

CP321 Data Visualization

- Visualize Amount
and
complete the visualization

Jiashu (Jessie) Zhao

- Visualize Amount
 - Bars, Grouped Bars, Stacked Bars, Dots, Heatmap
- Complete your visualization: Illustrate the figure with Title, Caption and Annotations

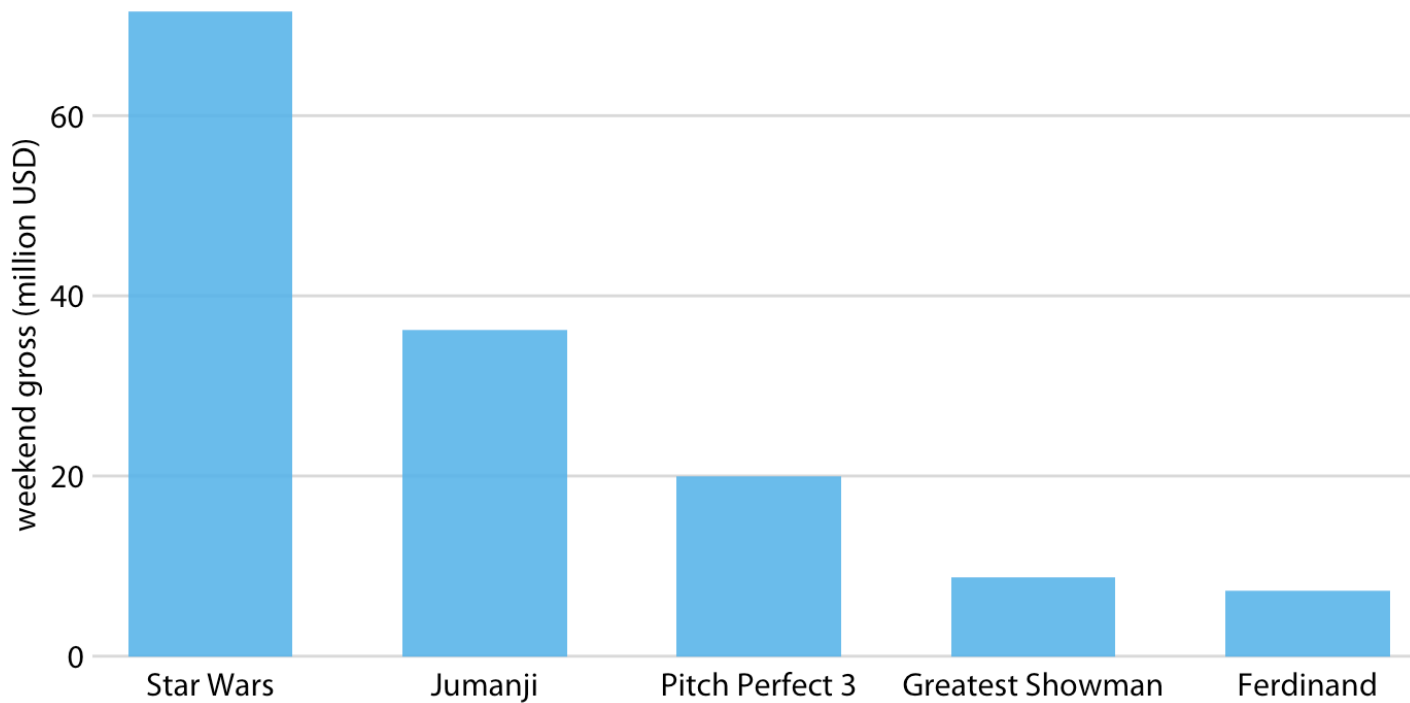
Visualizing amounts

- For the scenarios that we are interested in the magnitude of some set of numbers
- We have a set of categories (e.g., brands of cars, cities, or sports) and a quantitative value for each category.
- The standard visualization in this scenario is the **bar chart**. Variation include **grouped and stacked bars**.
- Alternatives to the bar plot are the **dot plot** and the **heatmap**.

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

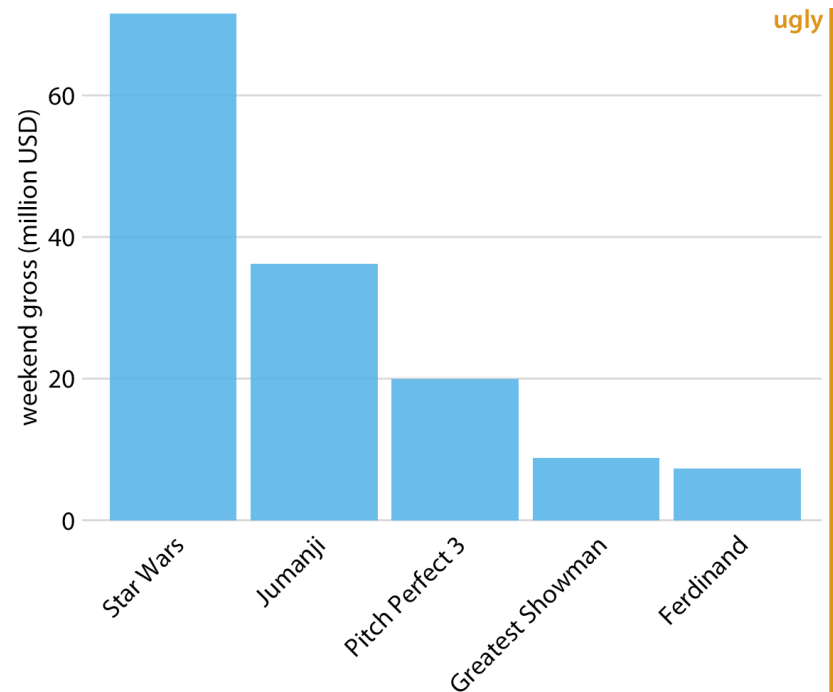
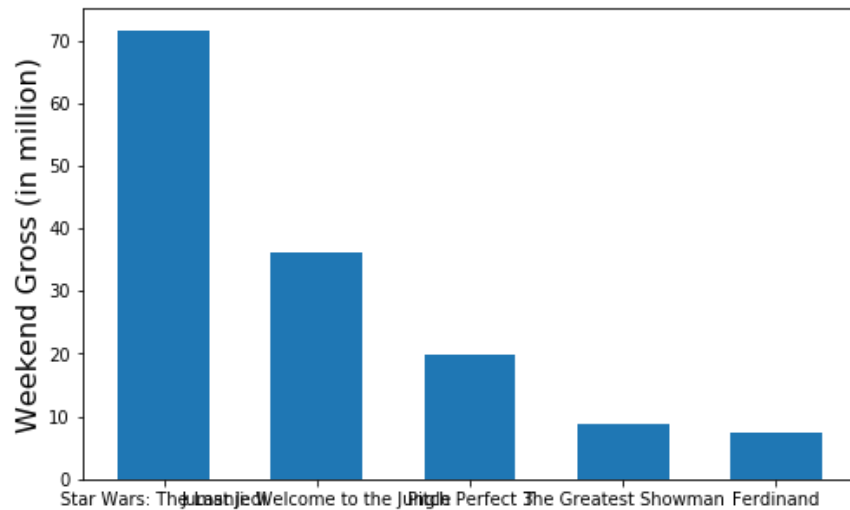
Bar Plot/Chart

- Commonly visualized with vertical bars.

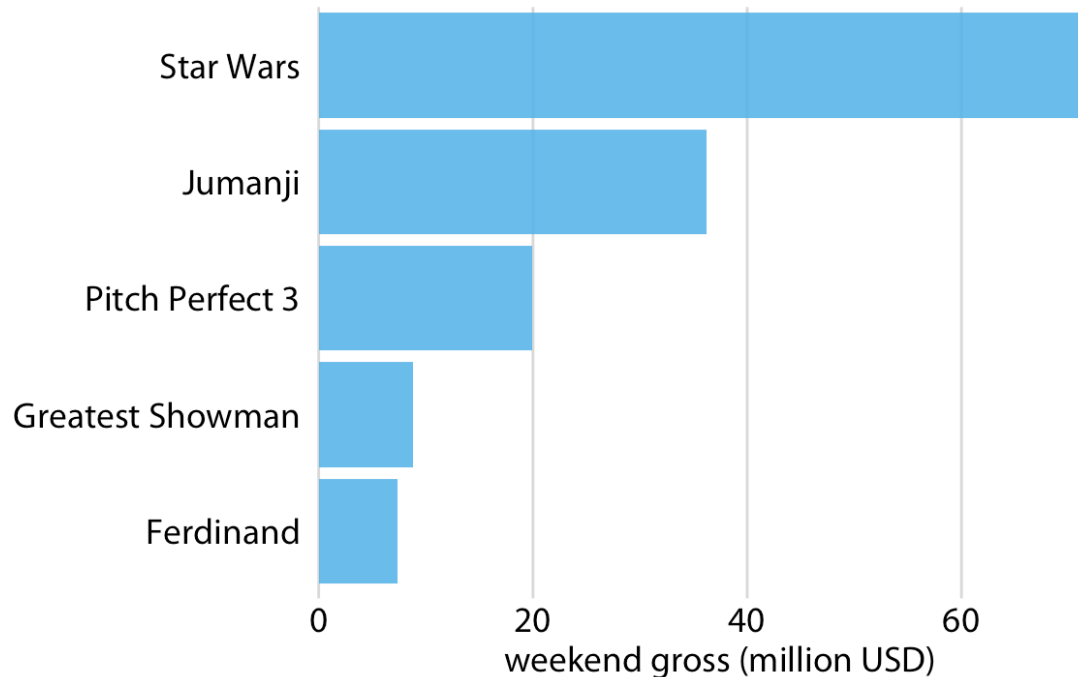


- A bar plot/chart
 - presents categorical data
 - with rectangular bars
 - the bars' heights or lengths are proportional to the values that they represent.
 - One axis of the chart shows the specific categories being compared
 - the other axis represents a measured value.
- The bars can be plotted vertically or horizontally.

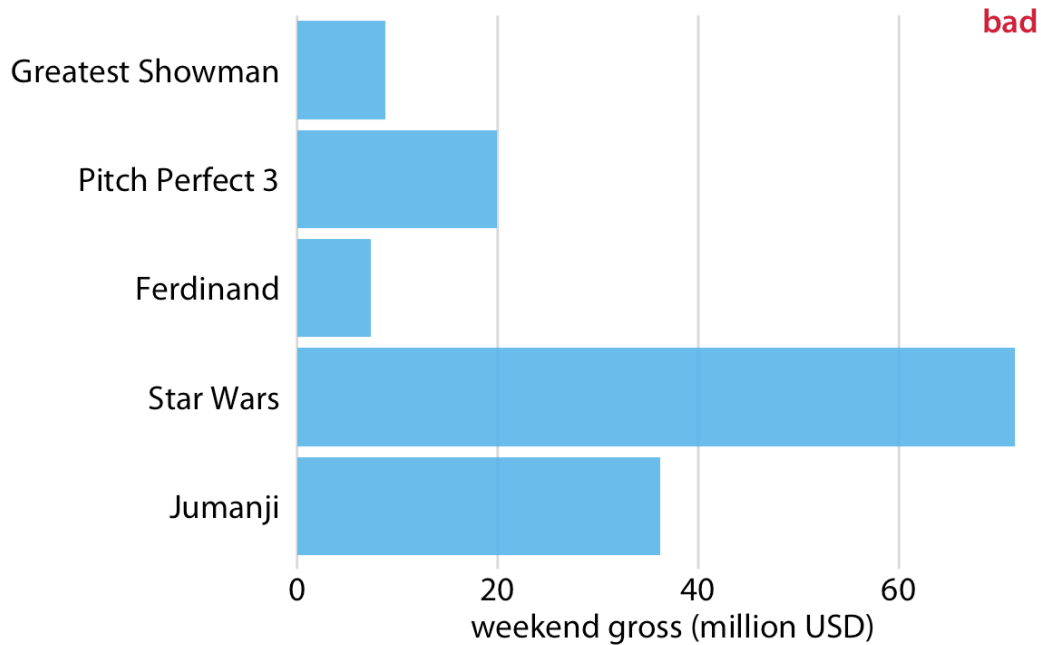
- One problem we commonly encounter with vertical bars is that the labels identifying each bar take up a lot of horizontal space.



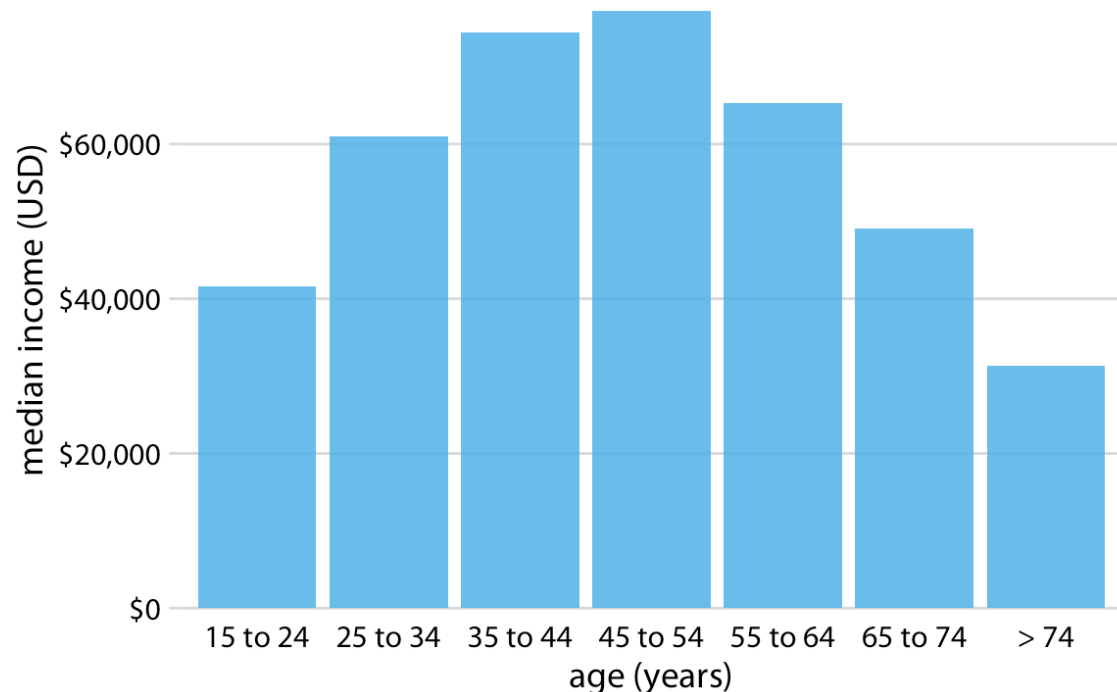
- The better solution for long labels is usually to swap the x and the y axis, so that the bars run horizontally

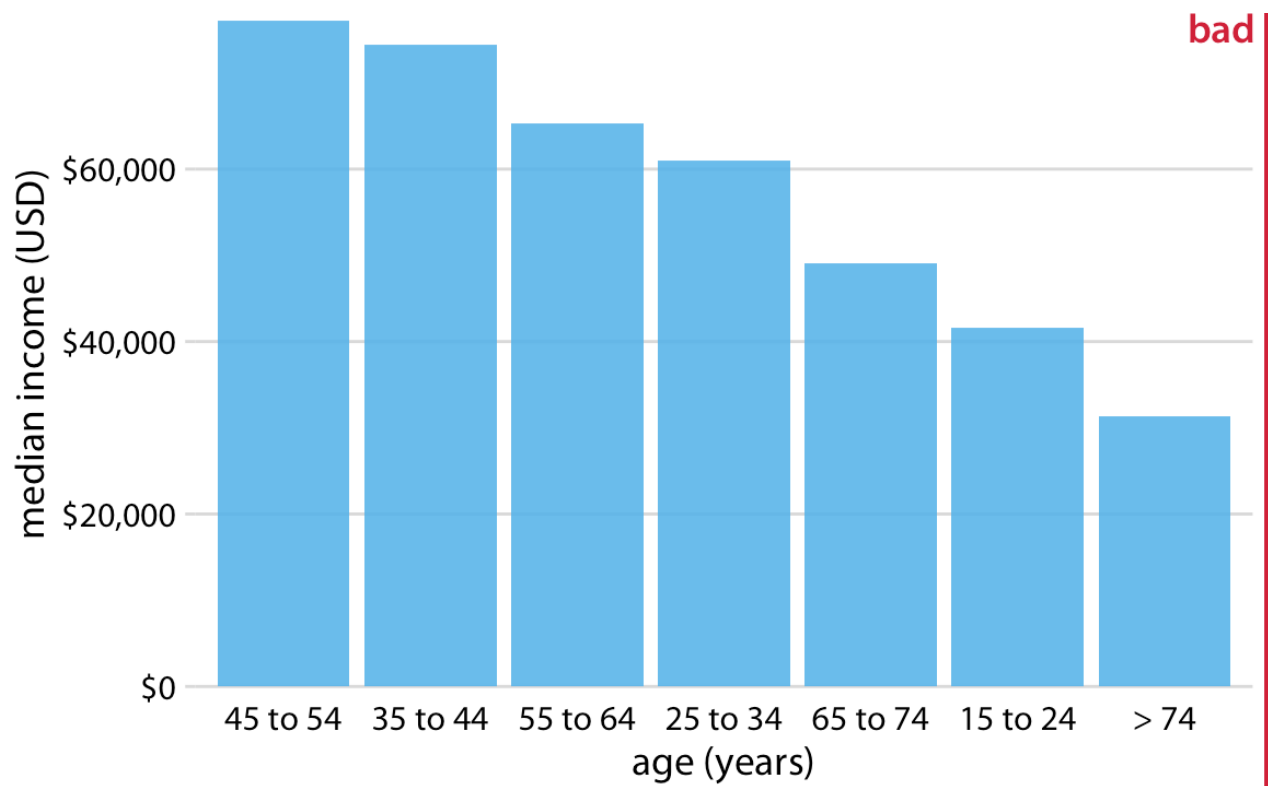


- Regardless of whether we place bars vertically or horizontally, we need to pay attention to the order in which the bars are arranged.



- Whenever there is a natural ordering (i.e., when our categorical variable is an ordered factor) we should retain that ordering in the visualization.

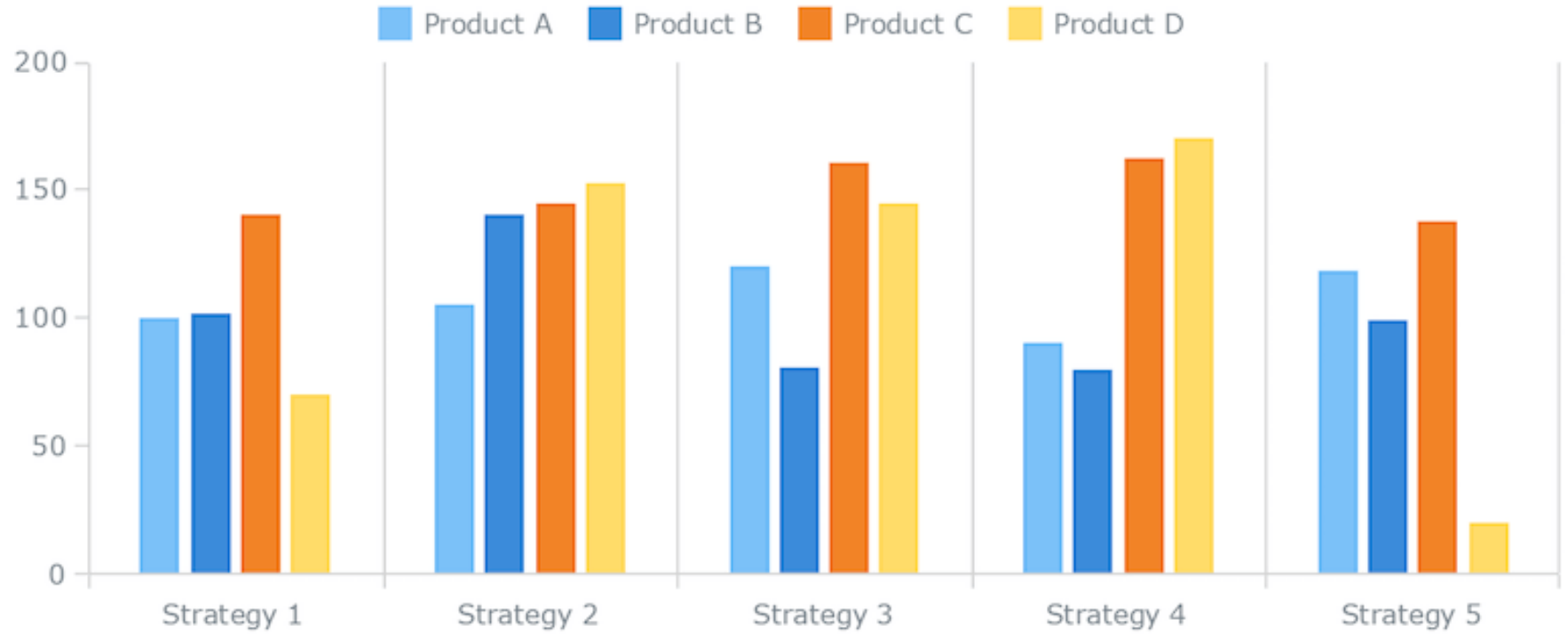




Grouped bars

- When we are interested in two categorical variables at the same time, we can visualize this dataset with a *grouped bar plot*.
 - we first draw a group of bars at each position along the x axis, determined by one categorical variable
 - then we draw bars within each group according to the other categorical variable

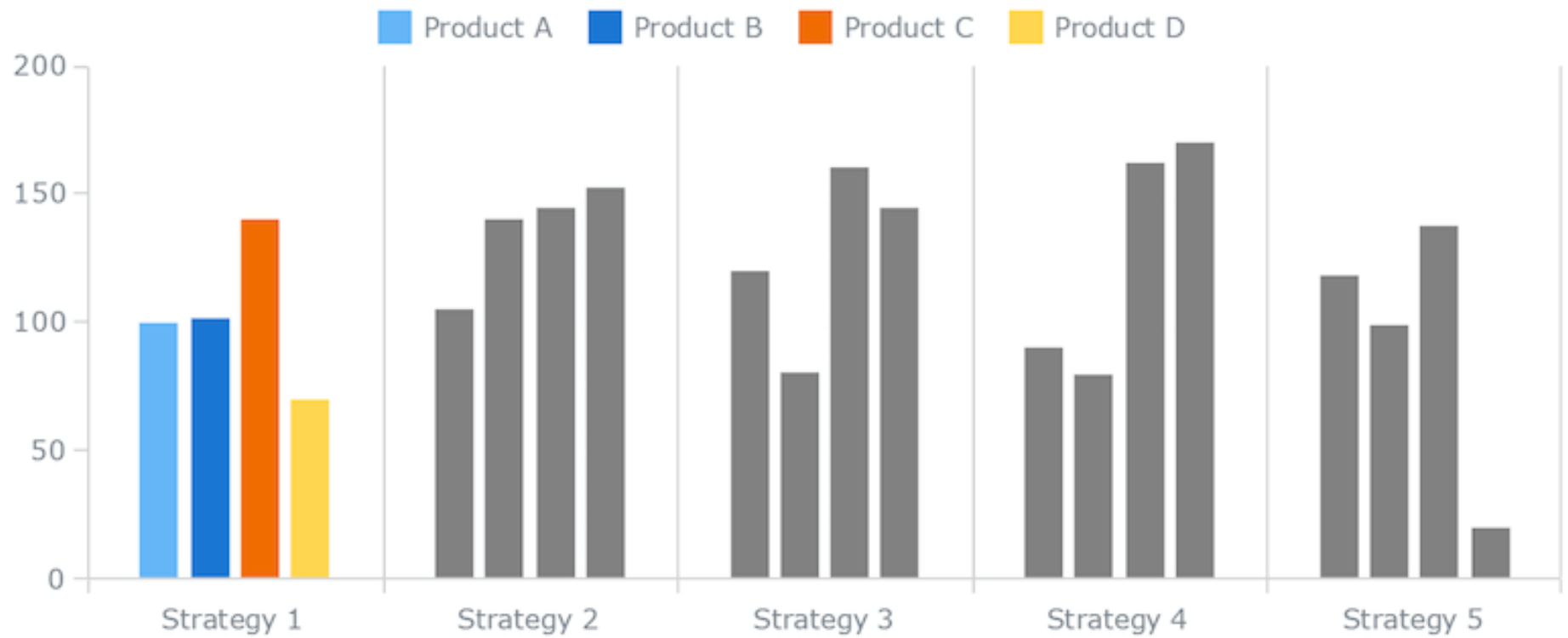
Compare sales strategy



Compare sales strategy



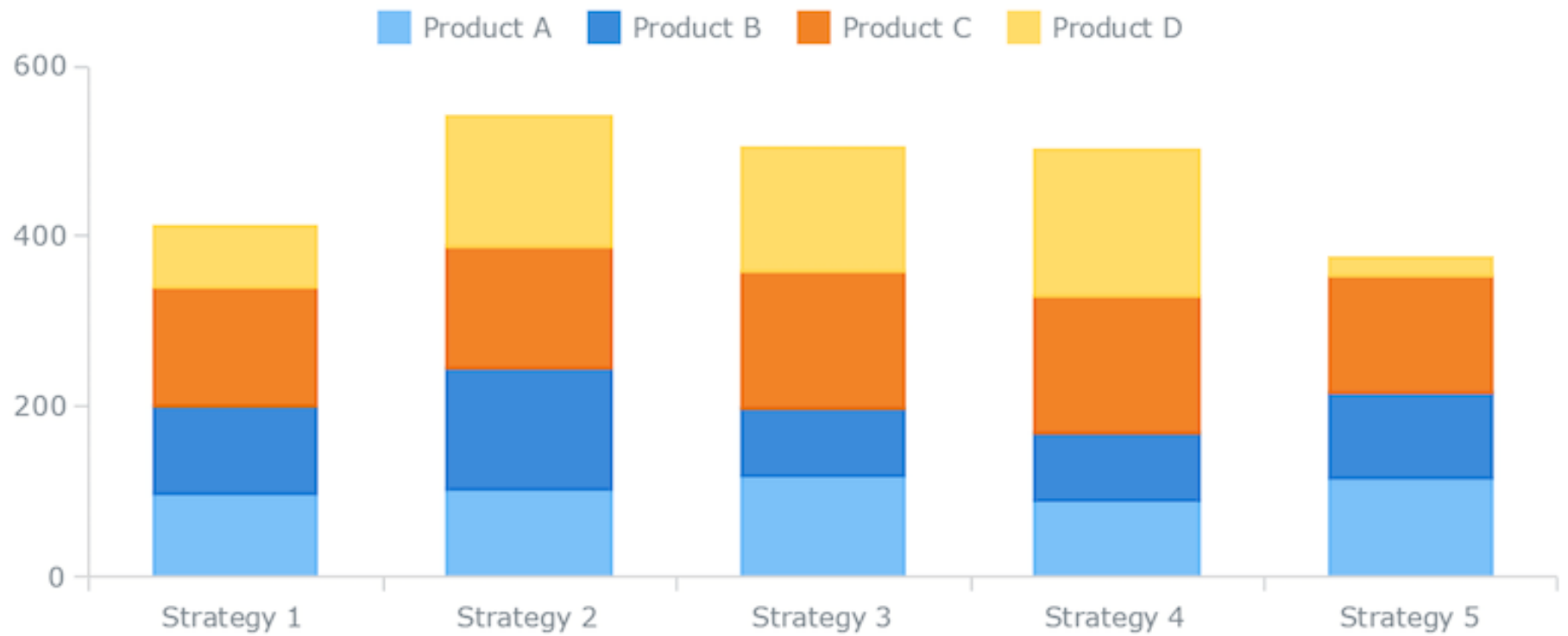
Compare sales strategy



Stacked Bars

- Instead of drawing groups of bars side-by-side, it is sometimes preferable to *stack bars* on top of each other.
- Stacking is useful *when the sum of the amounts represented by the individual stacked bars is in itself a meaningful amount.*
- Stacked bar charts are designed to help you simultaneously compare totals and notice sharp changes at the item level that are likely to have the most influence on movements in category totals.

Compare sales strategy



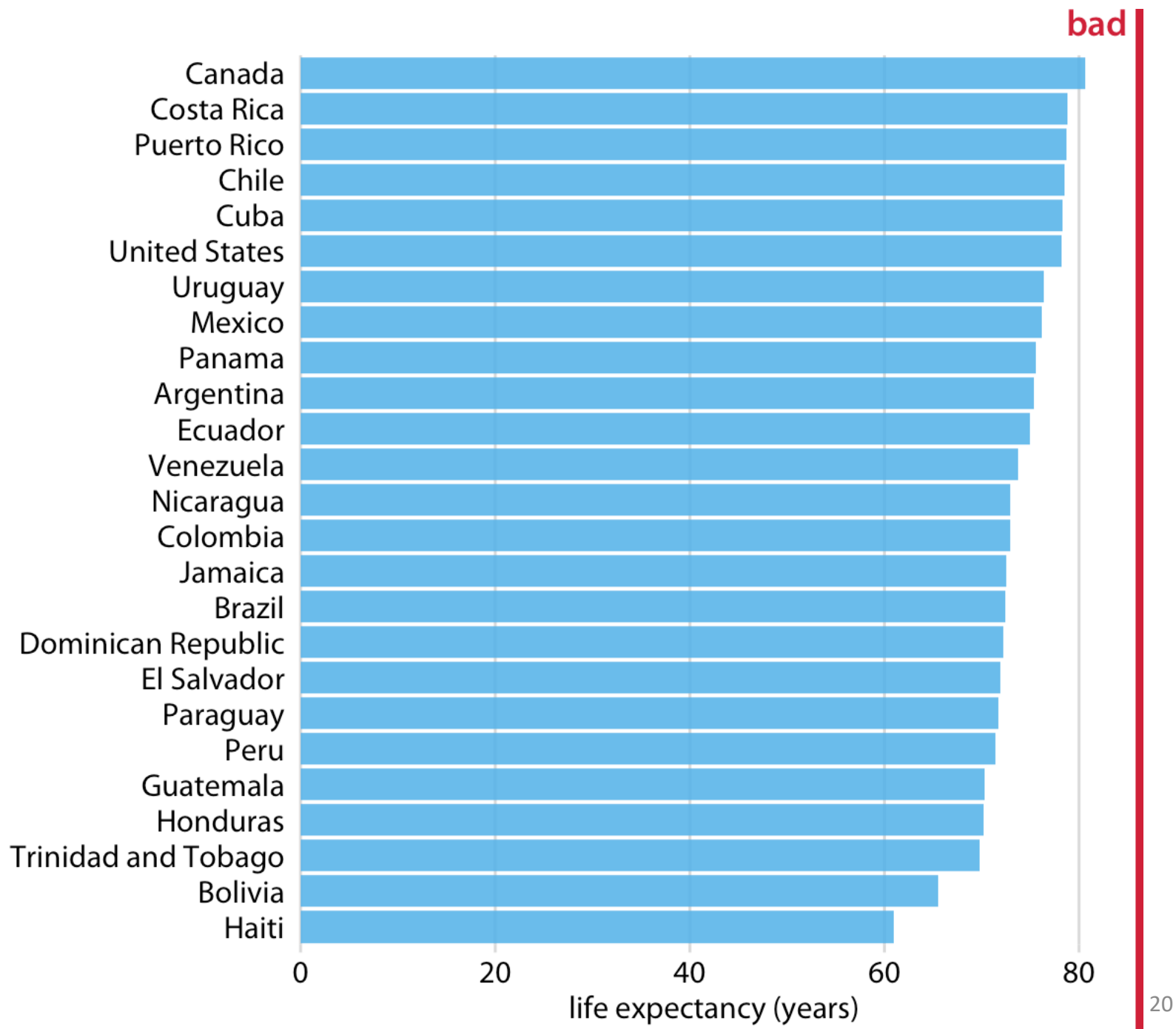
Compare sales strategy

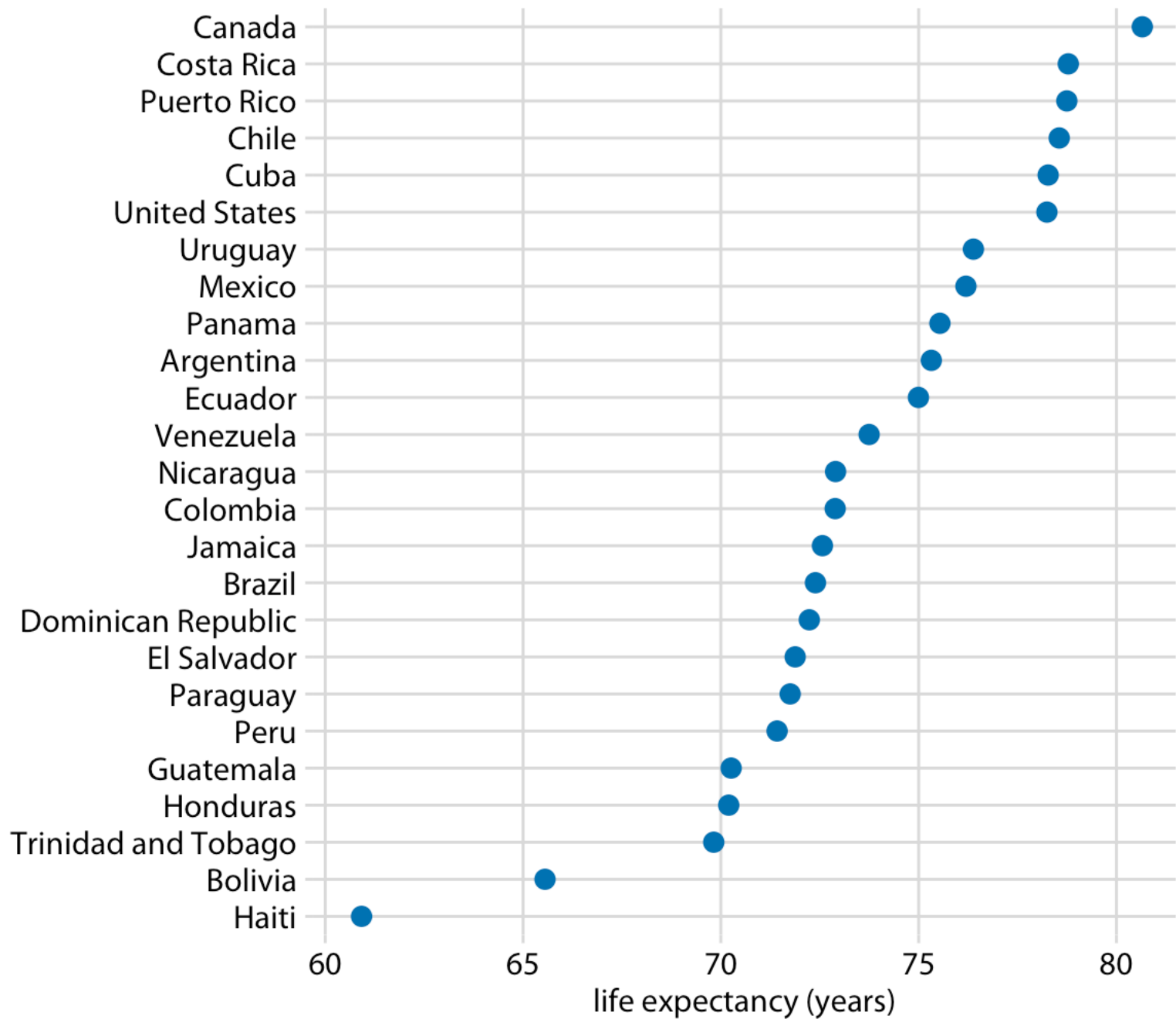
Product A Product B Product C Product D

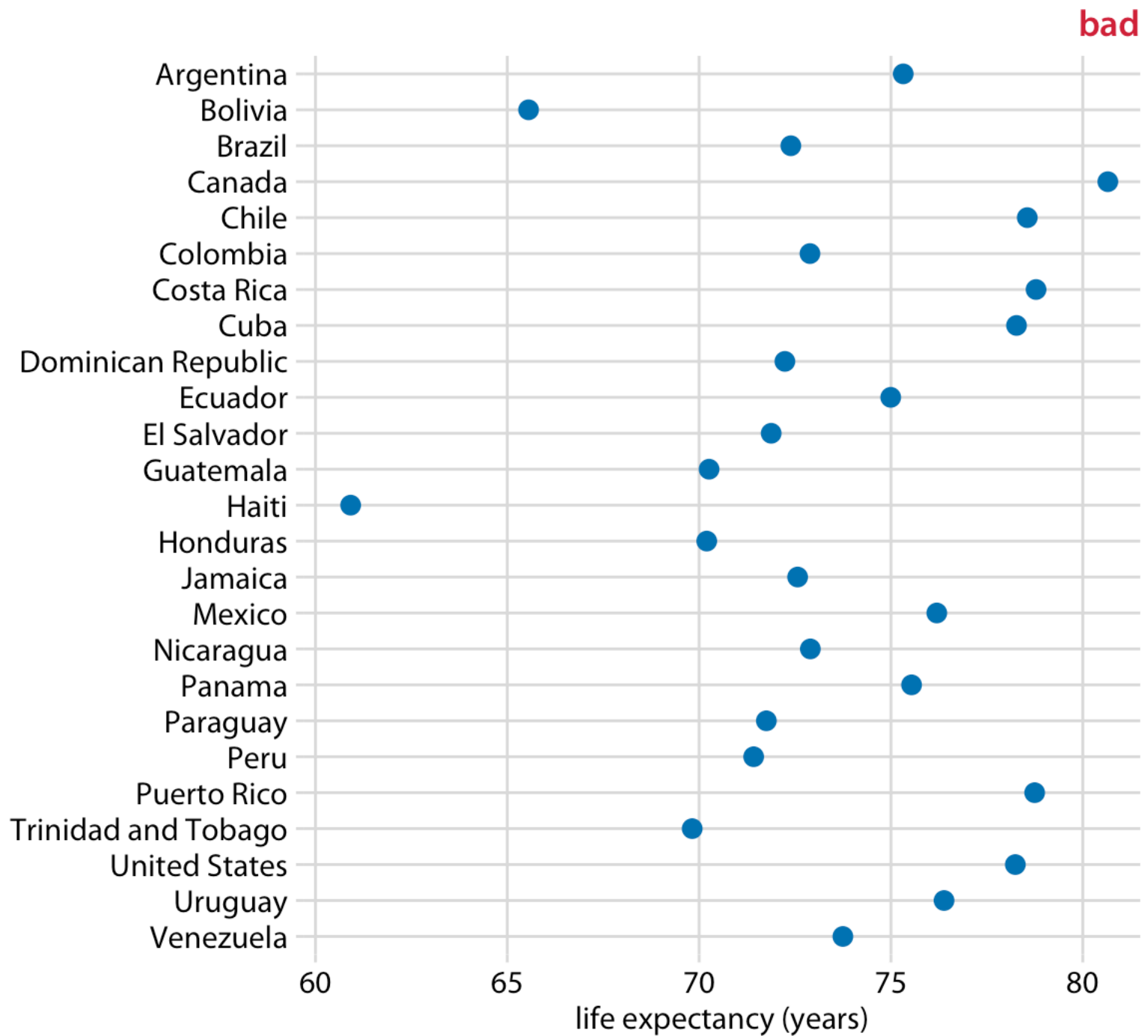


Dot plots and heatmaps

- Bars are not the only option for visualizing amounts.
- One important limitation of bars is that they need to start at zero, so that the bar length is proportional to the amount shown.
- In this case, we can indicate amounts by placing dots at the appropriate locations along the x or y axis.

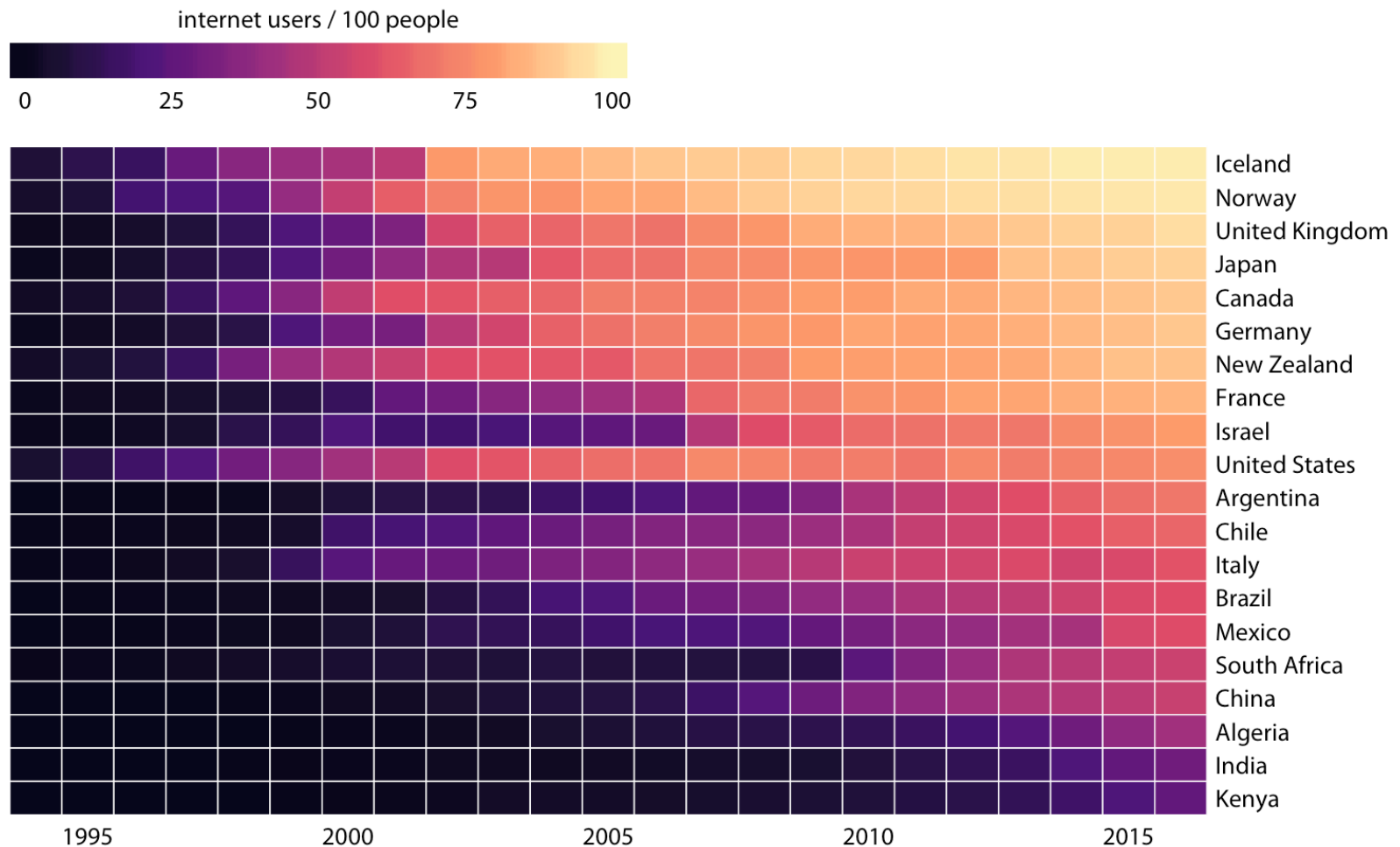




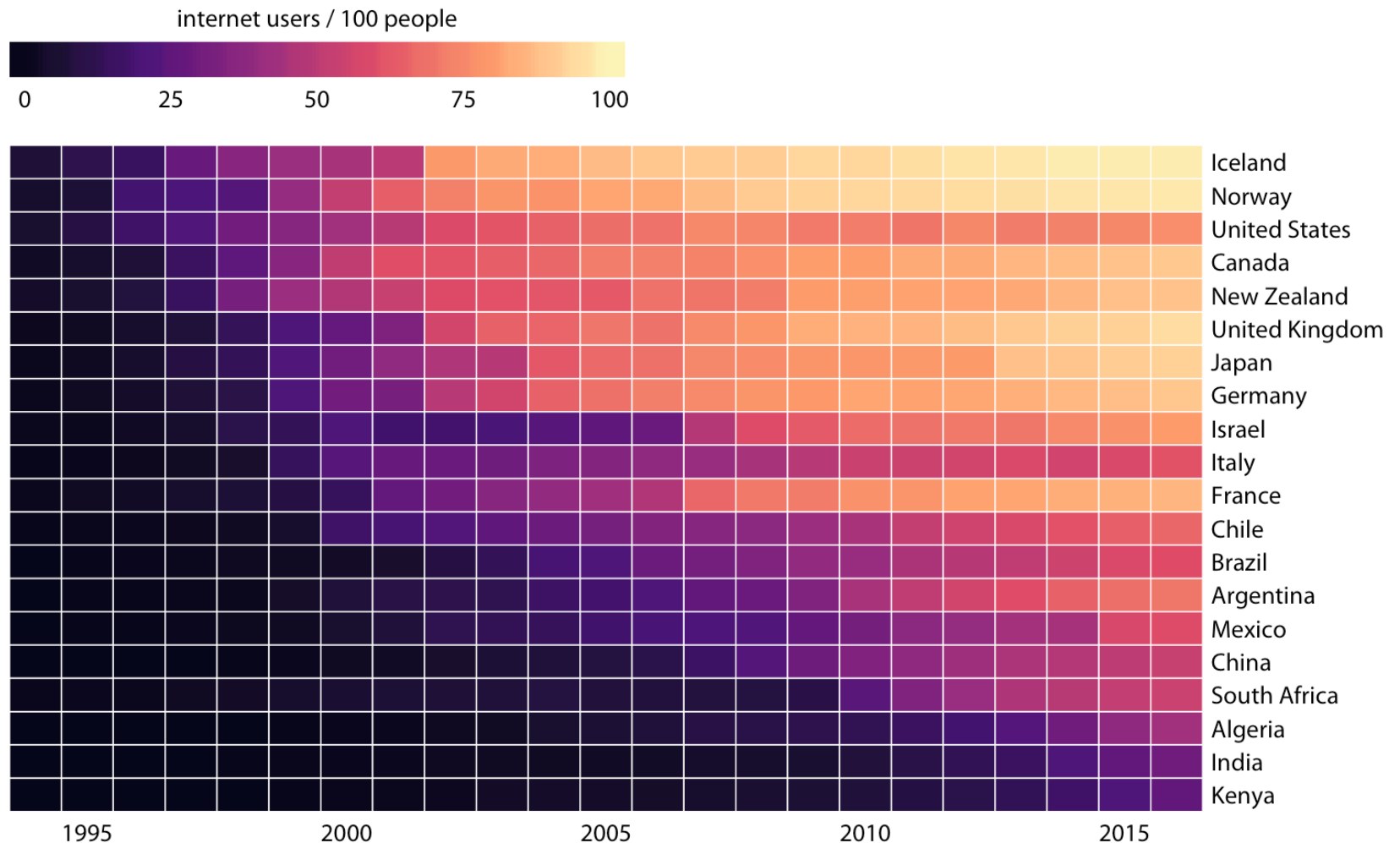


Heatmap

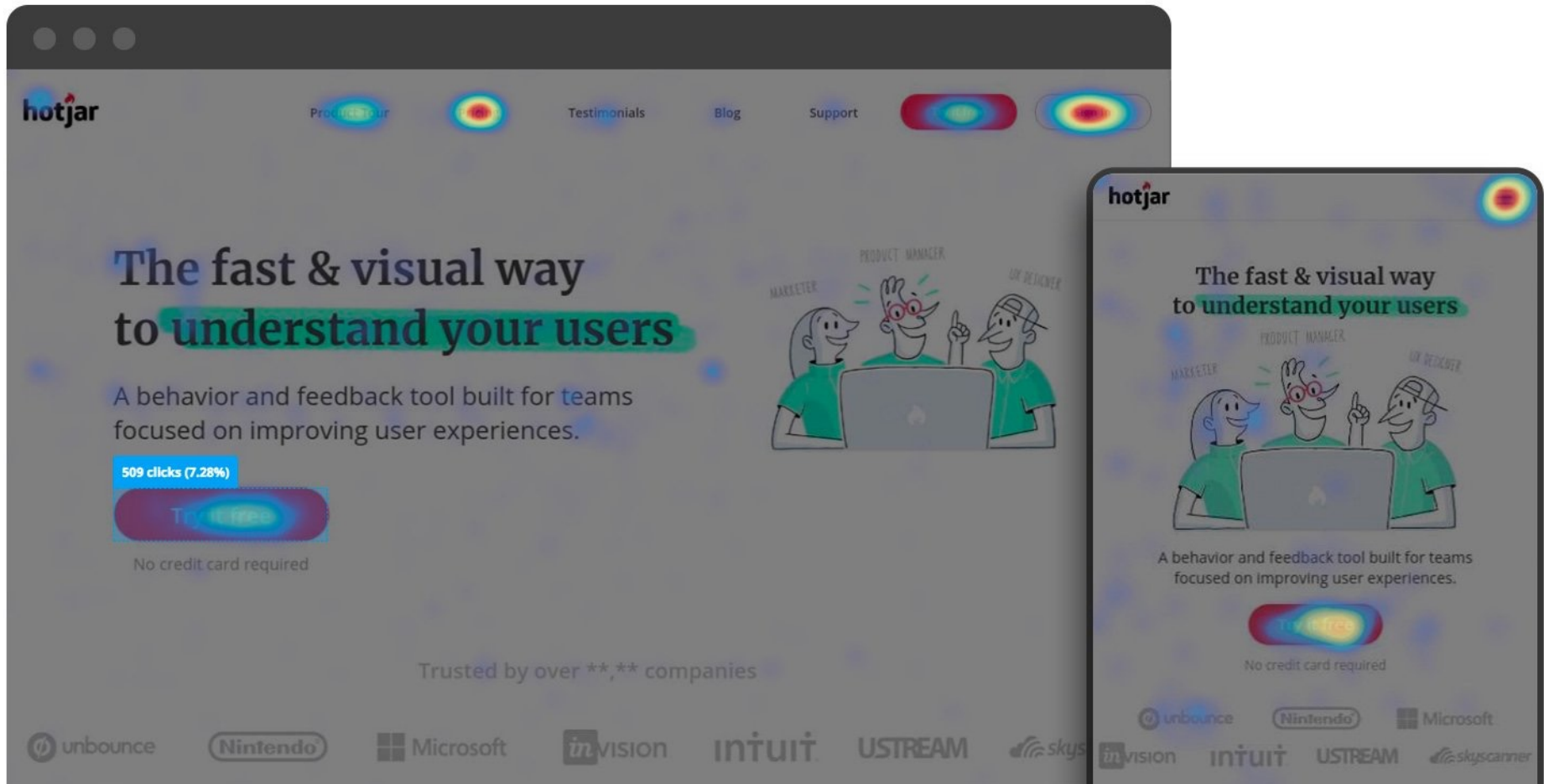
- As an alternative to mapping data values onto positions via bars or dots, we can map data values onto colors. Such a figure is called a *heatmap*.
- Heat maps make it easy to visualize complex data and understand it at a glance.



Internet adoption over time, for select countries. Color represents the percent of internet users for the respective country and year. Countries were ordered by percent internet users in 2016. Data source: World Bank



Internet adoption over time, for select countries. Countries were ordered by the year in which their internet usage first exceeded 20%. Data source: World Bank



A Click map of user clicks on web vs mobile app

Complete your visualization: Illustrate the figure with Title, Annotations and Caption

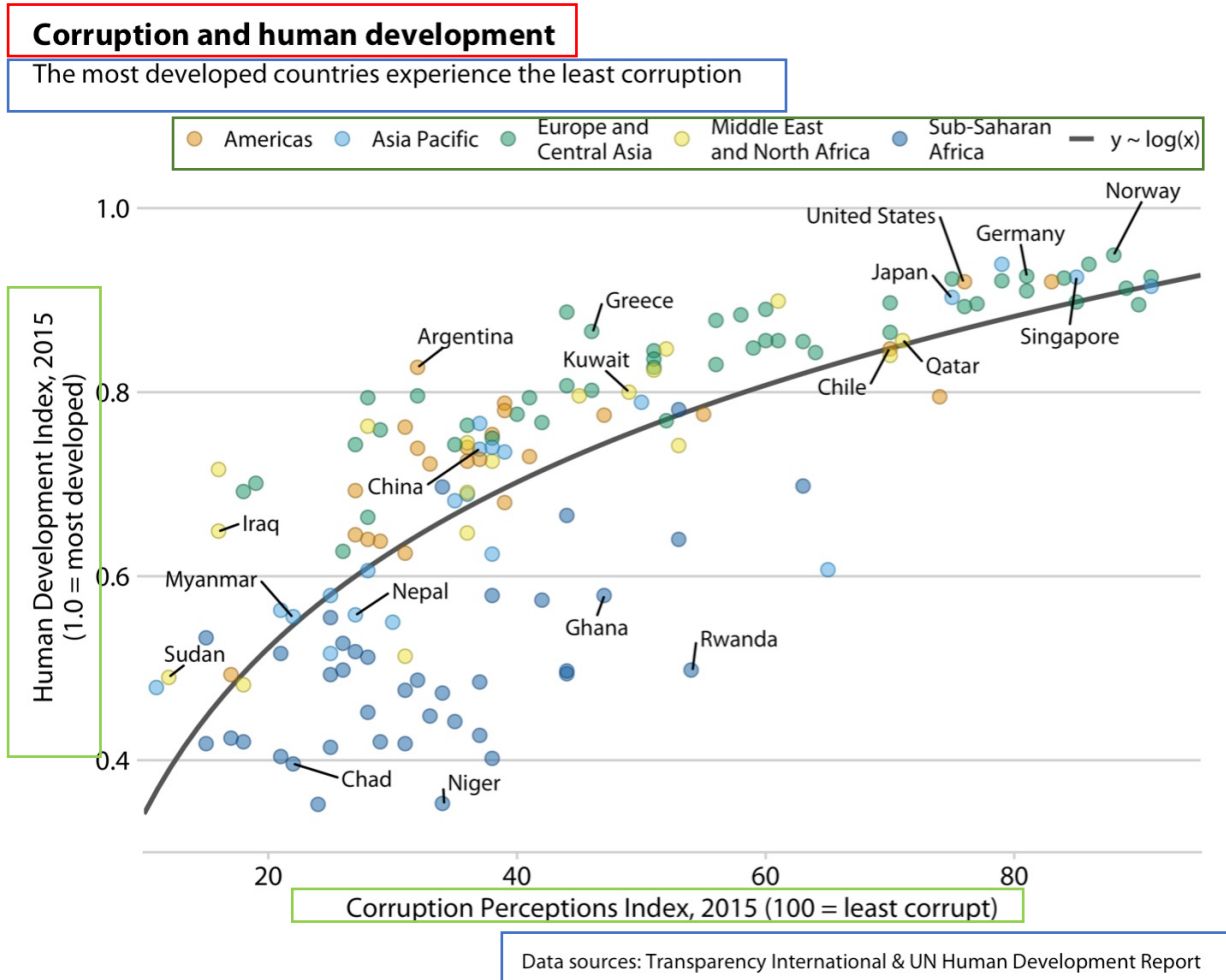


Figure 22.2: Infographic version of Figure 22.1. The title, subtitle, and data source statements have been incorporated into the figure. This figure could be posted on the web as is or otherwise used without separate caption block.

A complete visualization should be self-explanatory.

- Title

- Every figure needs the title.
 - In the figure: Usually for Standalone Figures
 - Integrated as the first element of the caption: in the context of conventional book or article publishing
- A figure can have only one title.

- Caption

- A brief description appears next to the image and identifies or describes the image, and credits the source.

- Annotations

- **Axis labels** and **legend titles** explain what the displayed data values are and how they map to plot aesthetics.
- More: Subtitle, Data Source Information, etc.

- Figure title in the Caption

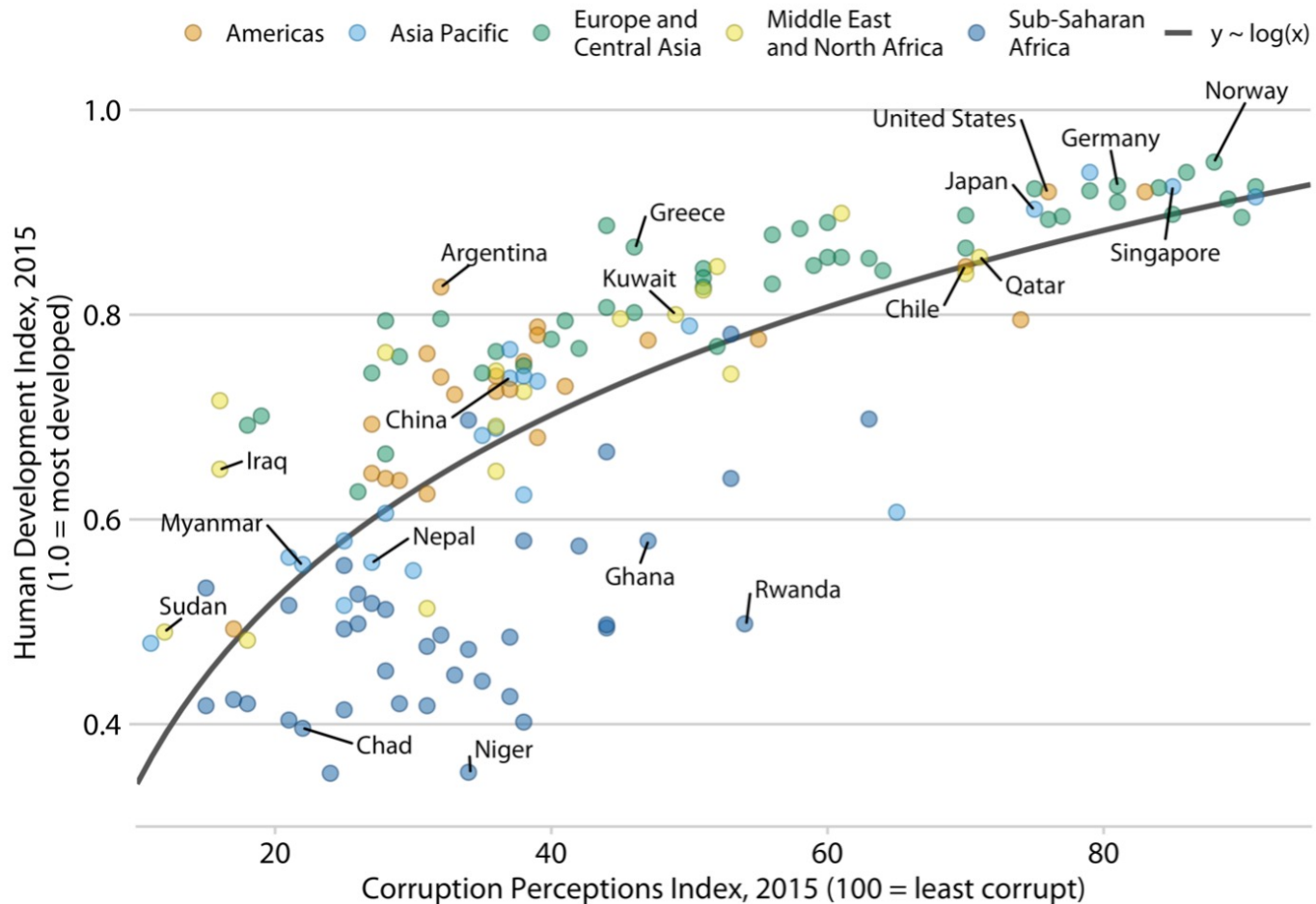


Figure 22.1: Corruption and human development: The most developed countries experience the least corruption. This figure was inspired by a posting in The Economist online (2011). Data sources: Transparency International & UN Human Development Report

- Omitting axis and legend titles
 - when axis or legend titles can be omitted, namely when the labels themselves are fully explanatory.
 - be careful when omitting axis or legend titles, because it is easy to misjudge what is and isn't obvious from the context.

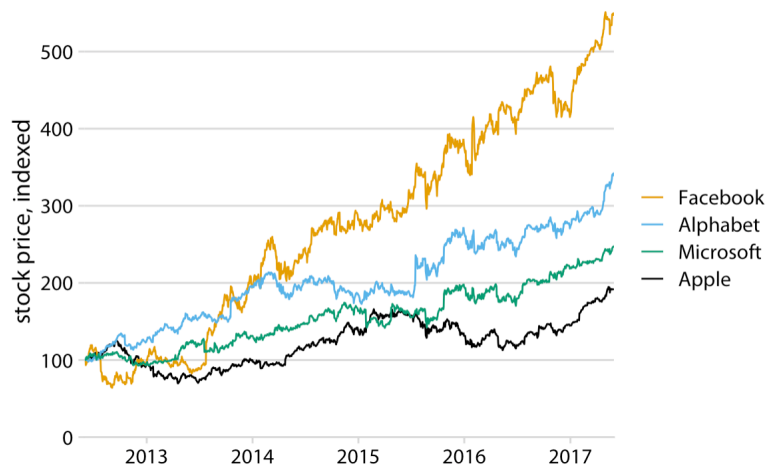


Figure: Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012.

