Fundamentals of Data Visualization

Chapter 6

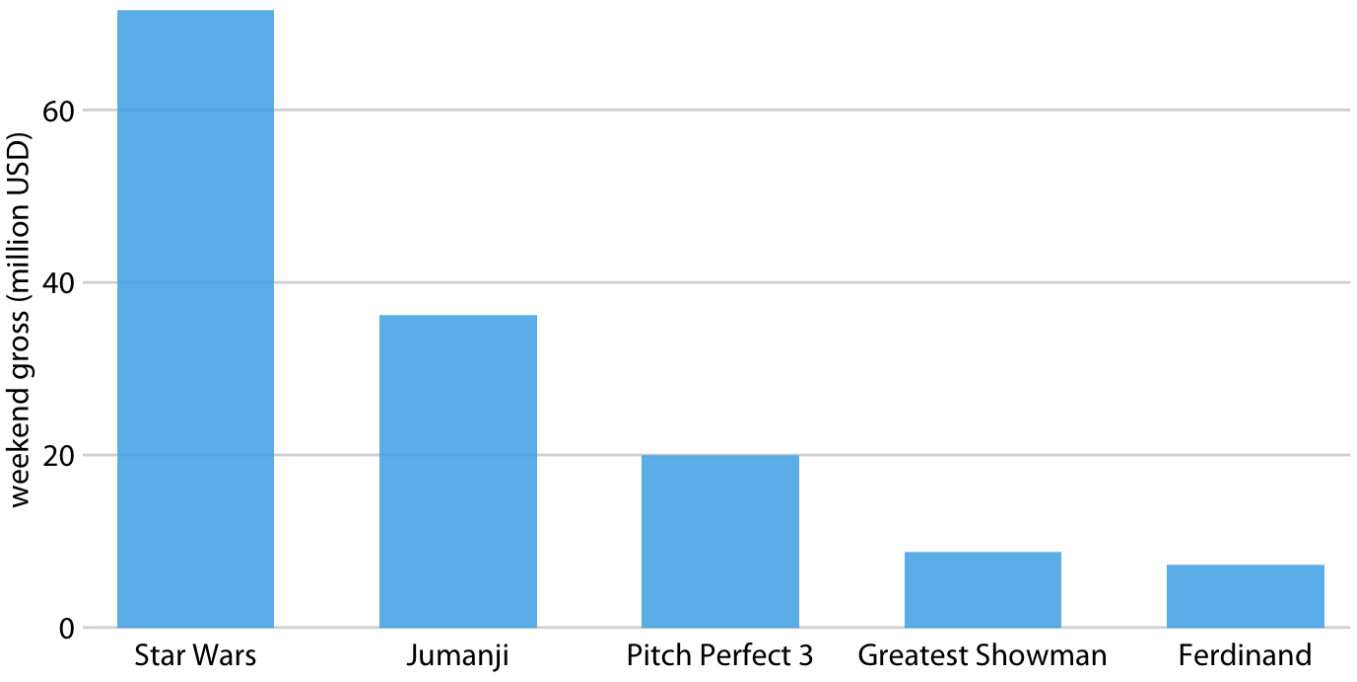
May 1, 2023

Visualizing Amounts

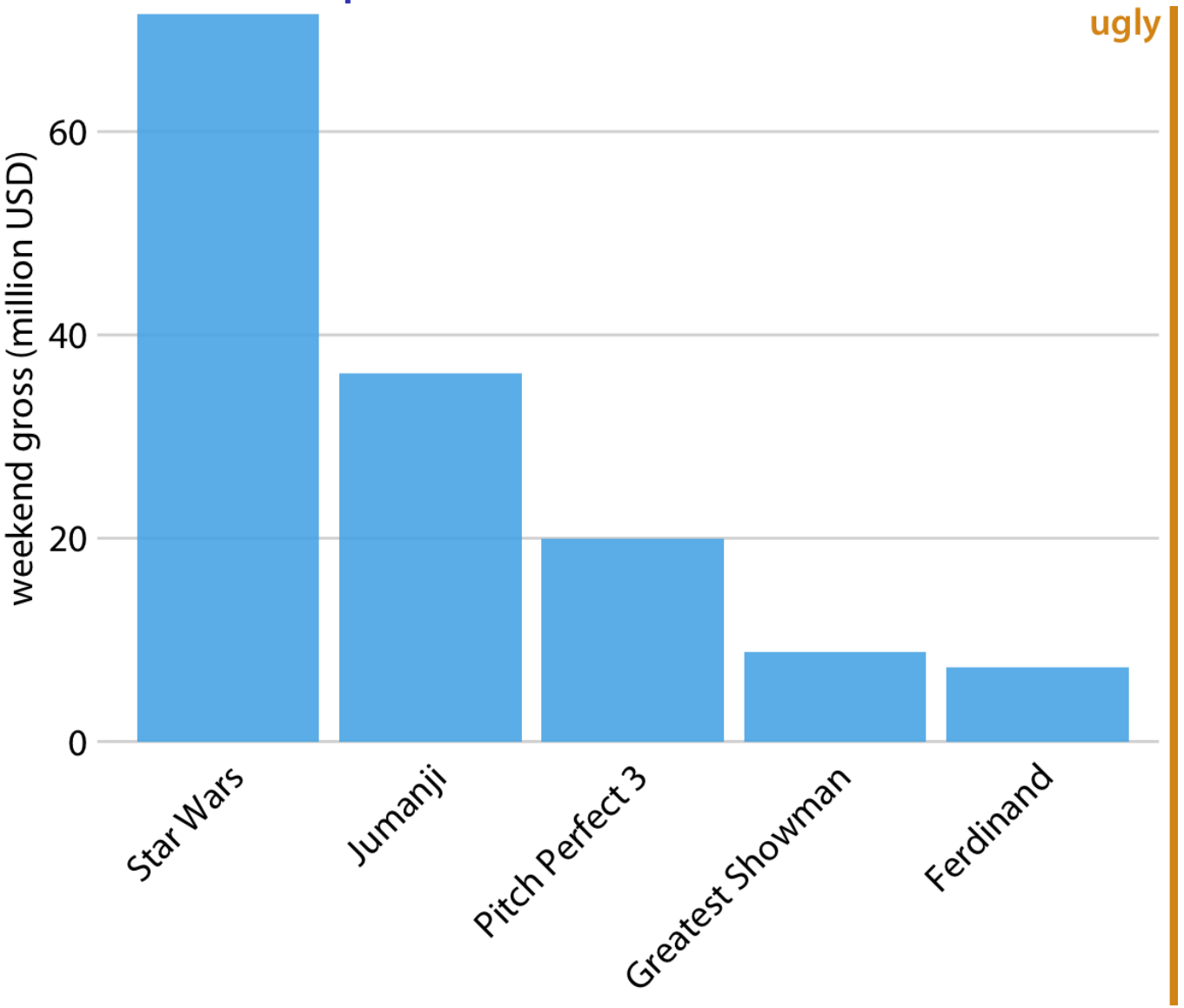
- In many scenarios, we are interested in the magnitude of some set of numbers.
  - total sales volume of different brands of cars
  - total number of people living in different cities
  - age of olympians performing different sports
- The standard visualization is the bar plot
- Alternatives are the dot plot and the heatmap.

Bar chart

Rank	Title	Weekend gross
1	Star Wars: The Last Jedi	$71,565,498
2	Jumanji: Welcome to the Jungle	$36,169,328
3	Pitch Perfect 3	$19,928,525
4	The Greatest Showman	$8,805,843
5	Ferdinand	$7,316,746

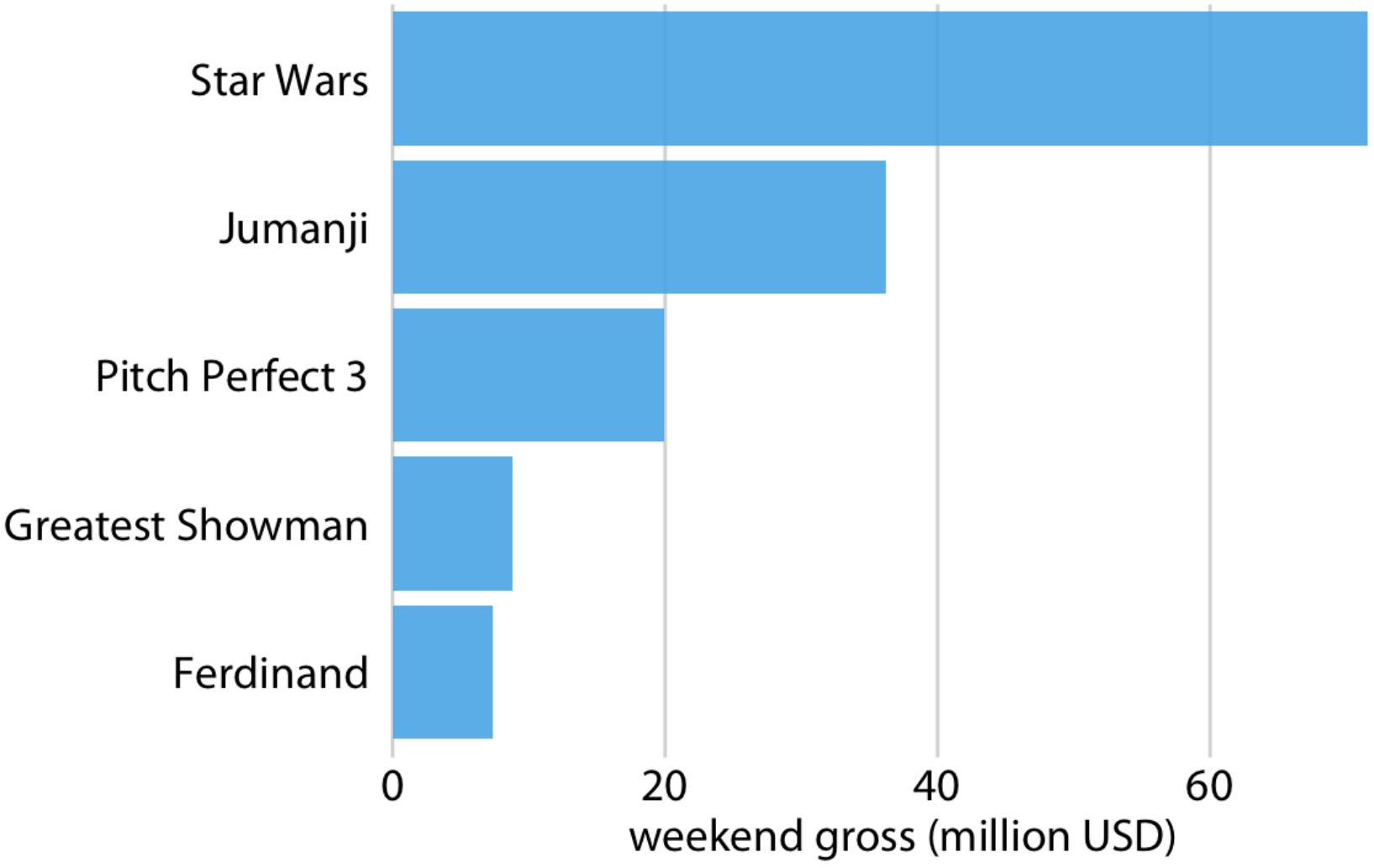


Bar chart problem: labels for each bar

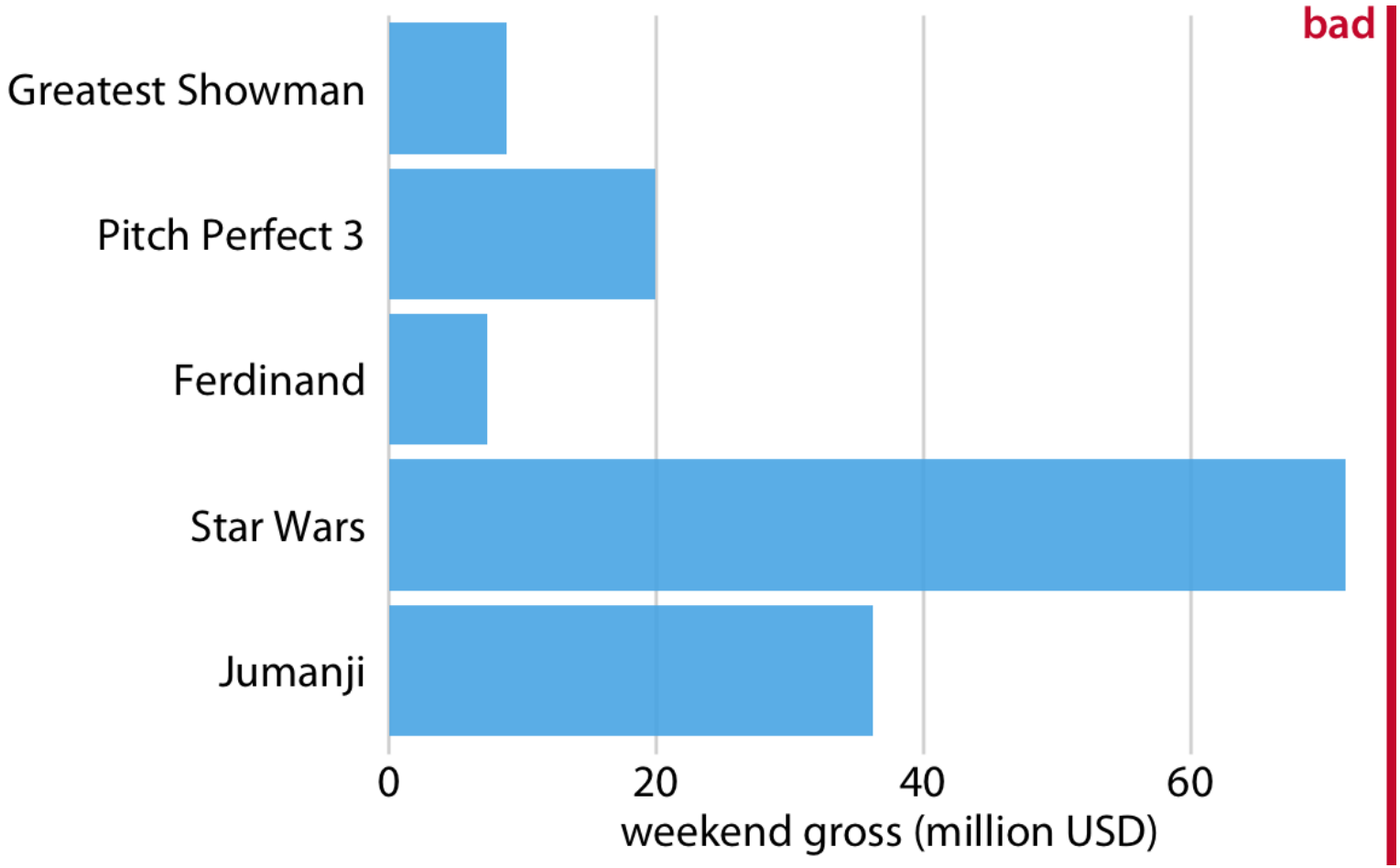


- Rotated labels are ugly

Better: horizontal bars

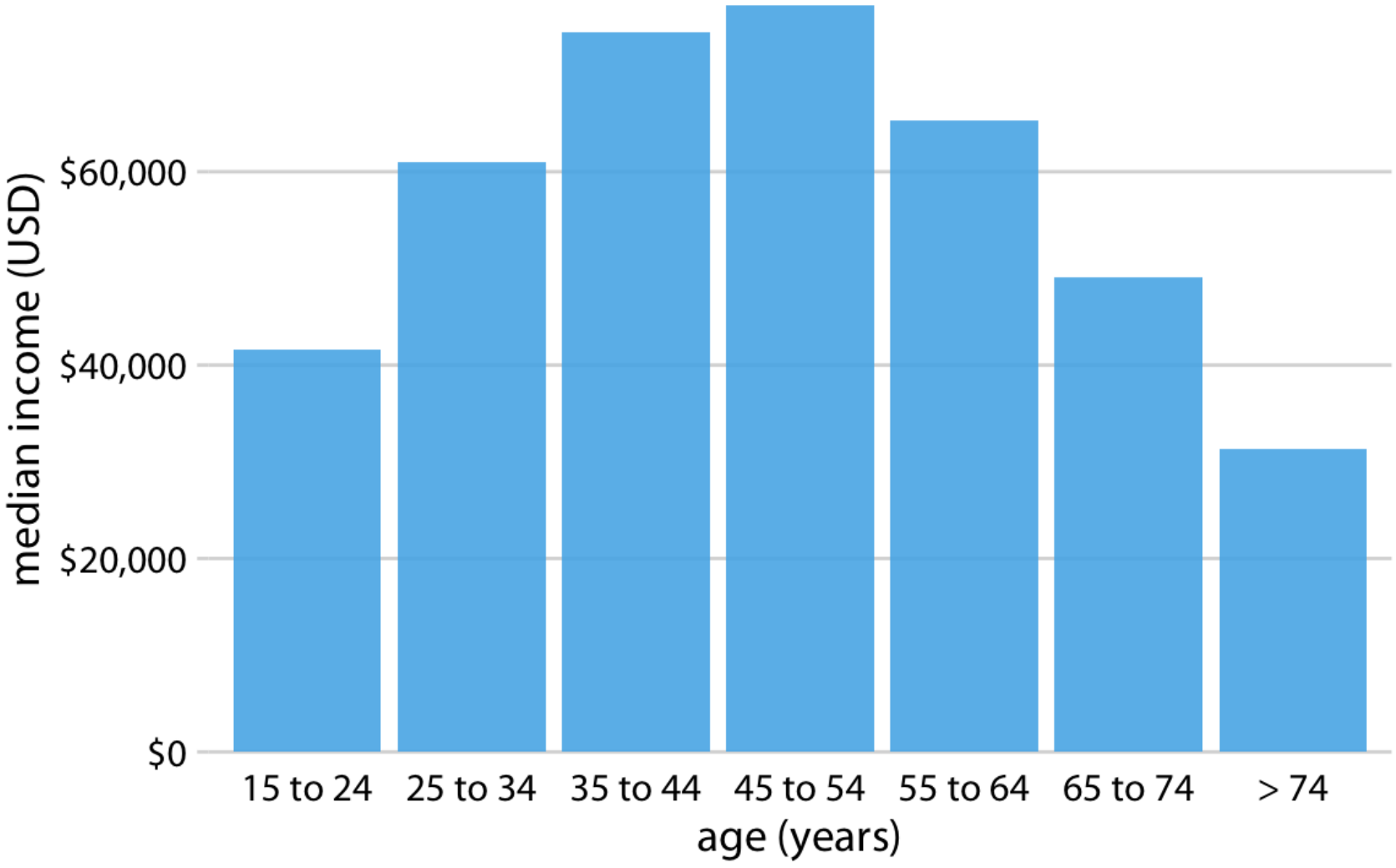


Order matters

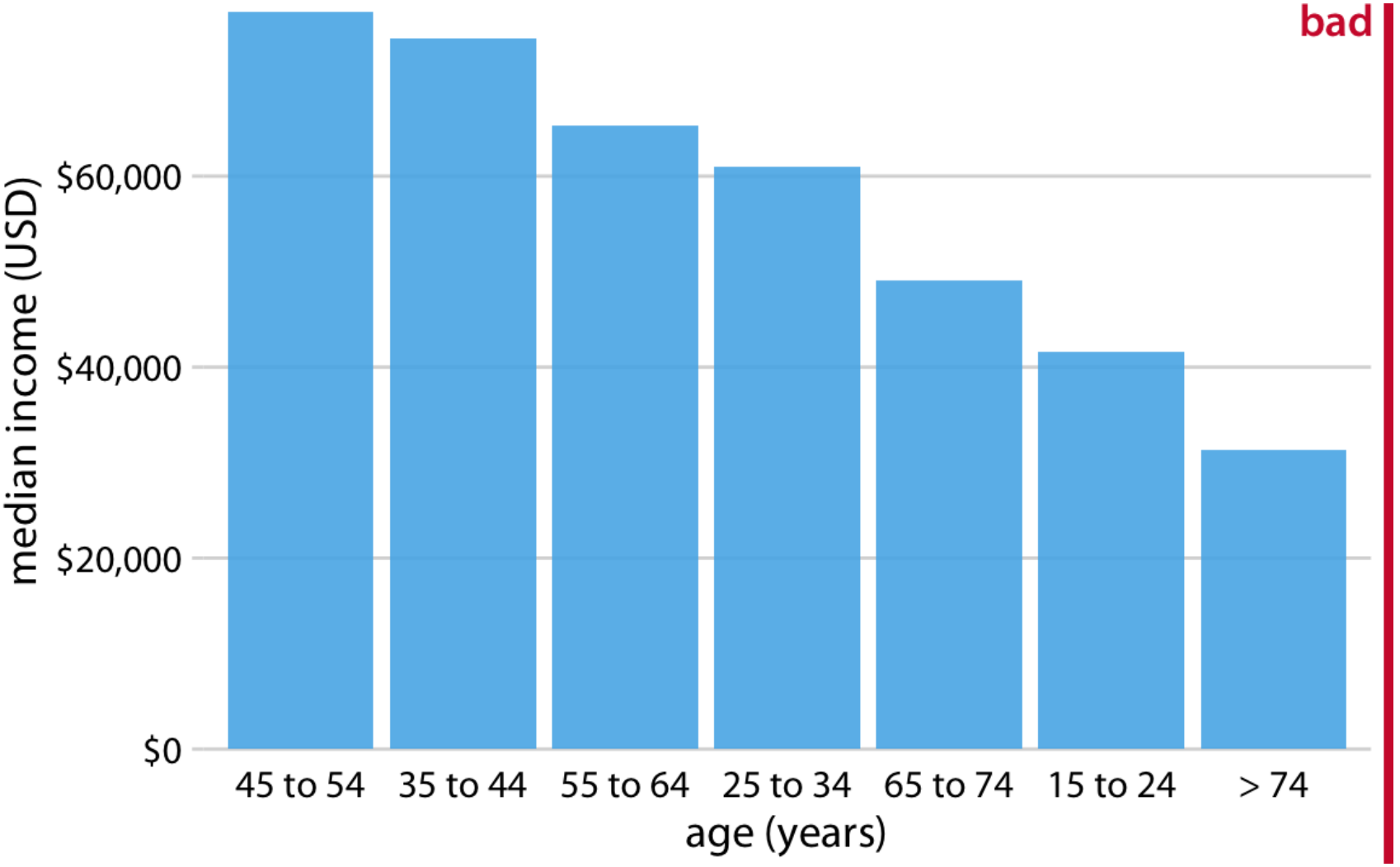


- Confusing and difficult to read
- If the bars represent unordered categories, order them by ascending or descending data values.

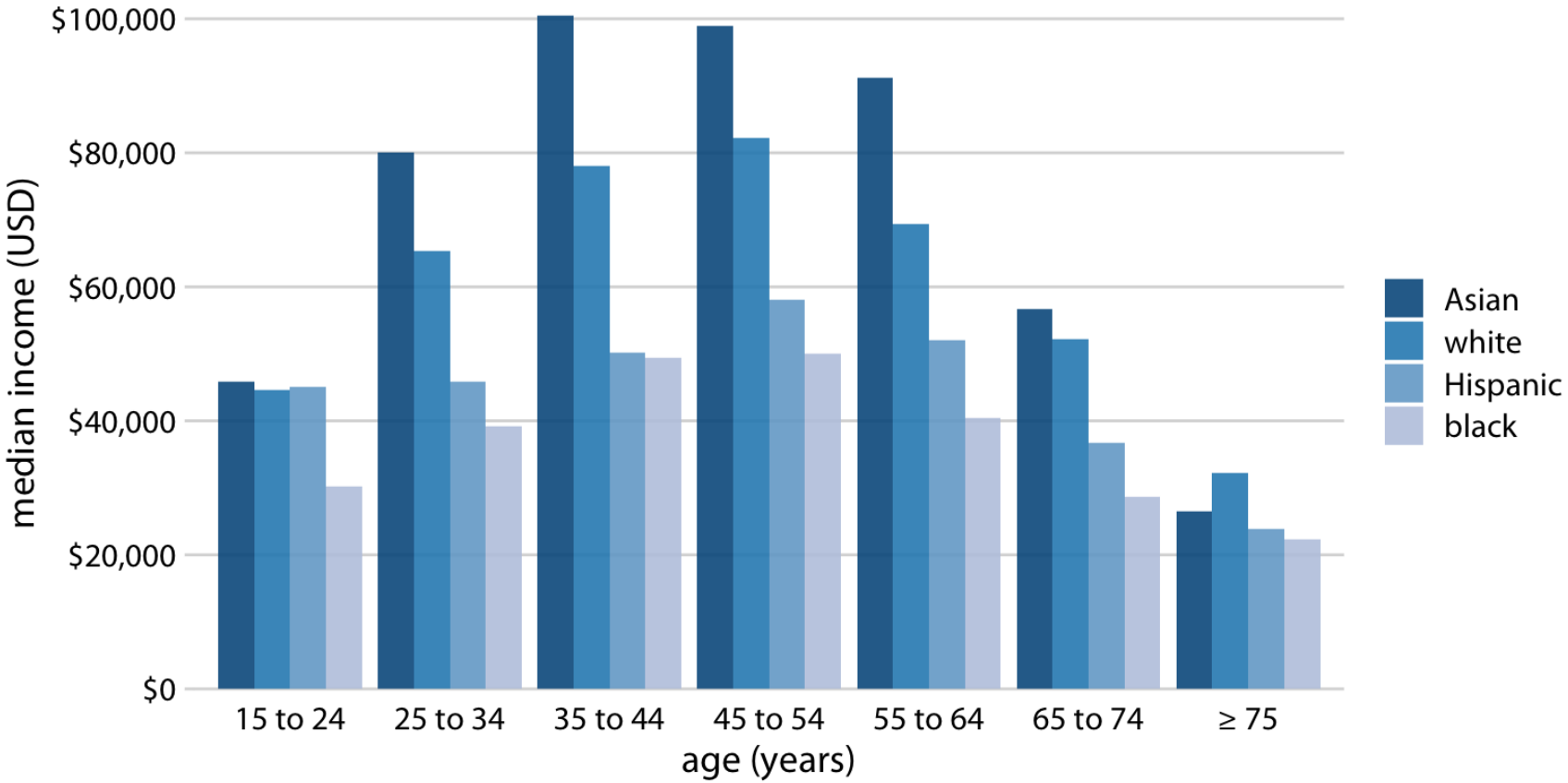
Use natural orders



Sometimes sorting by bar length makes no sense

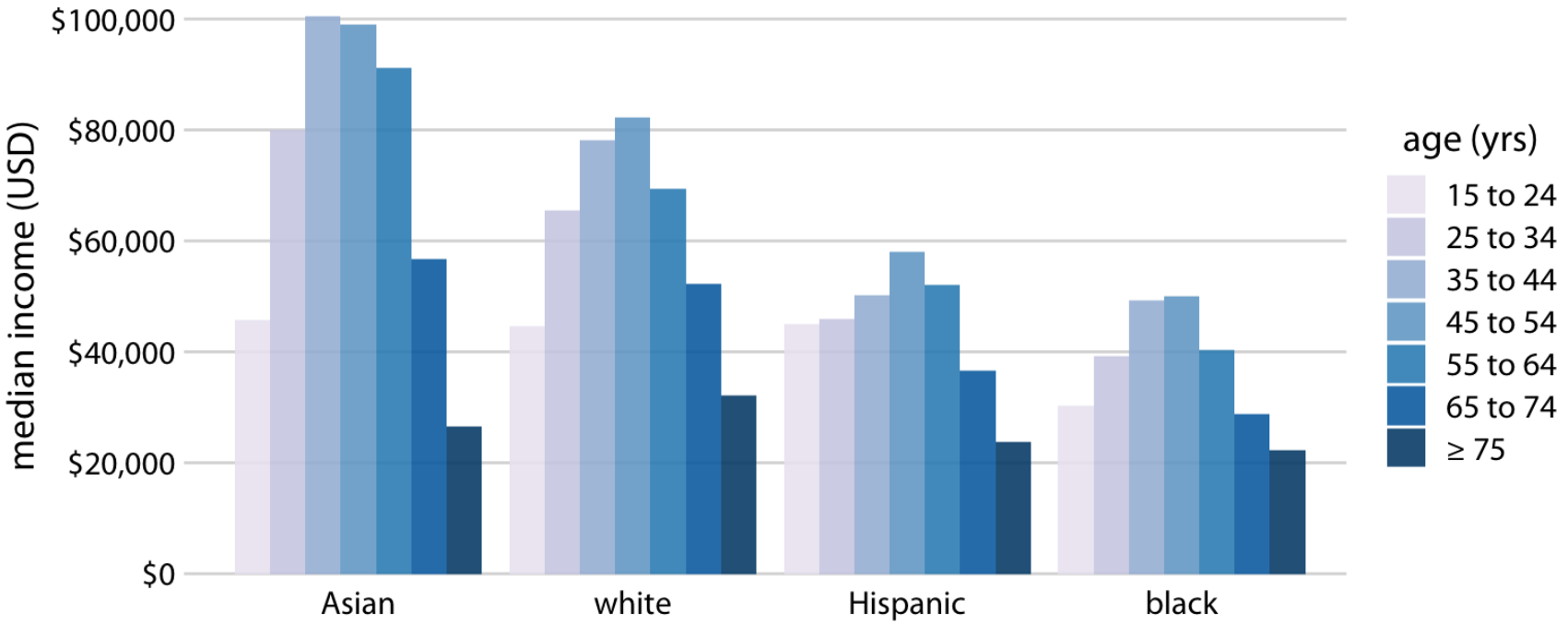


Grouped bar chart



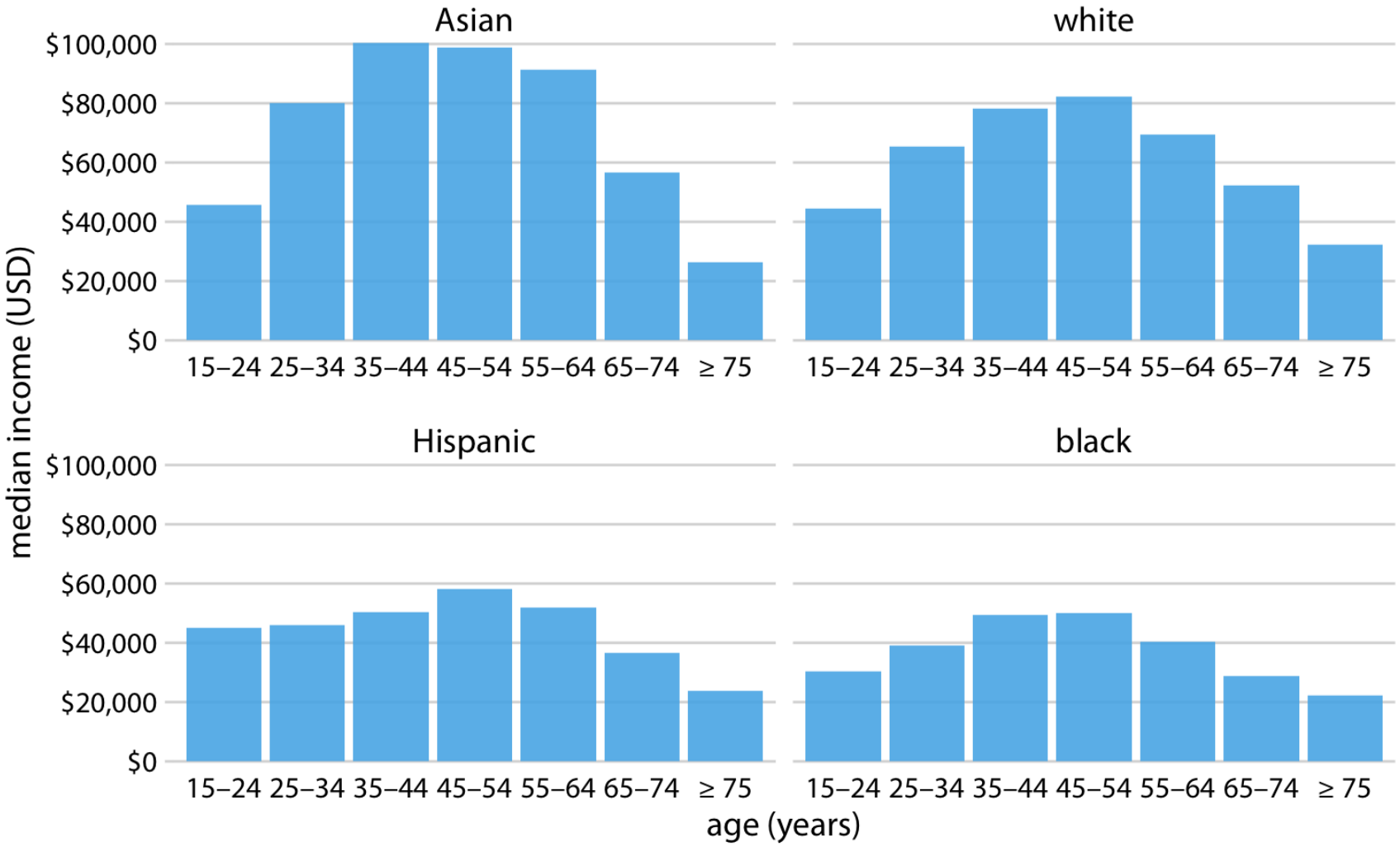
- More than one category
- Can be confusing
  - e.g. difficult to compare median incomes across age groups for a given racial group

Replotting can help



- Makes it easy to compare incomes within racial groups

Use four plots instead of four colors



- Generally easier to see position than color

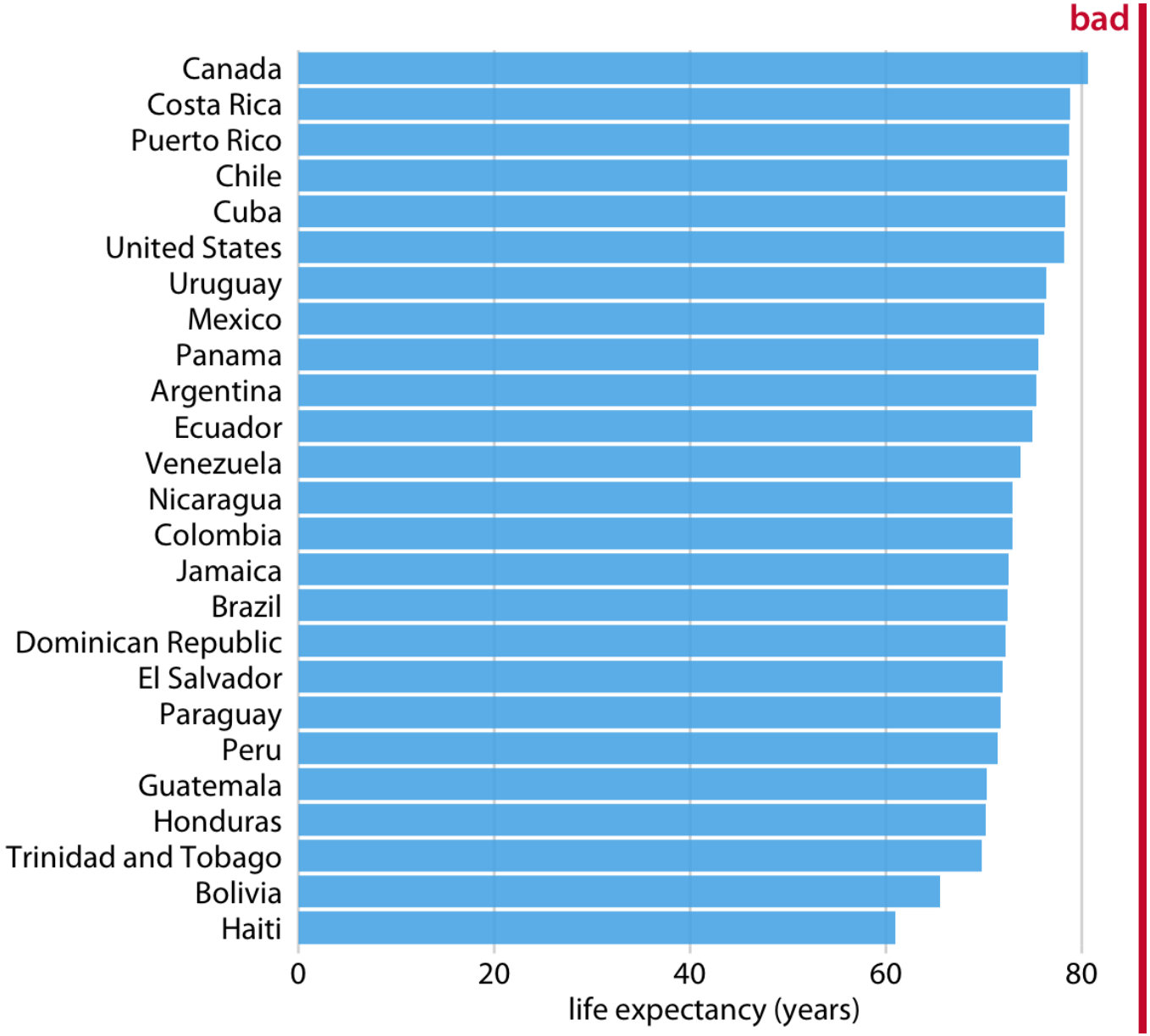
Stacked boxplots

- Plotted values must be summable
- The sum of two medians is meaningless
- The sum of two gross movie values is meaningful
- Stacking most useful for counts

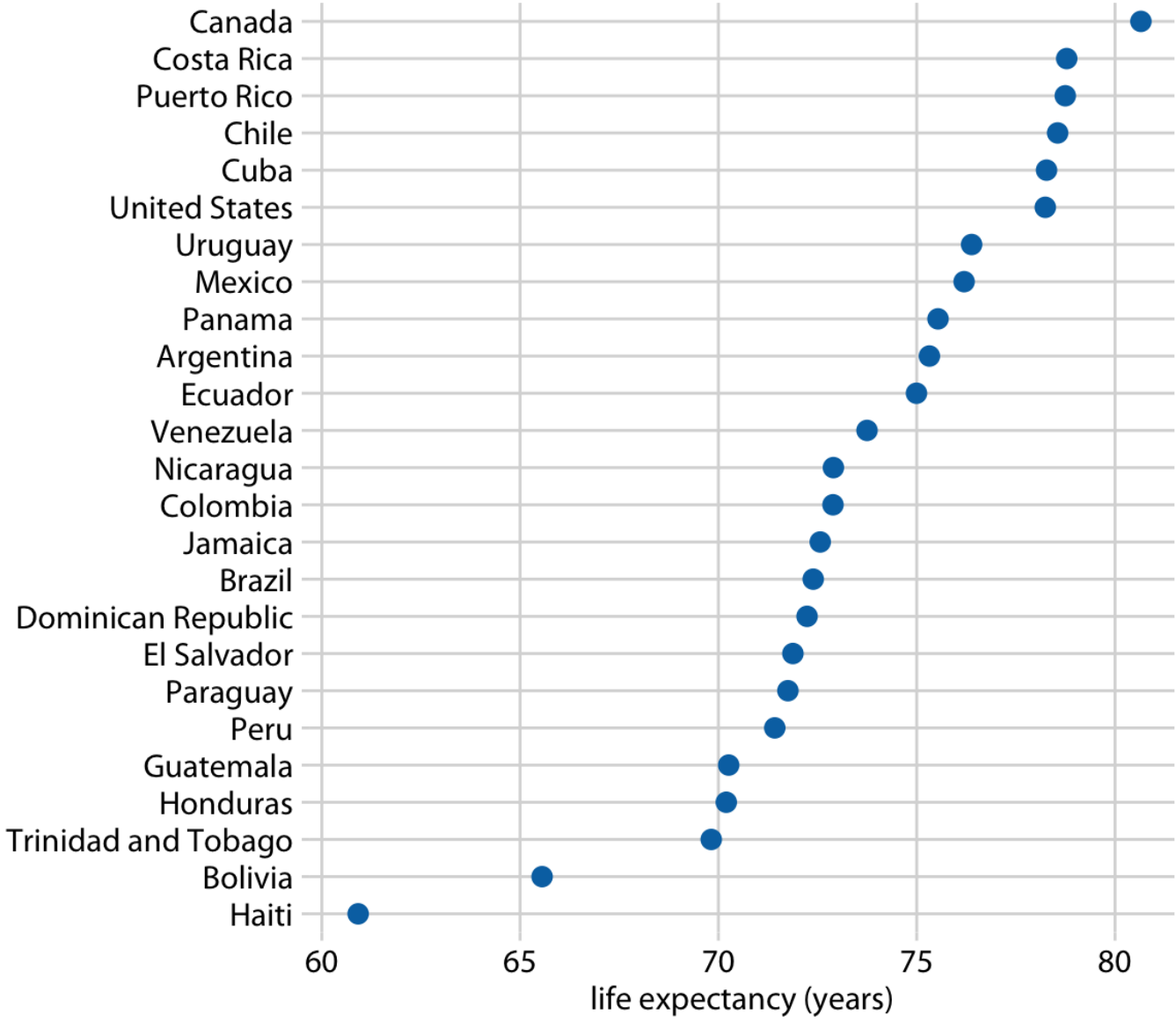
Stacked boxplots

- In this graph, y axis not needed
- Whenever the data is a small number of values, show the actual values.

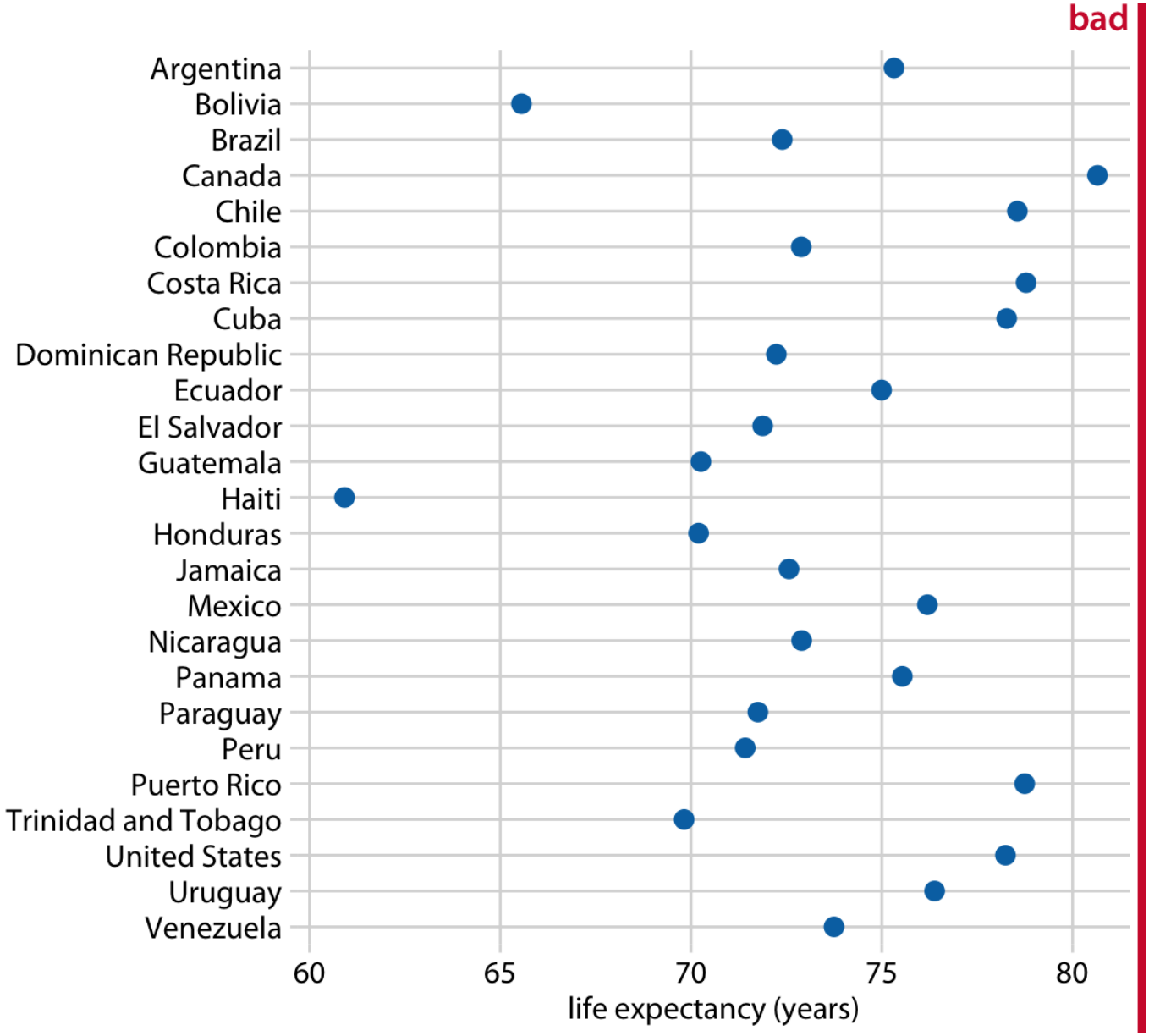
Boxplots must start at zero



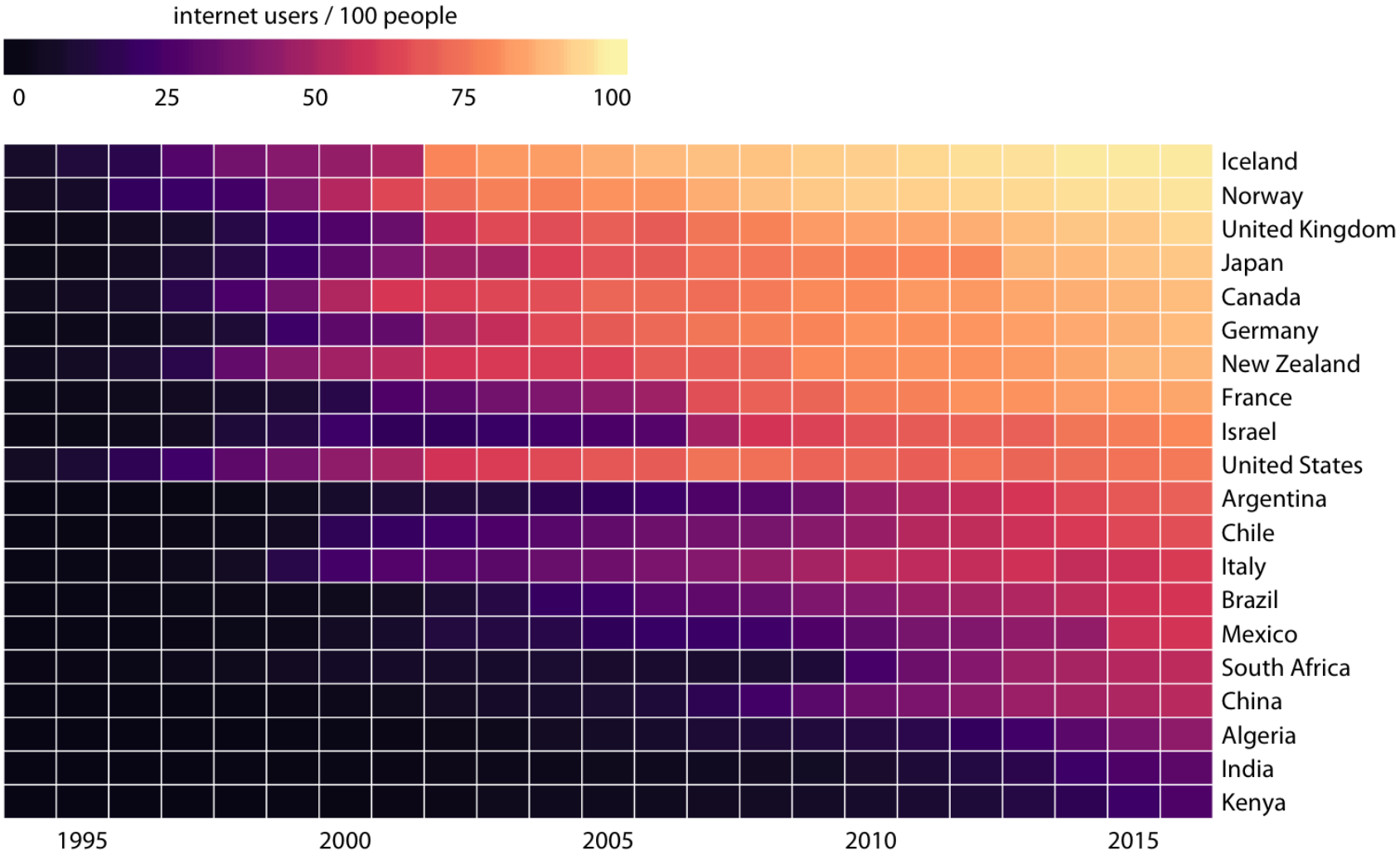
Use dot plot if you don't start at zero



Pay attention to factor ordering

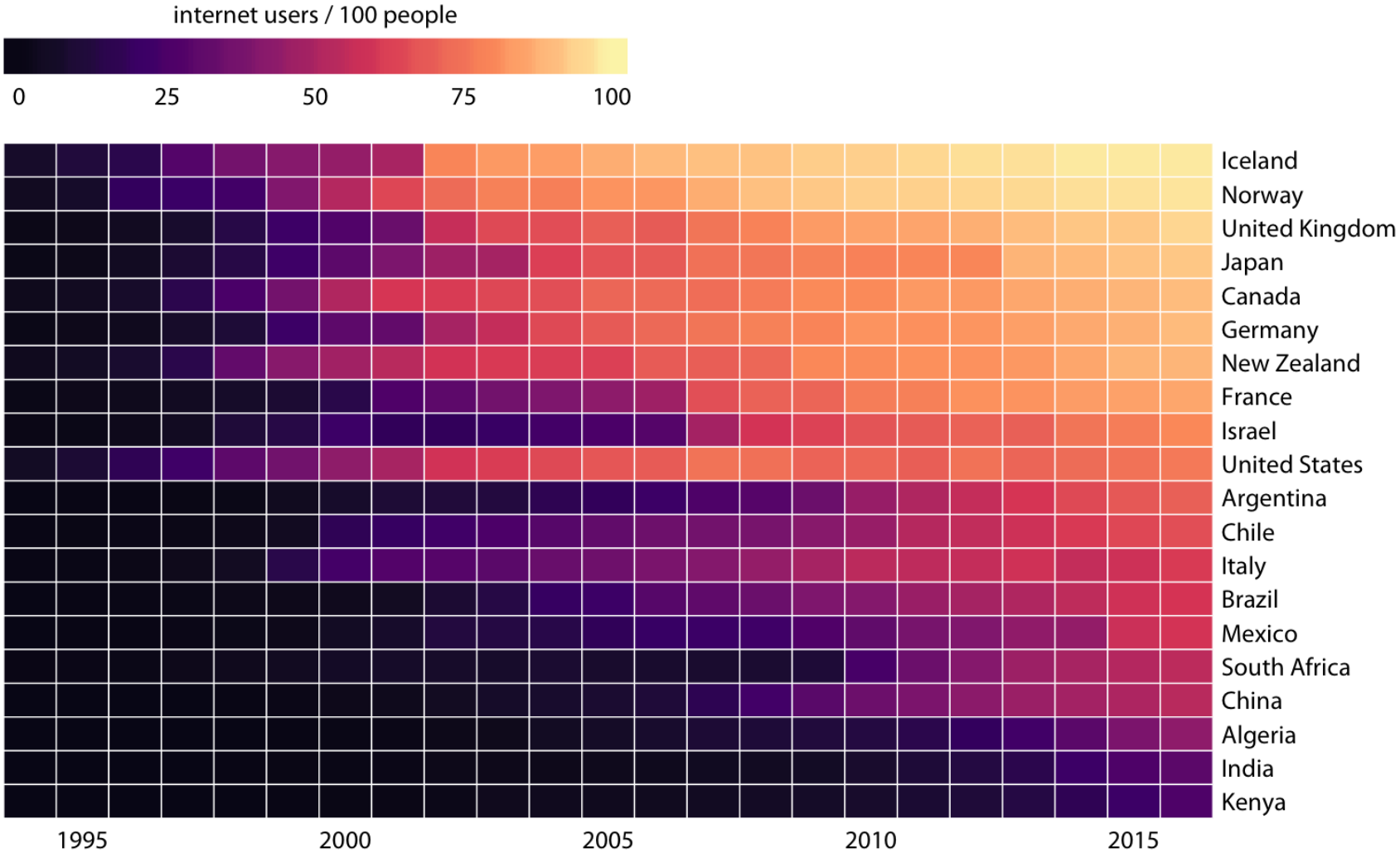


Heat maps



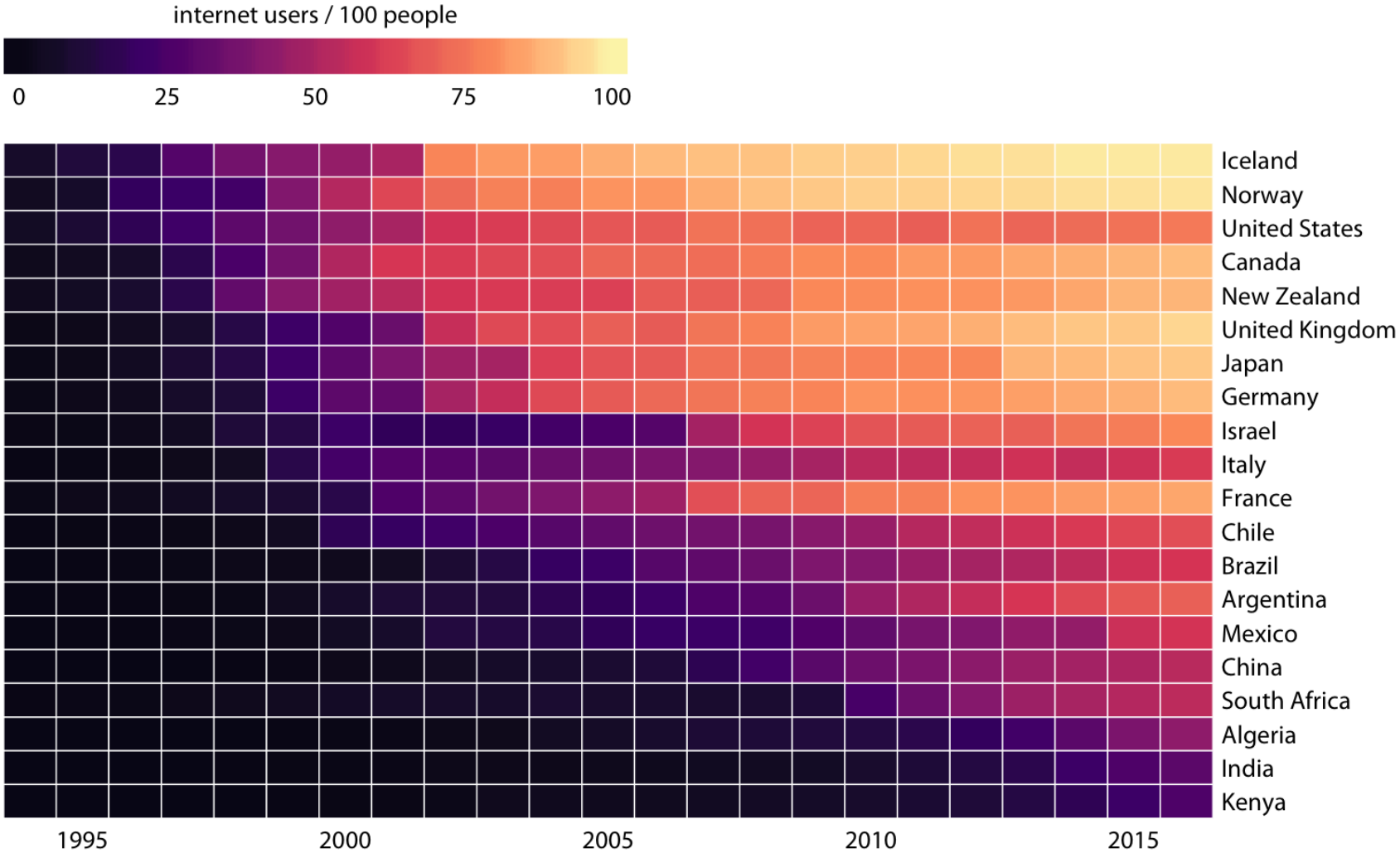
- Instead of position, we use color to represent value.
- Not as good for exact values, but gives a strong visual impression.

Pay attention to ordering



- Ordered by percentage in 2016

Pay attention to ordering



- Ordered by year in which usage was first over 20%