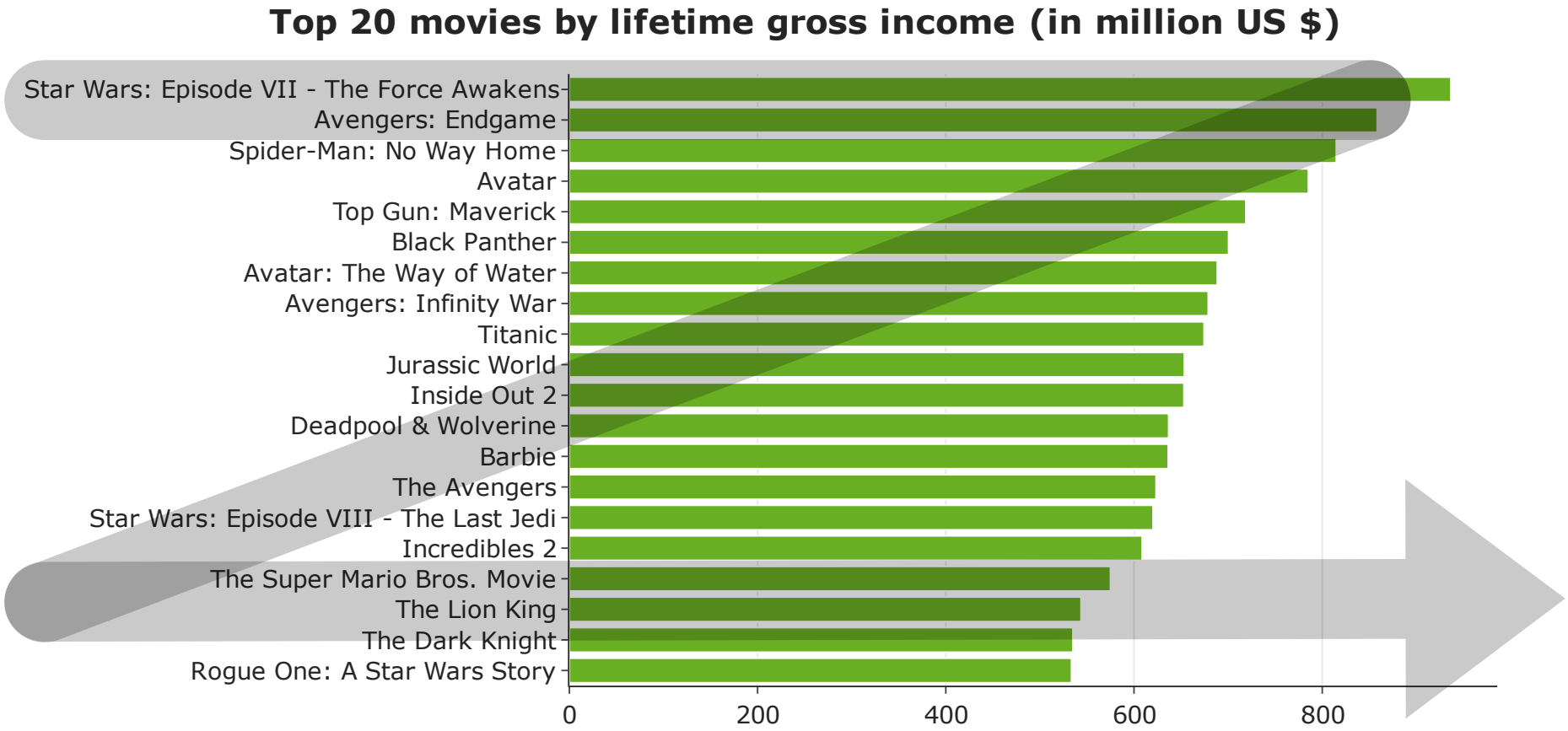
Descending or ascending order?





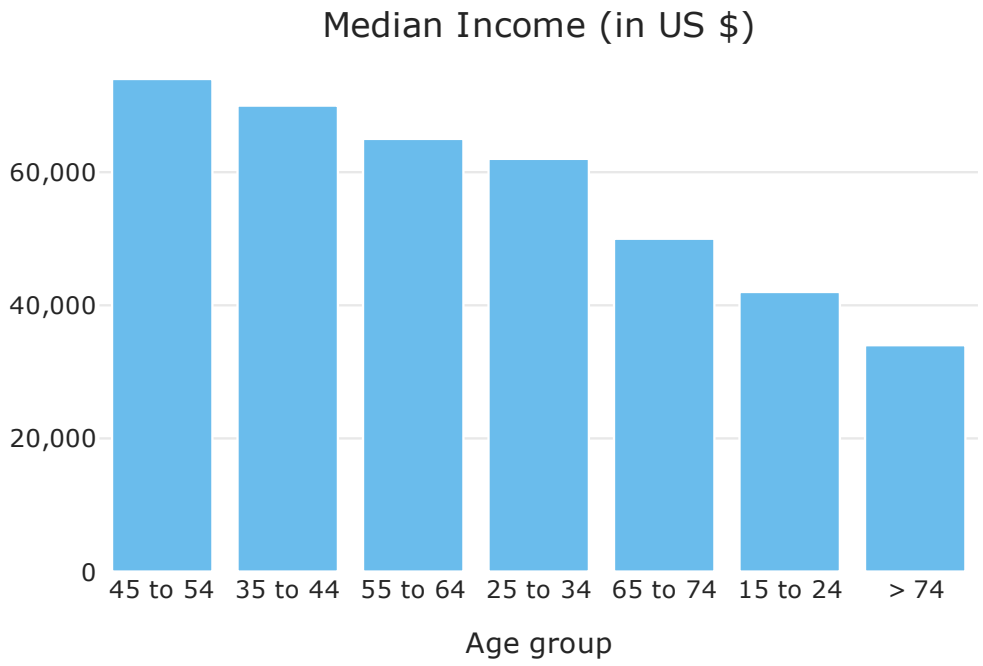
Largest bar on top

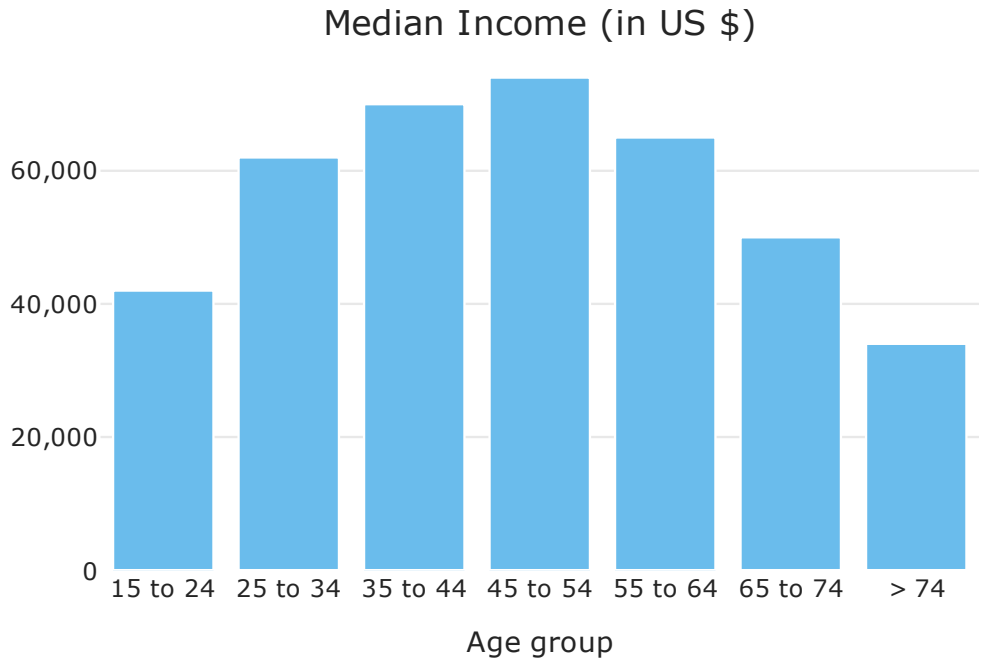
- People (in western countries) follow Z-shaped viewing pattern
- The most important information (title, top category, …) should appear on the top left



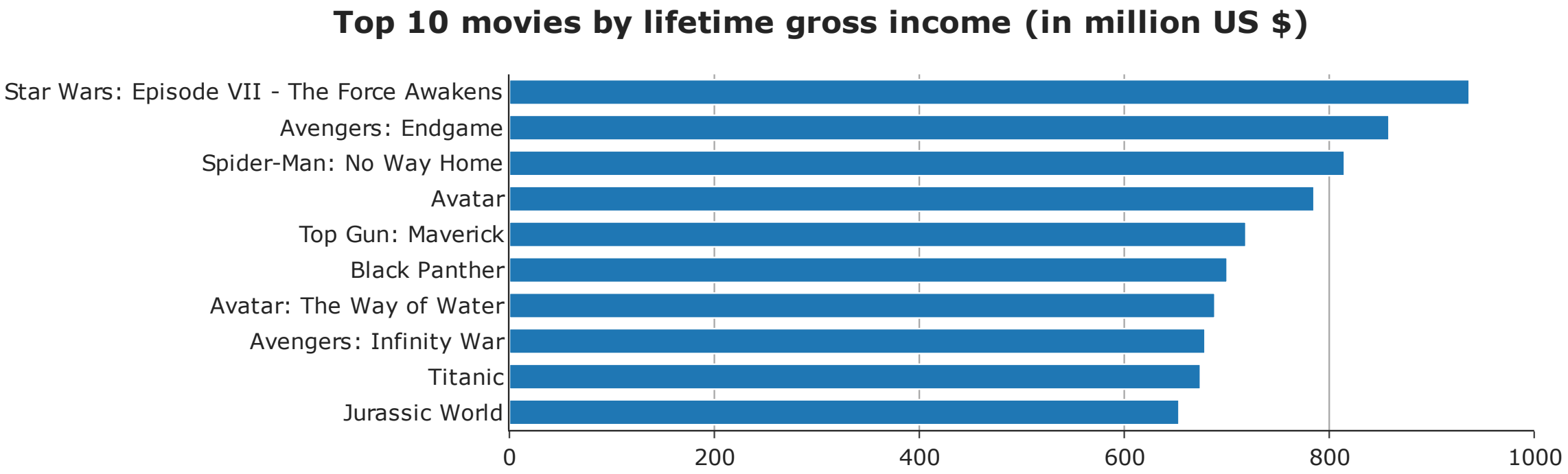
Always sort according to bar length?

No, if the categories are ordered (time, bins of a numerical variable, …) we should stick to this natural order.

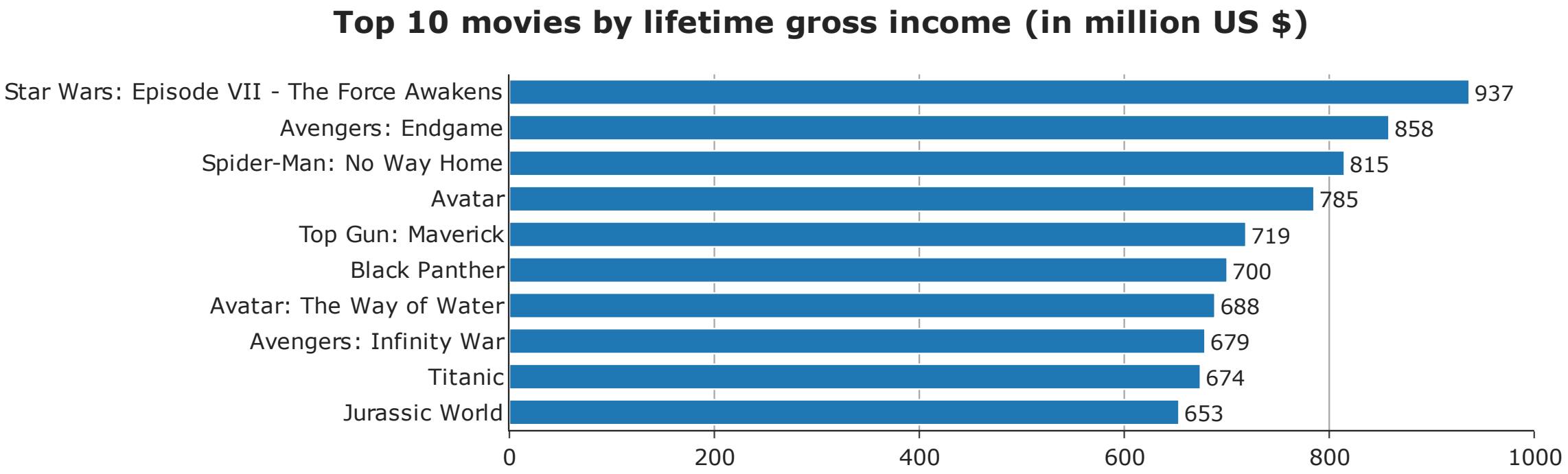




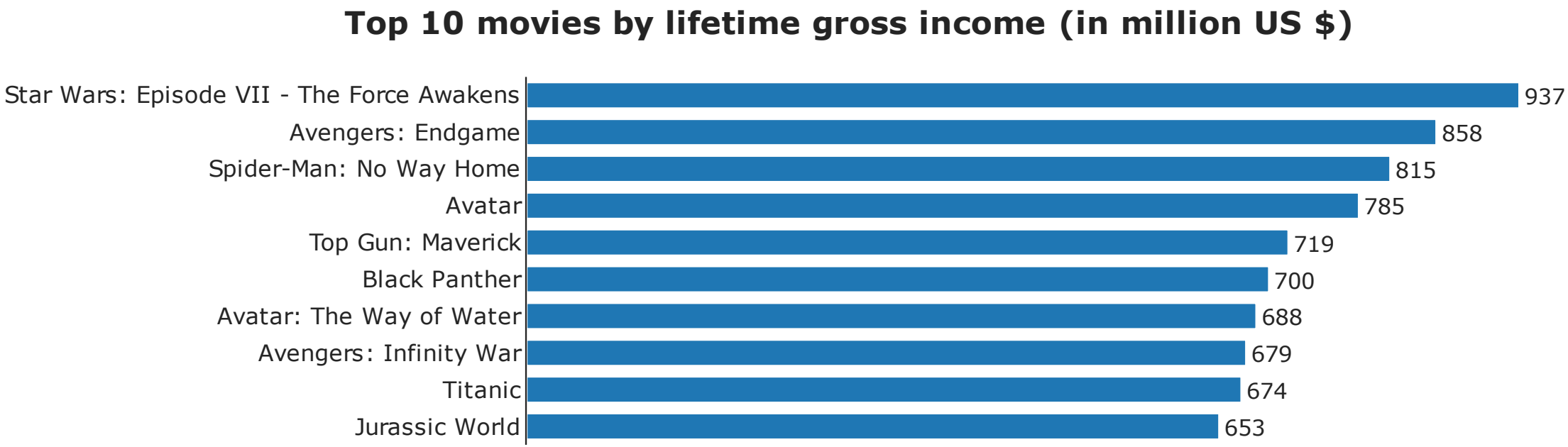
Data labels?


Top 10 movies by lifetime gross income (in million US $)
Star Wars: Episode VII - The Force Awakens
Avengers: Endgame
Spider-Man: No Way Home
Avatar
Top Gun: Maverick
Black Panther
Avatar: The Way of Water
Avengers: Infinity War
Titanic
Jurassic World
0
200
400
600
800
1000

Data labels?


Top 10 movies by lifetime gross income (in million US $)
Star Wars: Episode VII - The Force Awakens
937
Avengers: Endgame
858
Spider-Man: No Way Home
815
Avatar
785
Top Gun: Maverick
719
Black Panther
700
Avatar: The Way of Water
688
Avengers: Infinity War
679
Titanic
674
Jurassic World
653
0
200
400
600
800
1000

Data labels?

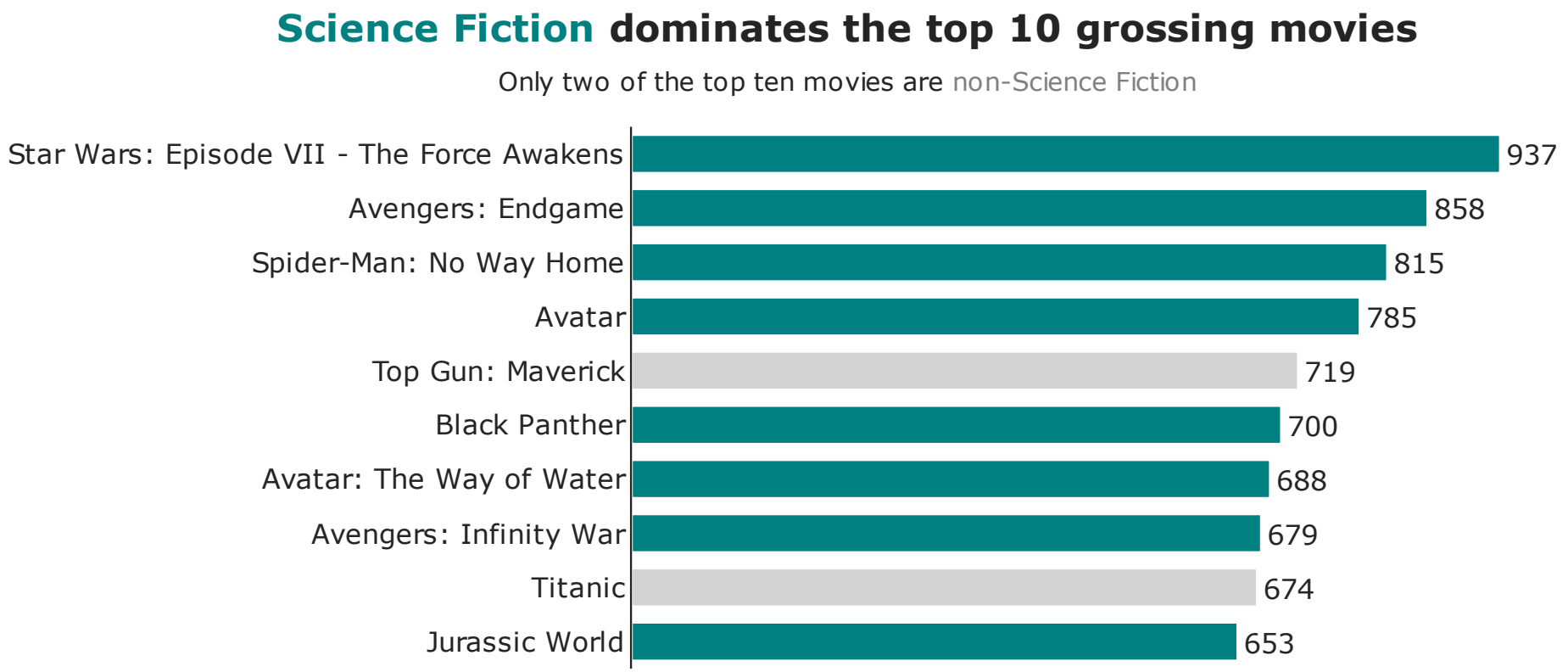

Top 10 movies by lifetime gross income (in million US $)
Star Wars: Episode VII - The Force Awakens
937
Avengers: Endgame
858
Spider-Man: No Way Home
815
Avatar
785
Top Gun: Maverick
719
Black Panther
700
Avatar: The Way of Water
688
Avengers: Infinity War
679
Titanic
674
Jurassic World
653

Data labels!

- Data labels remove the need for **y-axis and grindlines**
- Advantages:
  - Lower look-up times → proximity principle!
  - Exact data values
  - Improved figure-ground separation (less clutter!)
- Data labels work well – unless you have many bars

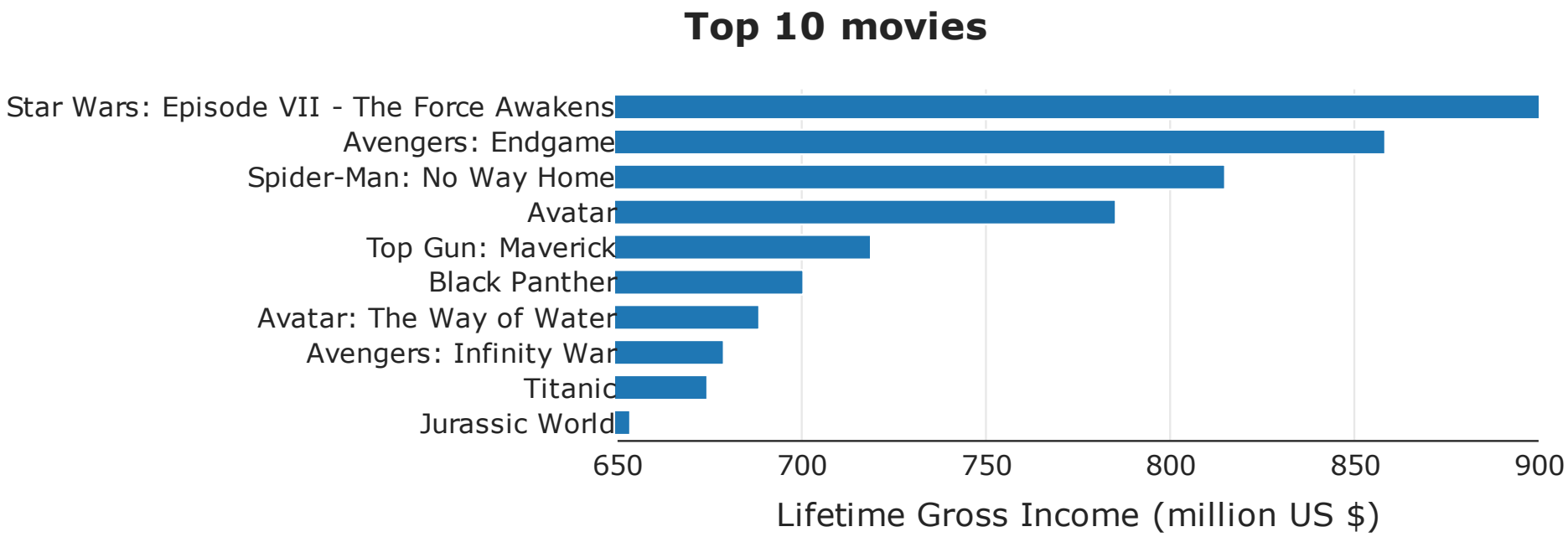
Pre-attentive attributes

We can use color to draw attention to the main pattern(s) of interest



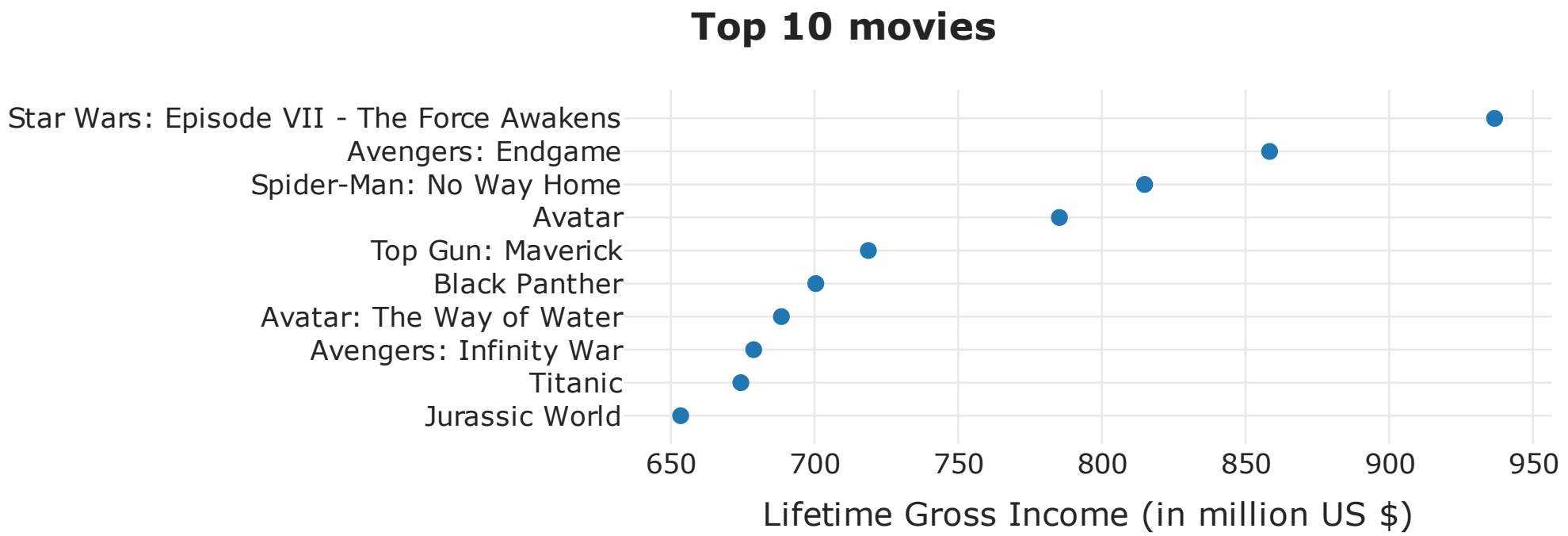
Bar plots and axis range

- The axis of a bar plot must always start at zero
- Reason: the bar length must be proportional to the magnitude that it represents (**Principle of proportional ink)**

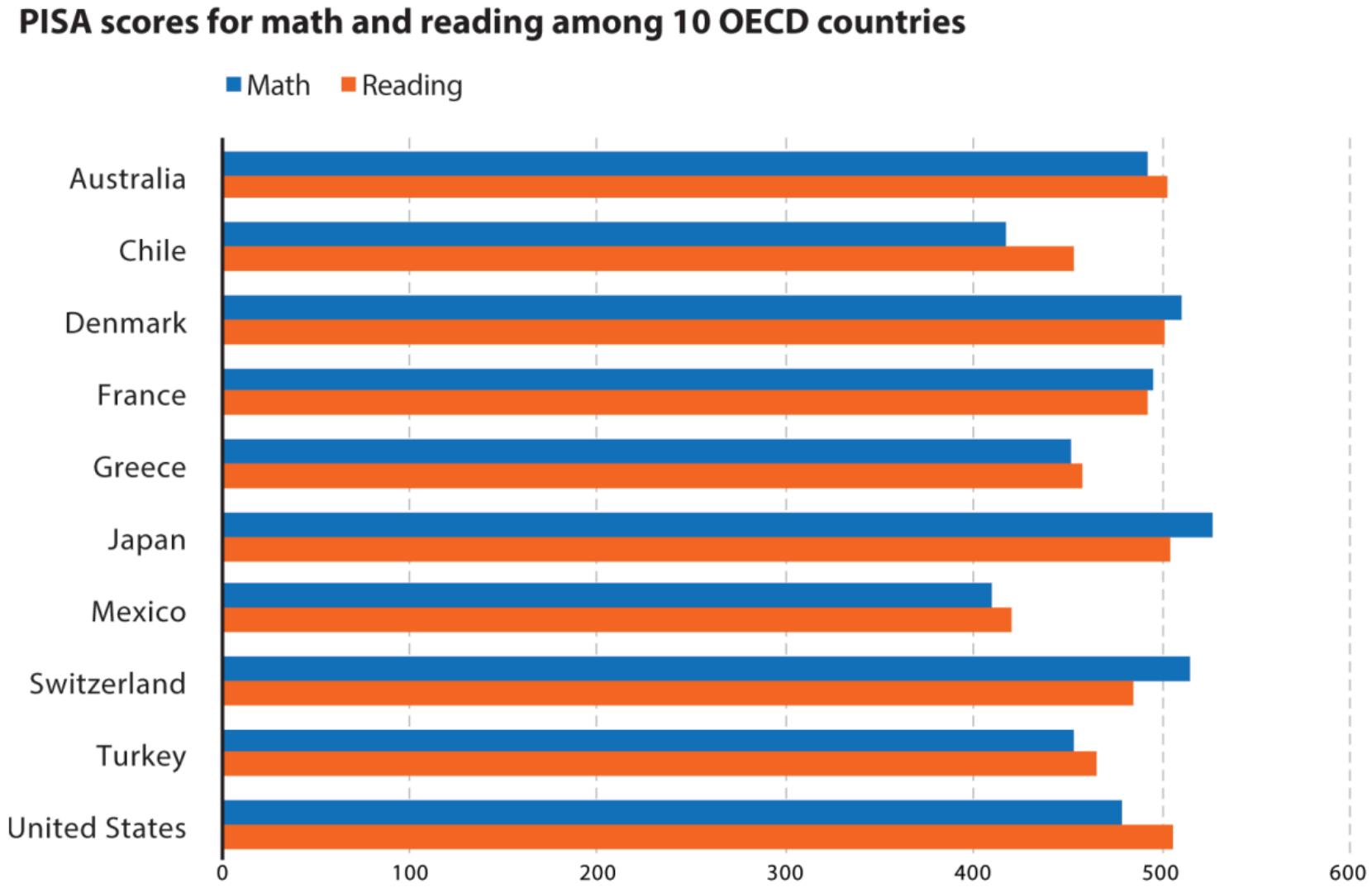


Scatterplot and axis range

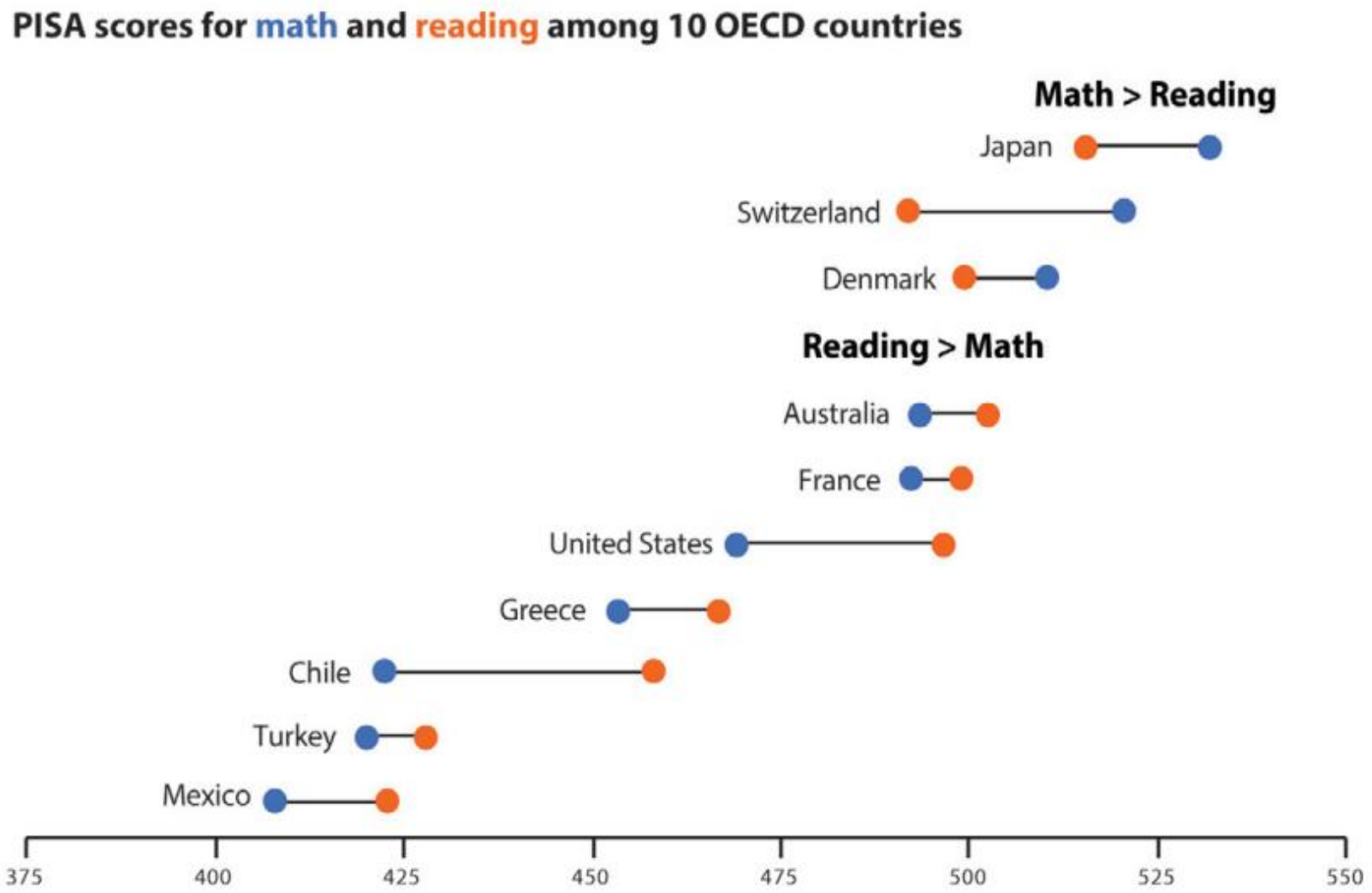
- The axis of a scatterplot is allowed to start anywhere, because it is the position of the marker – not its size or length – that encodes the value → the **principle of proportional ink** is not violated
- Scatterplots are great to highlight differences



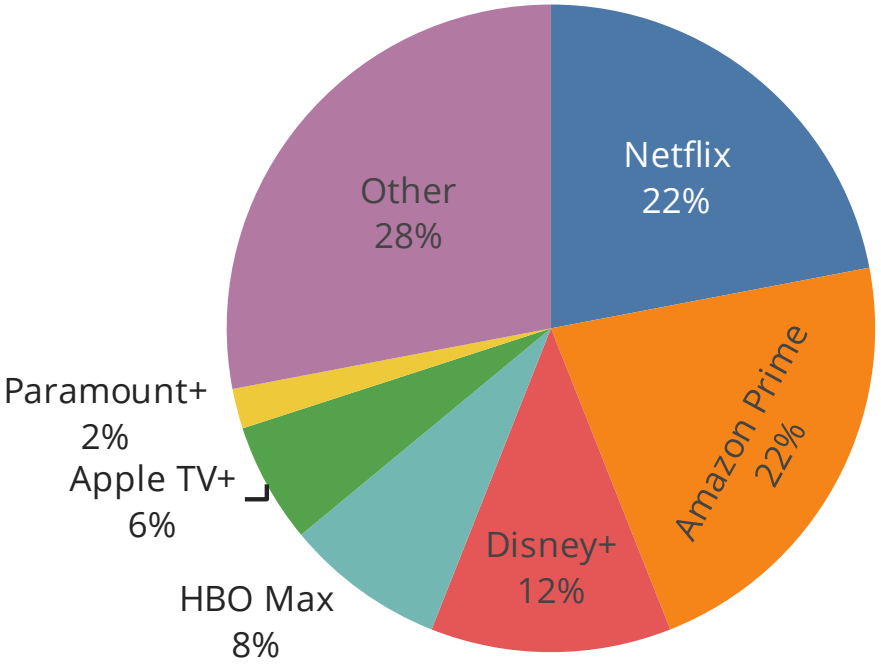
Bar plots can be perceptually heavy



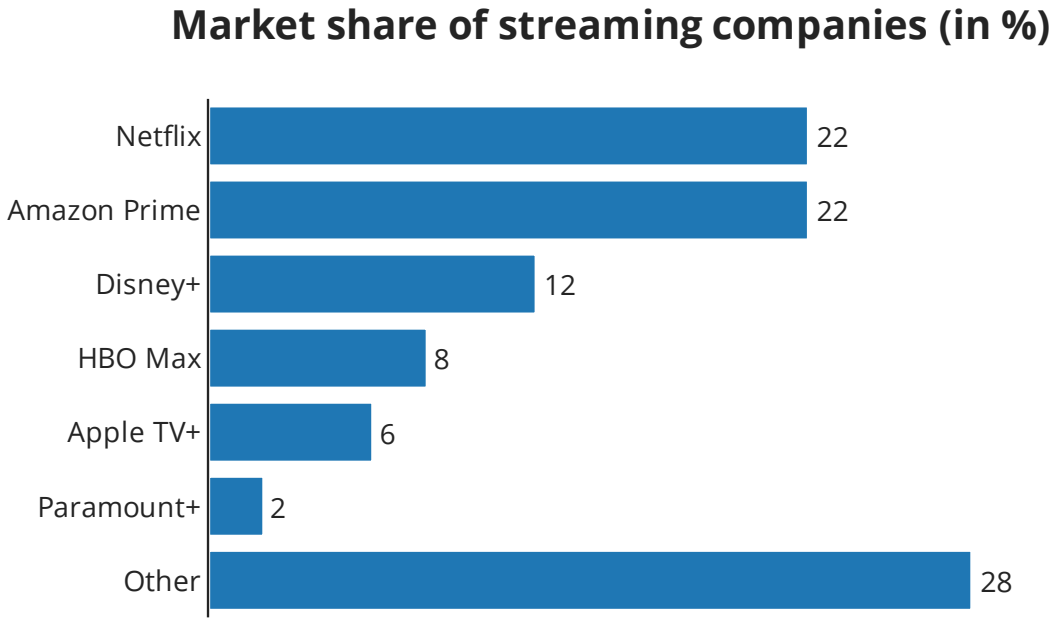
Scatterplots can be visually cleaner



What about the pie chart?

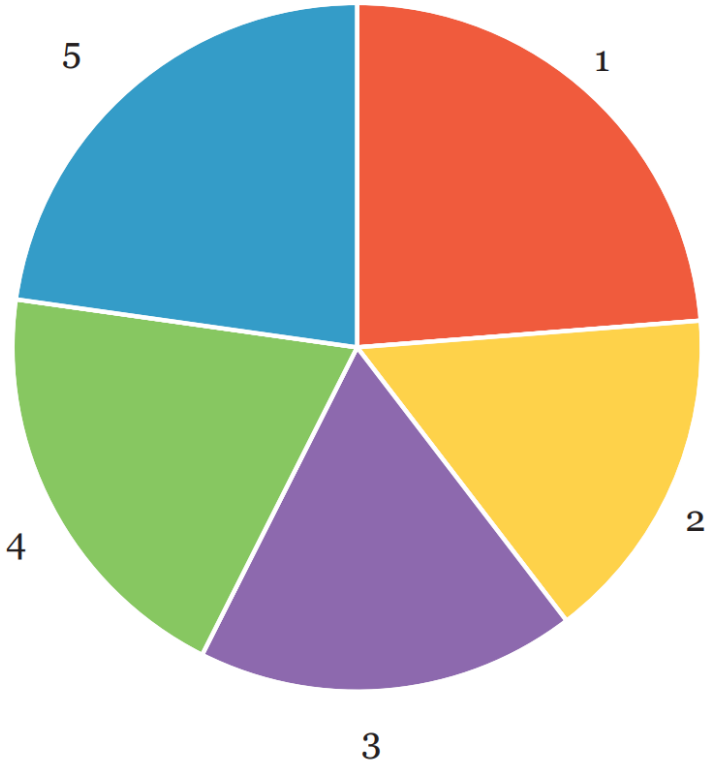


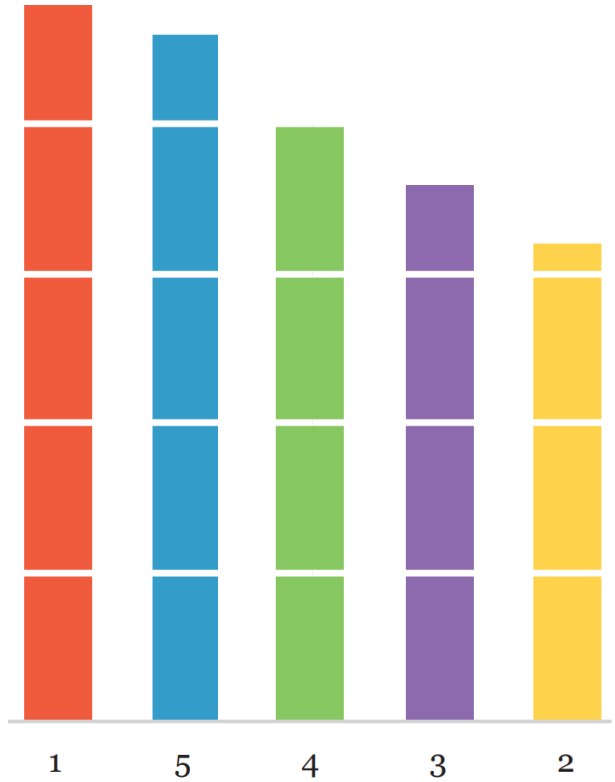
Simplifies % interpretation



Perceptually much easier

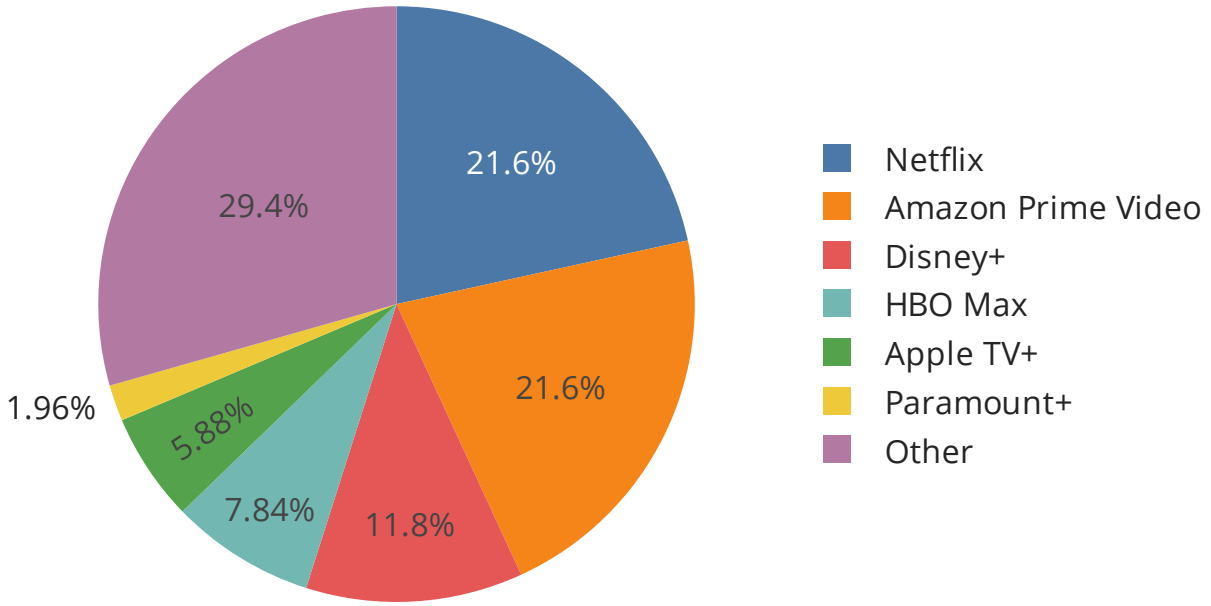
Rank the categories by size!





Perceptual problems of the pie chart

- **Proximity:** category names are often far away from the pie segments
- **Alignment:** consistent text alignment is difficult
- **Encoding effectiveness:** color and angle/area encodings are perceptually ineffective



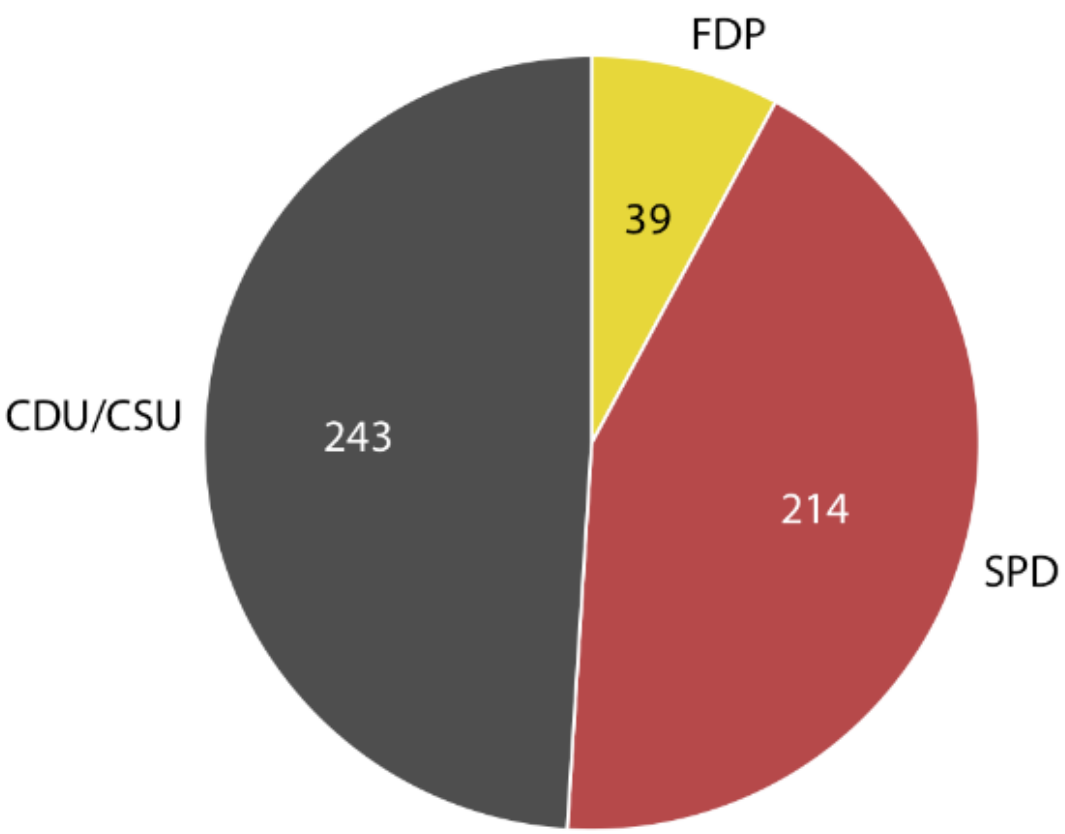
Famous statements on pie charts

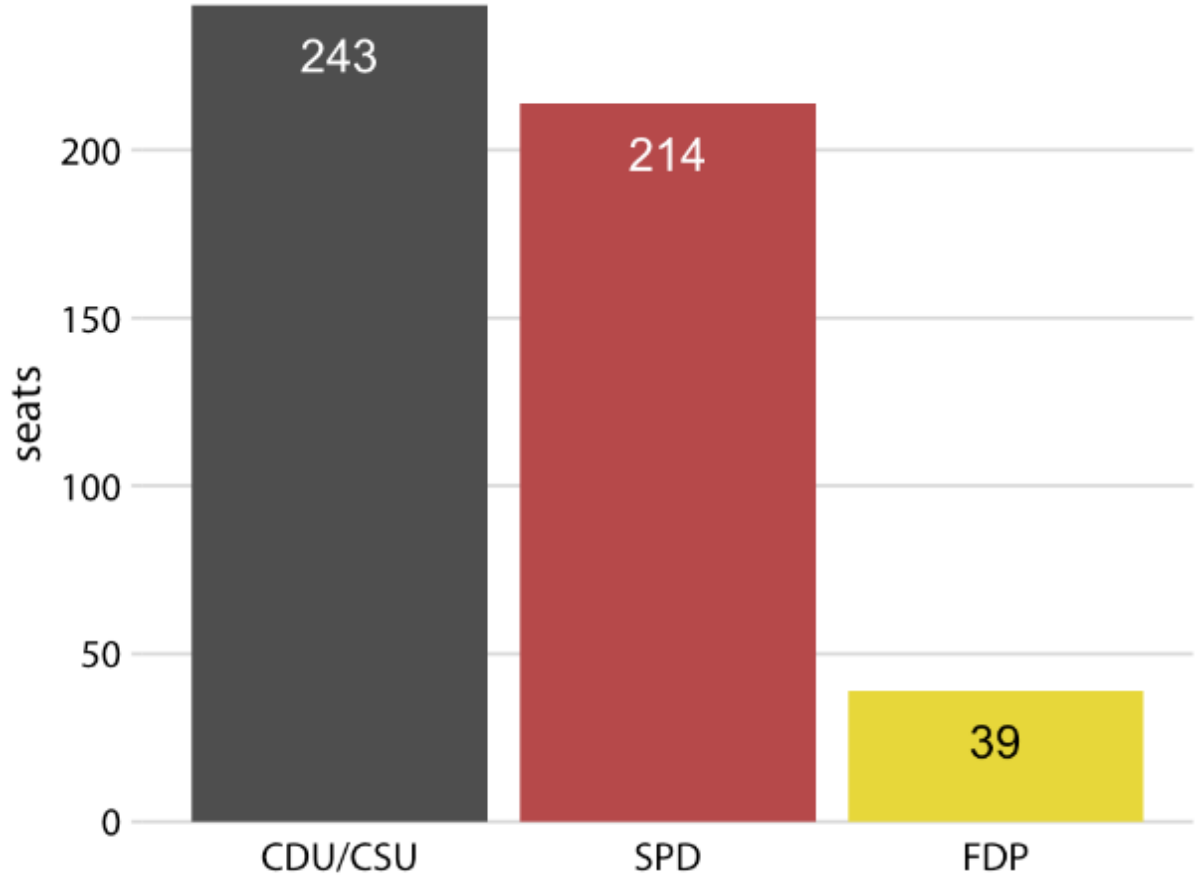
- **Edward Tufte**: „The only worse design than a pie chart is several of them…“
- **Cole Nussbaumer Knaflic**: „Pie charts are evil“

- **Alberto Cairo**: „A single pie chart with three or four segments is harmless. The problem is when we misuse pie charts, and we end up designing them with twelve or even twenty segments!
- **Jonathan Swabish: “**While the pie chart gets its fair share of complaints, it’s also a very familiar chart type for many people, and familiarity can be useful”.

A case for the pie chart

Here the pie chart highlights majority constellations in the German parliament, which the bar chart cannot do

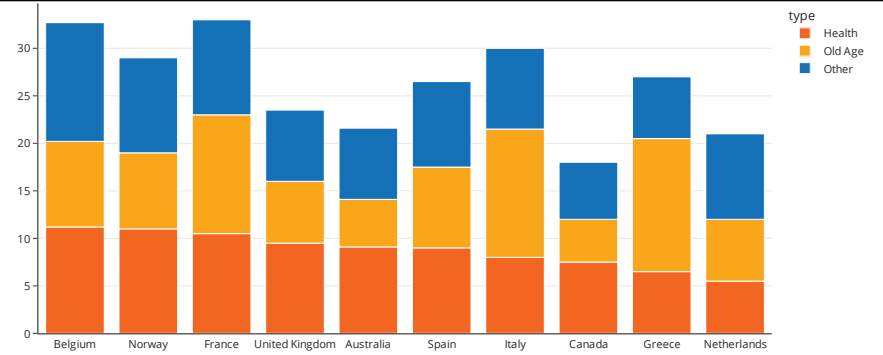




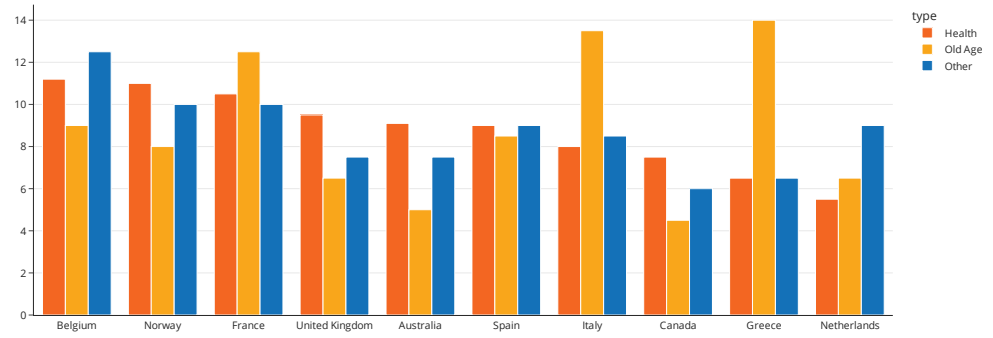
Two categorical variables – many design options

	country	type	expenditures
0	Australia	Health	9.1
1	Australia	Old Age	5.0
2	Australia	Other	7.5
3	Belgium	Health	11.2
4	Belgium	Old Age	9.0
5	Belgium	Other	12.5
6	Canada	Health	7.5
7	Canada	Old Age	4.5
8	Canada	Other	6.0
9	France	Health	10.5
10	France	Old Age	12.5

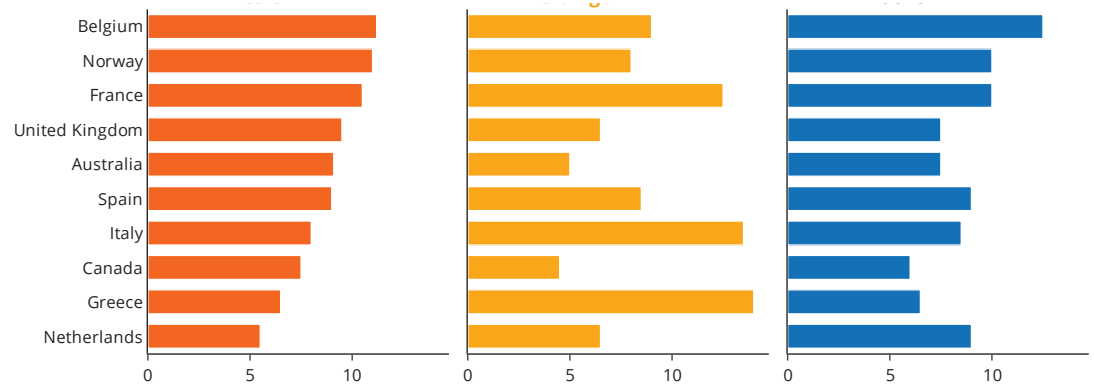
1. Stacked bars



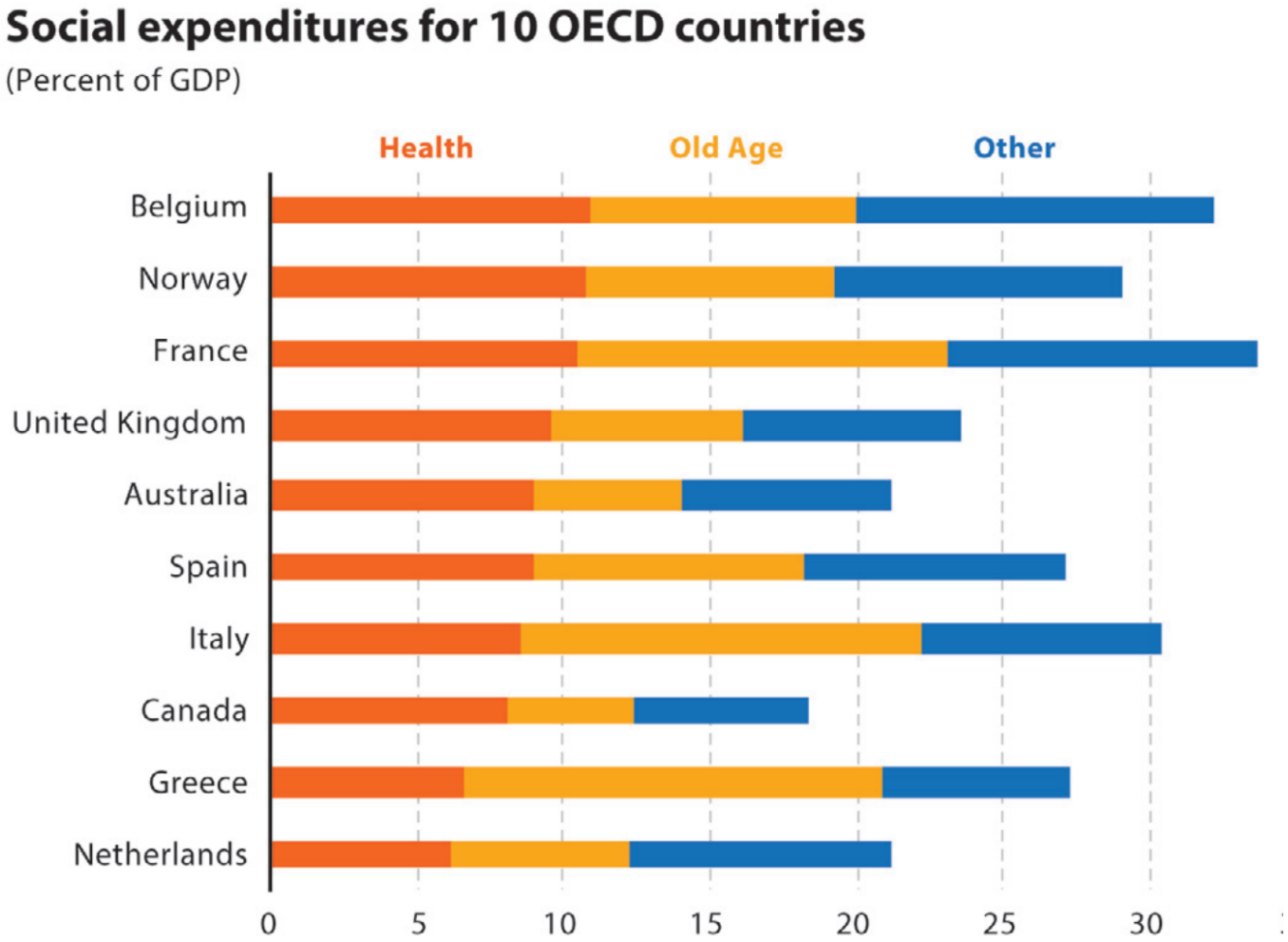
2. Grouped bars



3. Facets



Stacked bars



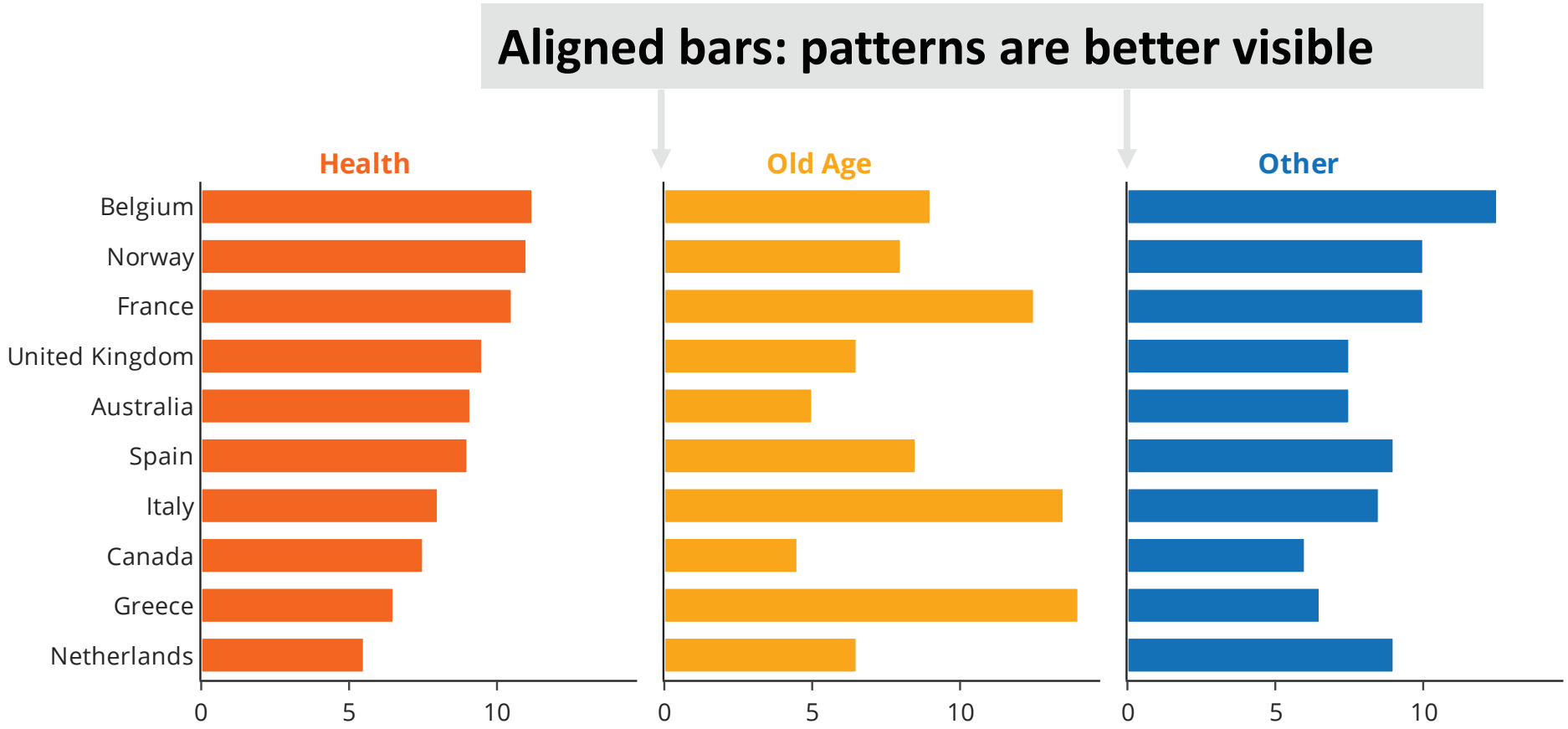
Good:

- Aligned and sorted bars for health → patterns are well perceived
- Total expenditures are immediately visible

Bad:

- Unaligned bars for old age and other → patterns are hard to see
- Information overload

Facets or Subplots



Preatentive Attributes

If a visualization contains a lot of information, the use of color or other **preattentive attributes** can be particularly effective to draw attention to important aspects