



# *Visualizing Comparison of Data*

---

DR. MUNEENDRA OJHA

ASSISTANT PROFESSOR

DEPARTMENT OF INFORMATION TECHNOLOGY

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD



# *Comparing Categories*

# *How to compare?*

- Bars, lines, and dots can all let our readers compare within and between groups.
- In some cases, we want our reader to see both levels and change, or some other variable combination; in other cases, we want to focus their attention on one comparison or another.
- The challenge when comparing categorical data is deciding what we want the chart to convey
- We need to prioritize what we want our charts to do.
- By putting every bar or dot in the graph, we can obscure the point we wish to convey.

# *Bar charts*

- The rectangles can be arranged along the vertical axis - called a **bar chart** or on the horizontal axis - called a column chart.
- The length or height of the rectangular bars in bar and column charts depict the value of your data.
- With rectangles sitting on the same straight axis, it's easy to compare the values quickly and accurately.

## The total population in Brazil exceeds that of other countries

(Millions of people)

250

200

150

100

50

0

Brazil Ethiopia France Germany Italy Japan Mexico Turkey Russia Vietnam

Source: The World Bank

- The bar chart is a familiar chart that's easy to read and make.
- It sits at the top of the perceptual ranking matrix.

## The total population in Brazil exceeds that of other countries

(Millions of people)

250

200

150

100

50

Brazil Russia Japan Mexico Ethiopia Vietnam Germany Turkey France Italy

Source: The World Bank

- When possible, sort the data in your bar charts.
- This makes it easier for your reader to find the highest and lowest values.

# *Start the axis at zero*

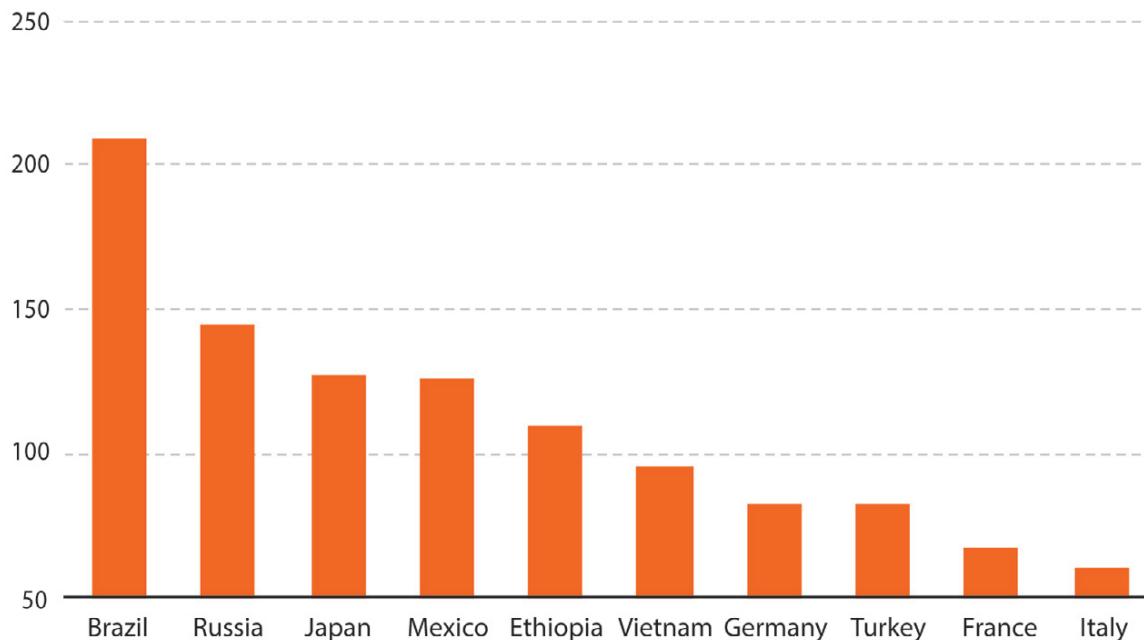
- Starting the axis of bar charts at zero is a rule of thumb upon which many data visualization experts and authors agree.
- Because we perceive the values in the bar chart from the length of the bars, starting the axis at something other than zero may overemphasize the difference between the bars and skew our perception.

# *Start the axis at zero*

- As none are lower than fifty million, we might be tempted to start the axis at fifty million.

The total population in Brazil exceeds that of other countries

(Millions of people)



It looks as though Brazil is orders of magnitude larger than Italy, when, in fact, it is only about three-and-a-half times greater.

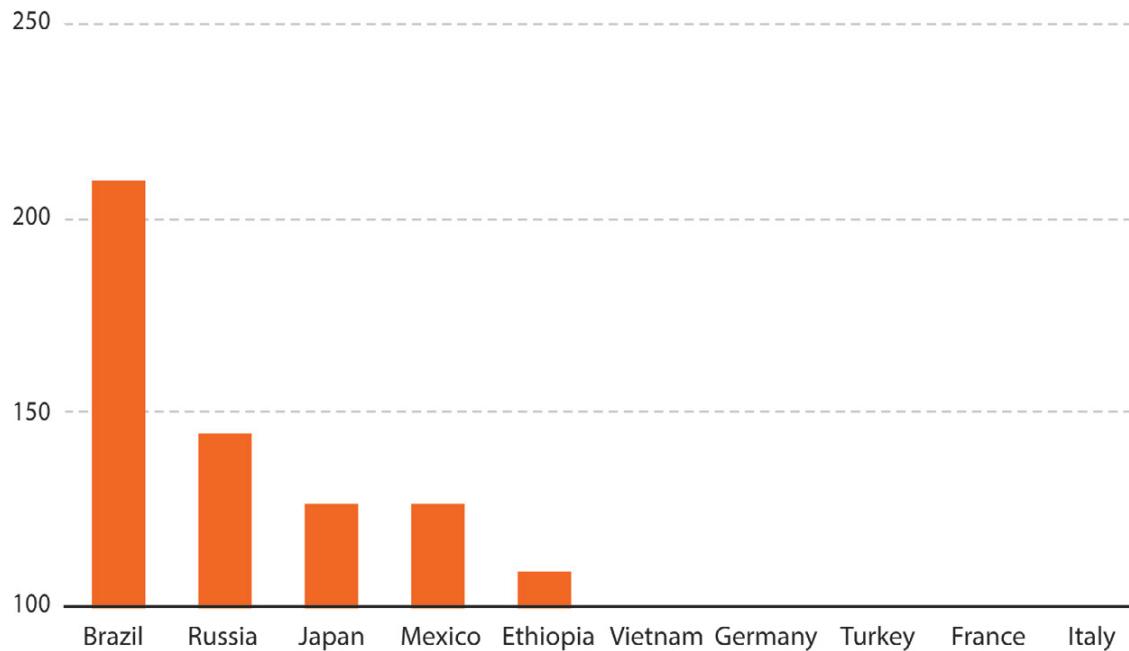
Source: The World Bank

# *Start the axis at zero*

- If you want to take a more extreme view of this, imagine starting the graph at a hundred million.

**The total population in Brazil exceeds that of other countries**

(Millions of people)

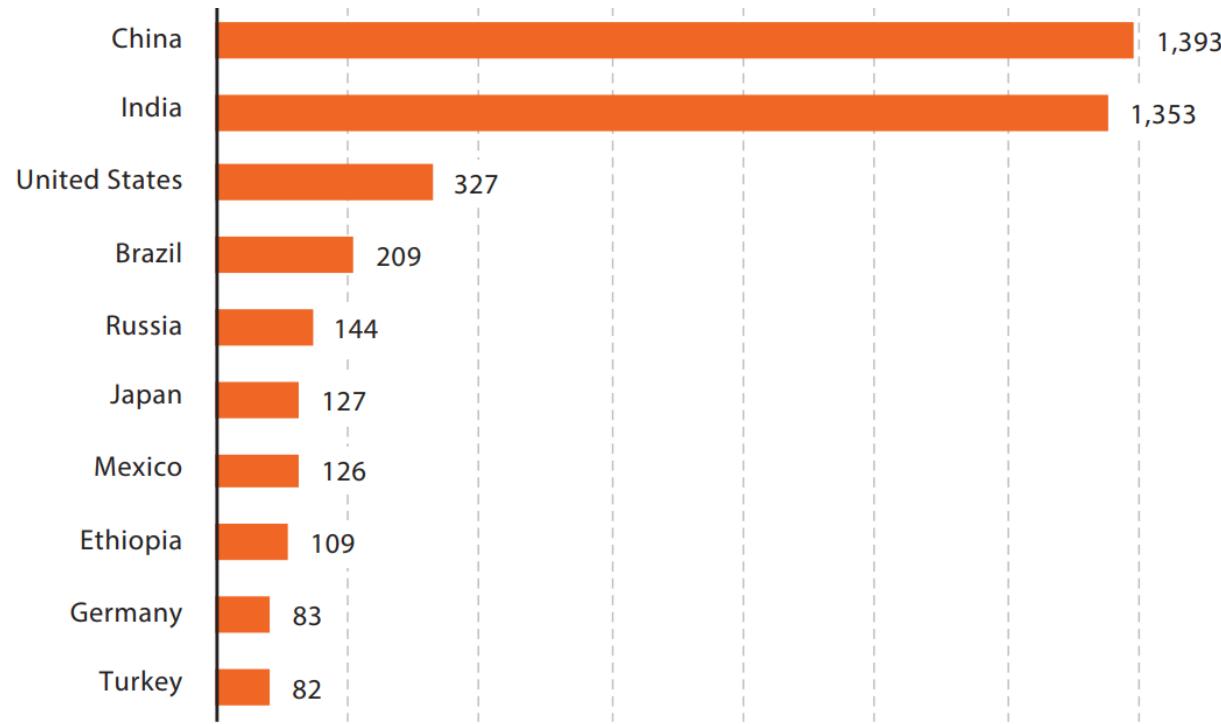


Now at a glance it looks like nobody lives in half of these countries!

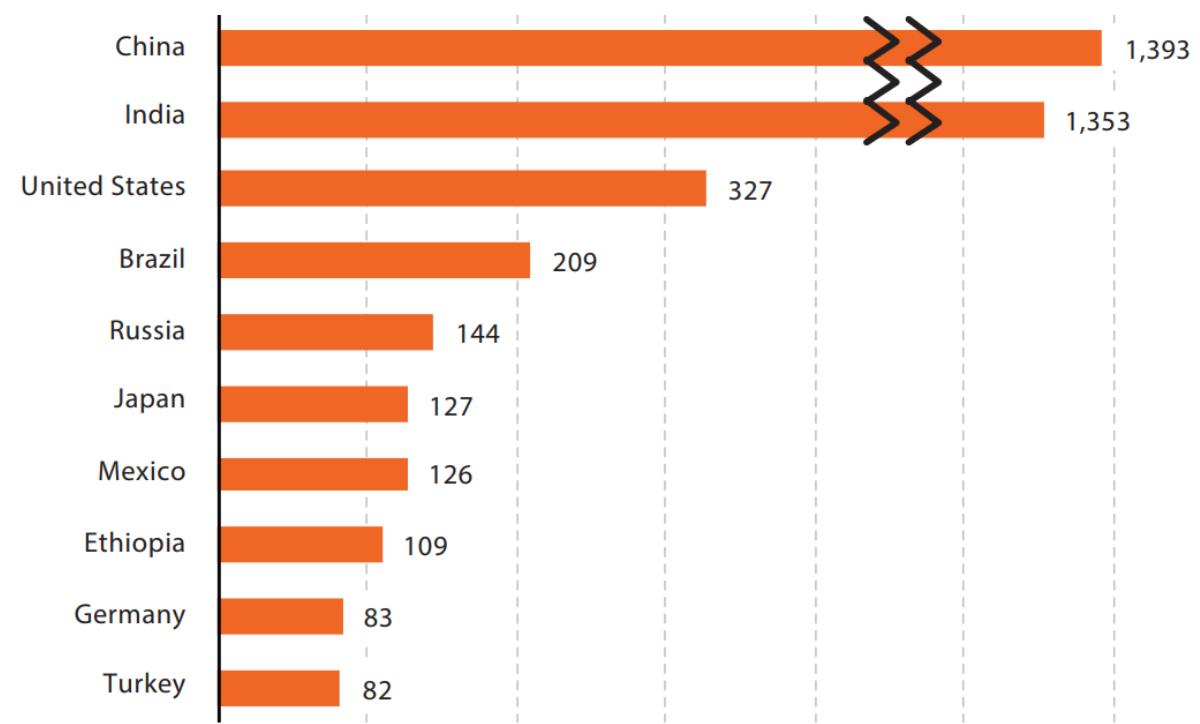
Source: The World Bank

# *Don't break the bar*

- Another cardinal sin of data visualization is what is called “breaking the bar”—that is, using a squiggly line or shape to show that you’ve cropped one or more of the bars.
- It’s tempting to do this when you have an outlier but it distorts the relative values between the bars.

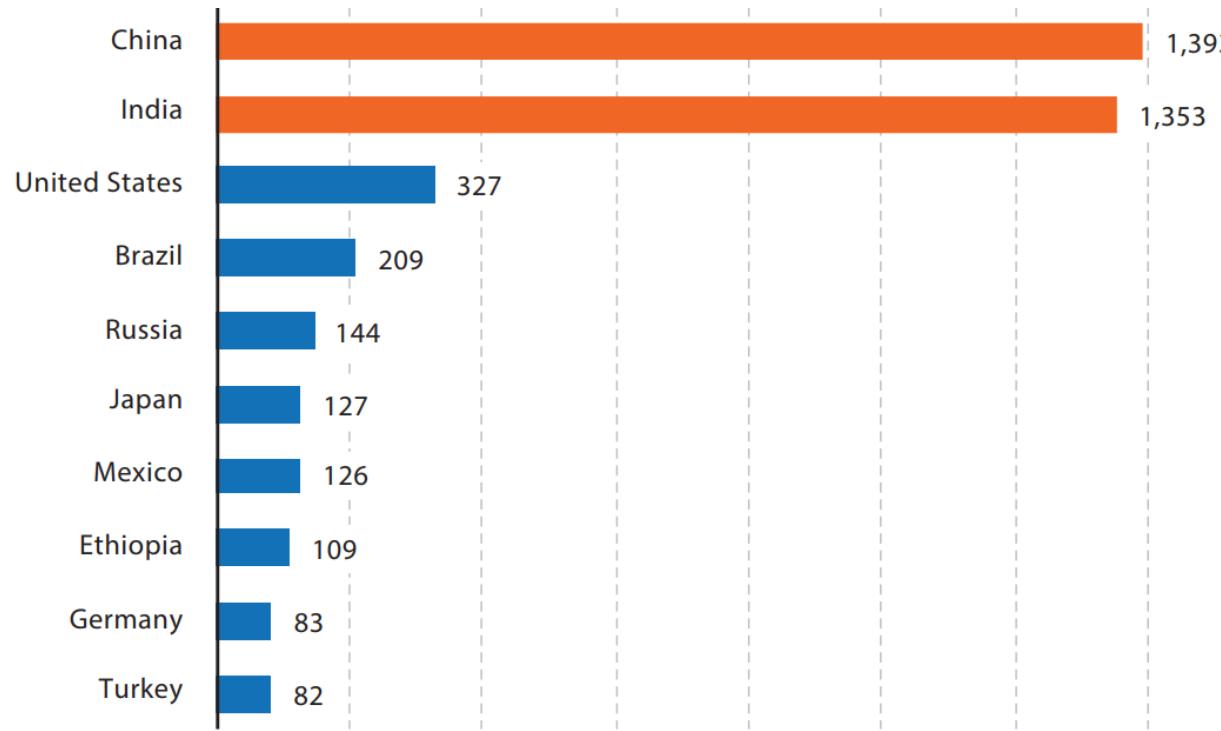


Source: The World Bank

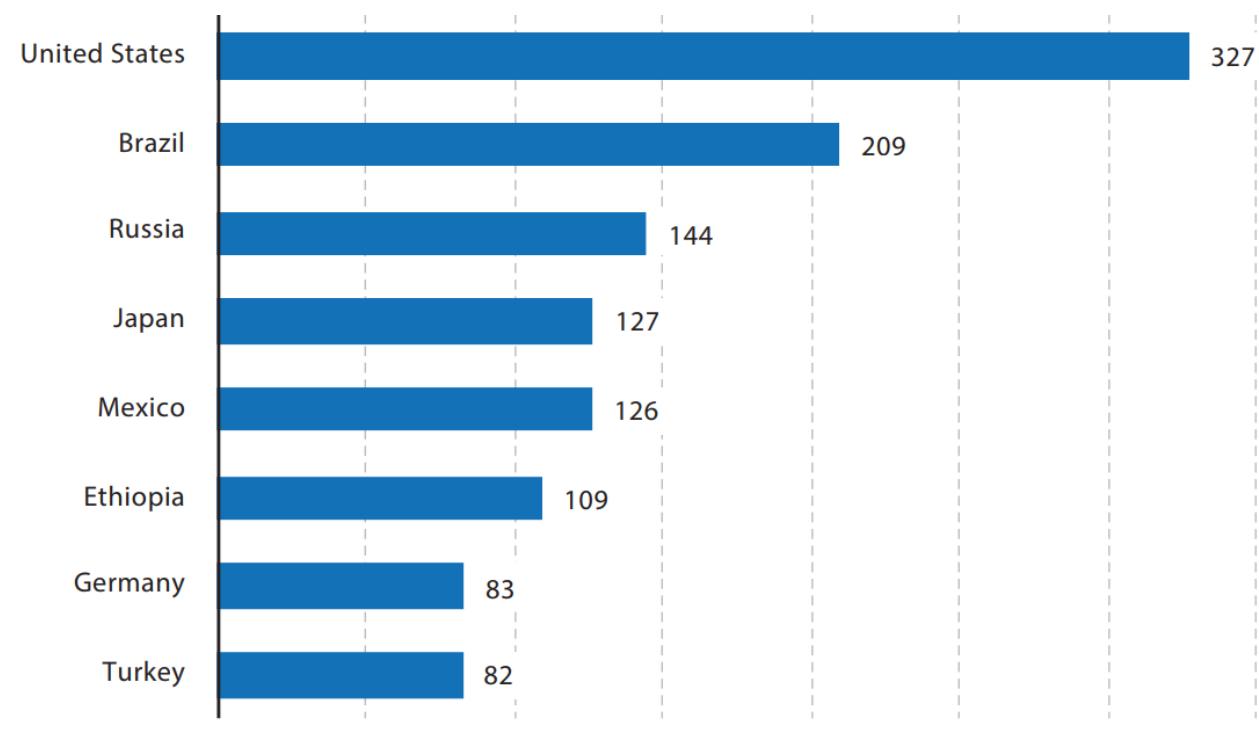


Source: The World Bank

- Don't break the bar in your bar charts.
- The break can be arbitrarily set anywhere and distort our perception of the data.



Source: The World Bank



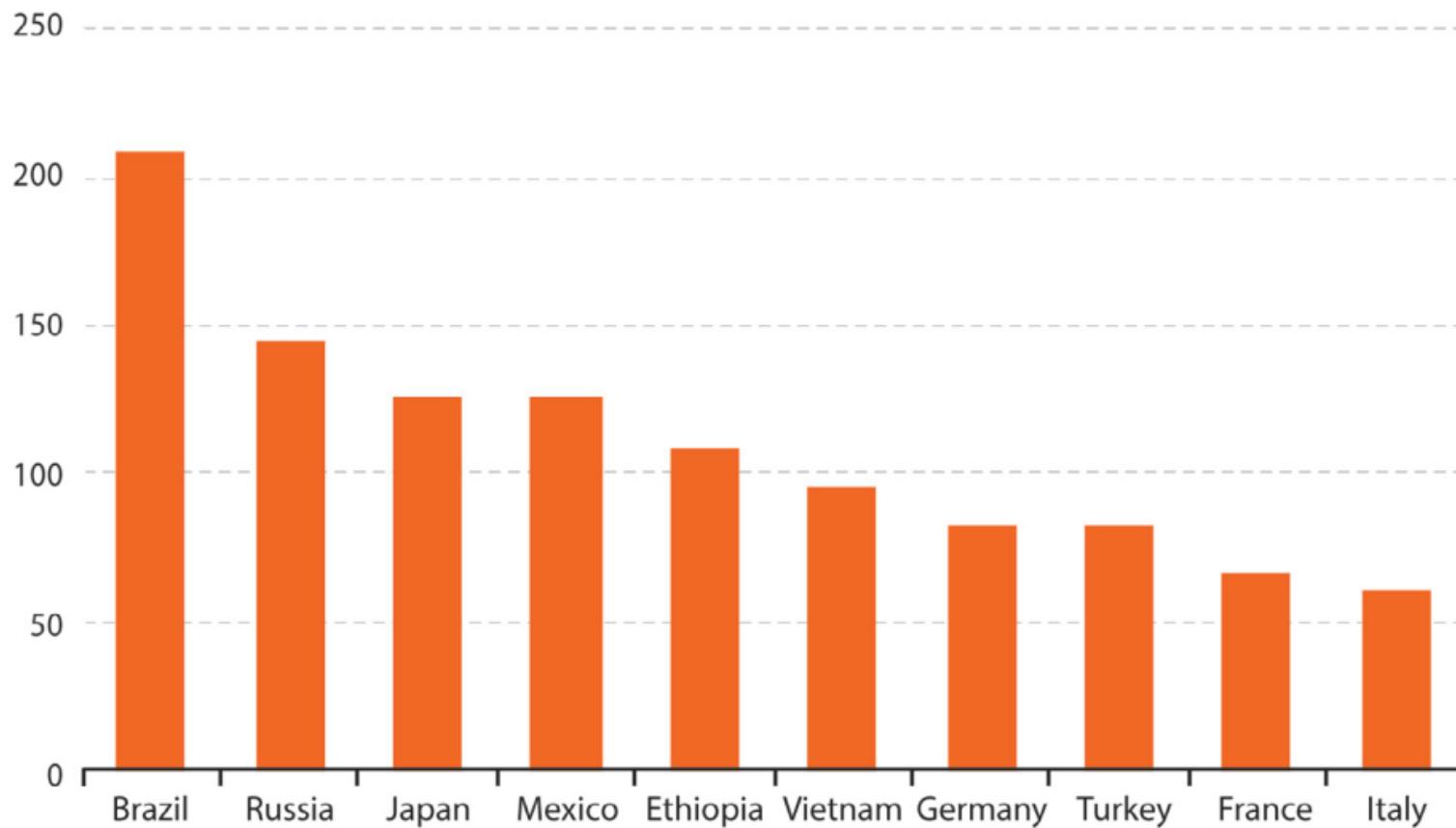
Source: The World Bank

- In cases where you have large values or outliers but want to show the detailed differences between the smaller values, try using more graphs.
- This is a “zoom in” and “zoom out” approach—show all of your data so your reader can see the magnitude of the largest values, and then zoom in for a detailed look that omits the outliers.

# *Use tick marks and gridlines judiciously*

**The total population in Brazil exceeds that of other countries**

(Millions of people)



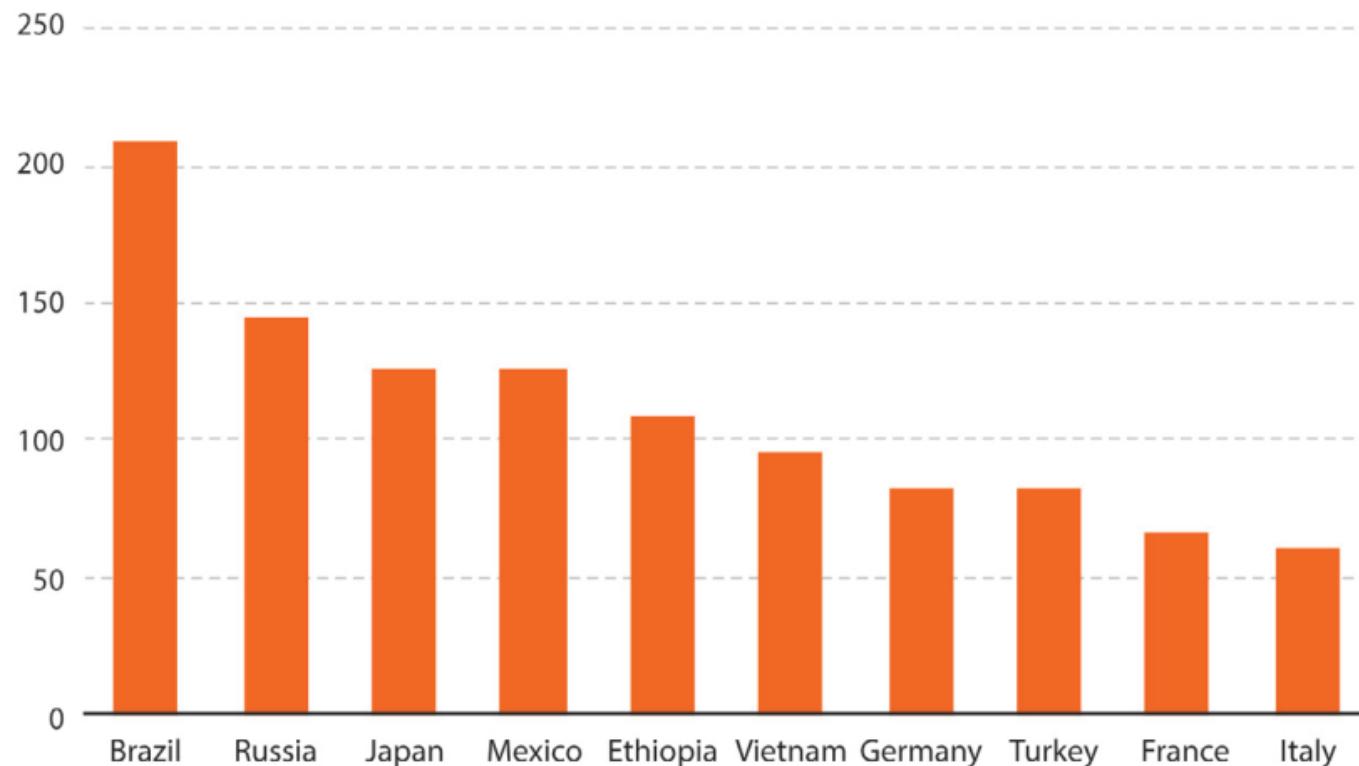
- Bar charts don't need tick marks between the bars.
- White space is an effective separator and deleting the tick marks reduces clutter

Source: The World Bank

# *Use tick marks and gridlines judiciously*

**The total population in Brazil exceeds that of other countries**

(Millions of people)



Source: The World Bank

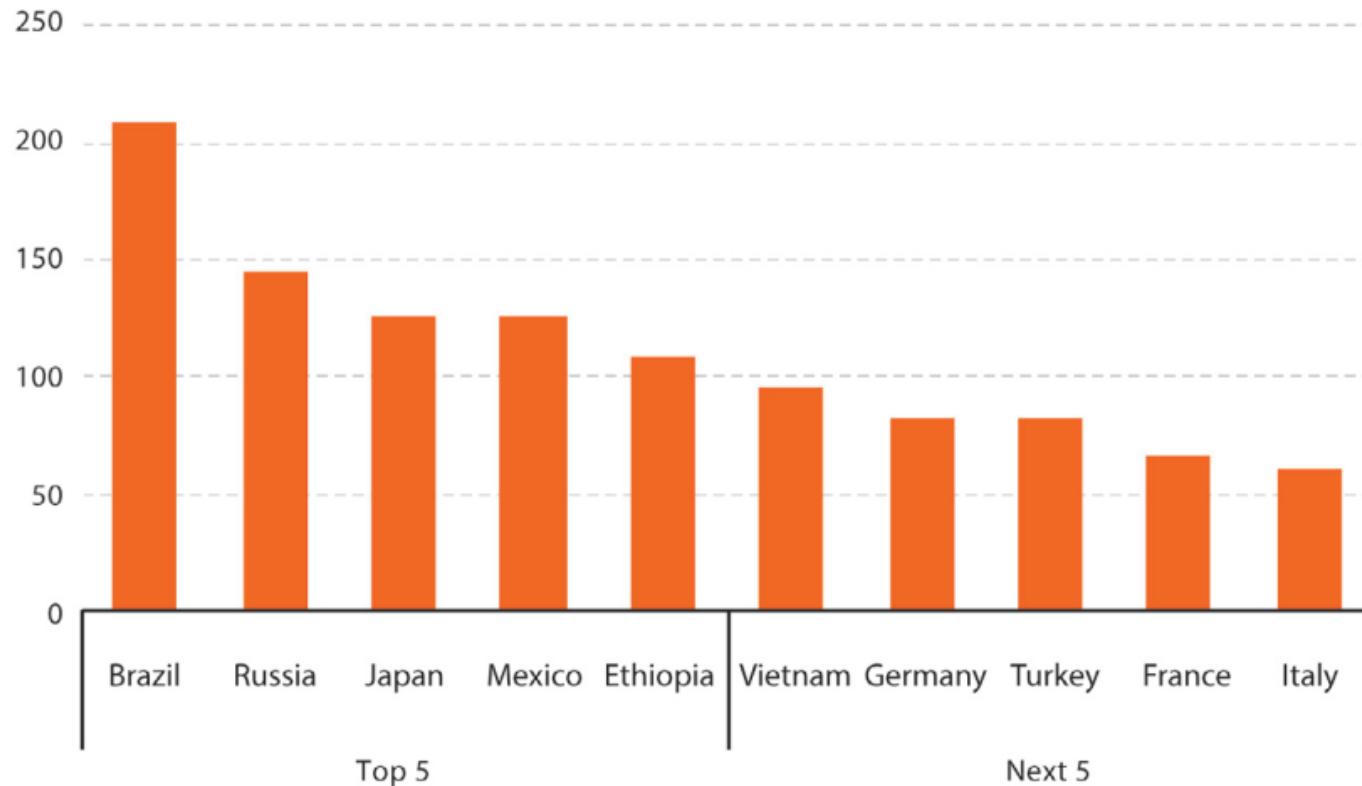
- In bar charts, tick marks are not necessary.
- The white space does the job of separating the bars.

Omitting tick marks is part of removing as many non-data elements as possible.

# *Use tick marks and gridlines judiciously*

**The total population in Brazil exceeds that of other countries**

(Millions of people)

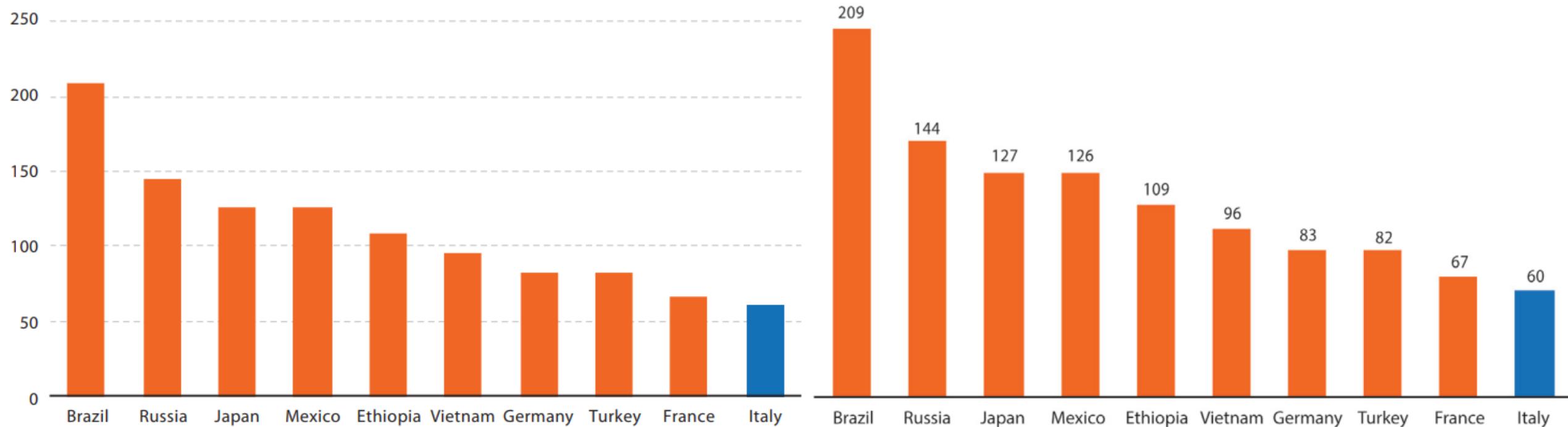


Source: The World Bank

- One exception is if you have a “major” category label that spans multiple bars.
- In such cases, larger tick marks can be helpful to group the labels

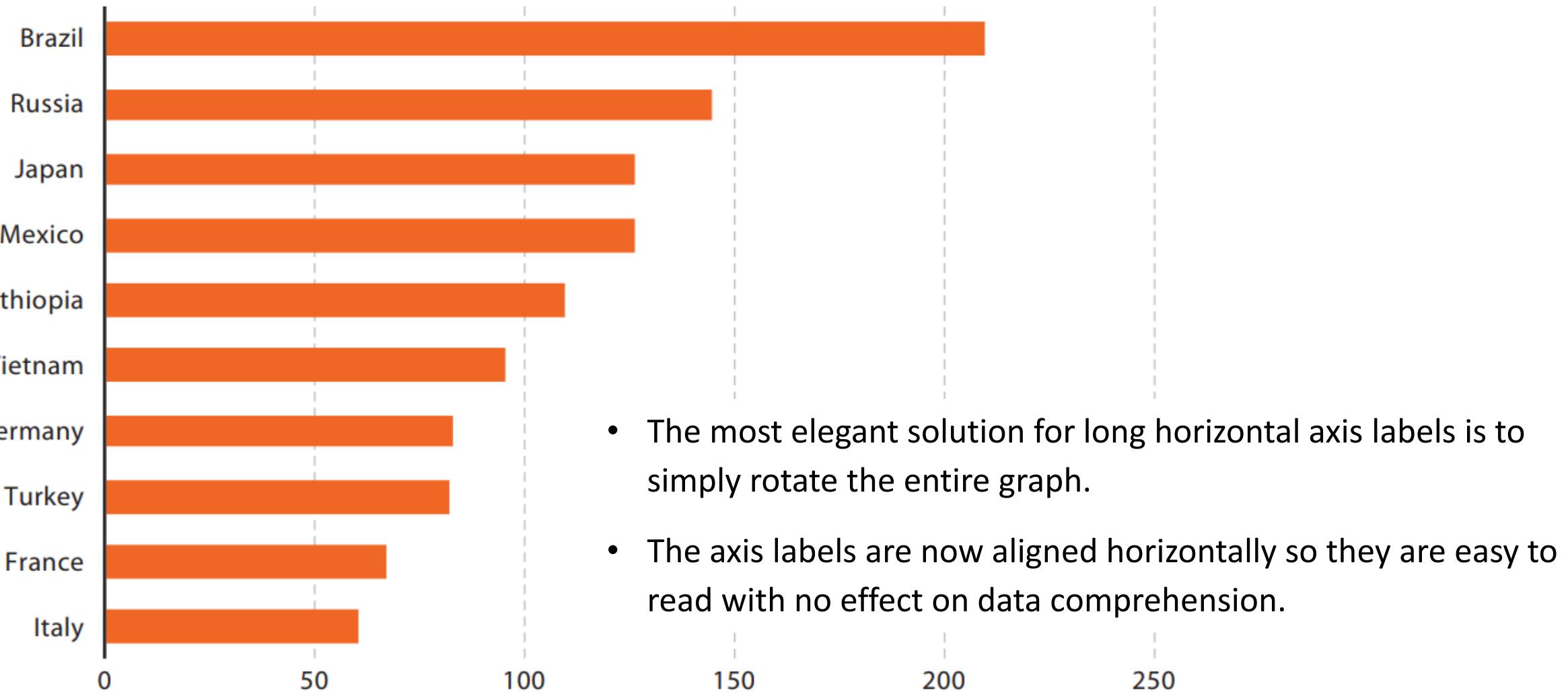
Tick marks may be necessary when you have a “major” category.

Consider Italy in the next two graphs (highlighted in blue).



- Without the labels, the gridline helps us see that there are more than fifty million people living in the country.
- With the label, it is clear that it's sixty million people and thus the gridlines are probably not necessary.

# *Rotate long axis labels*

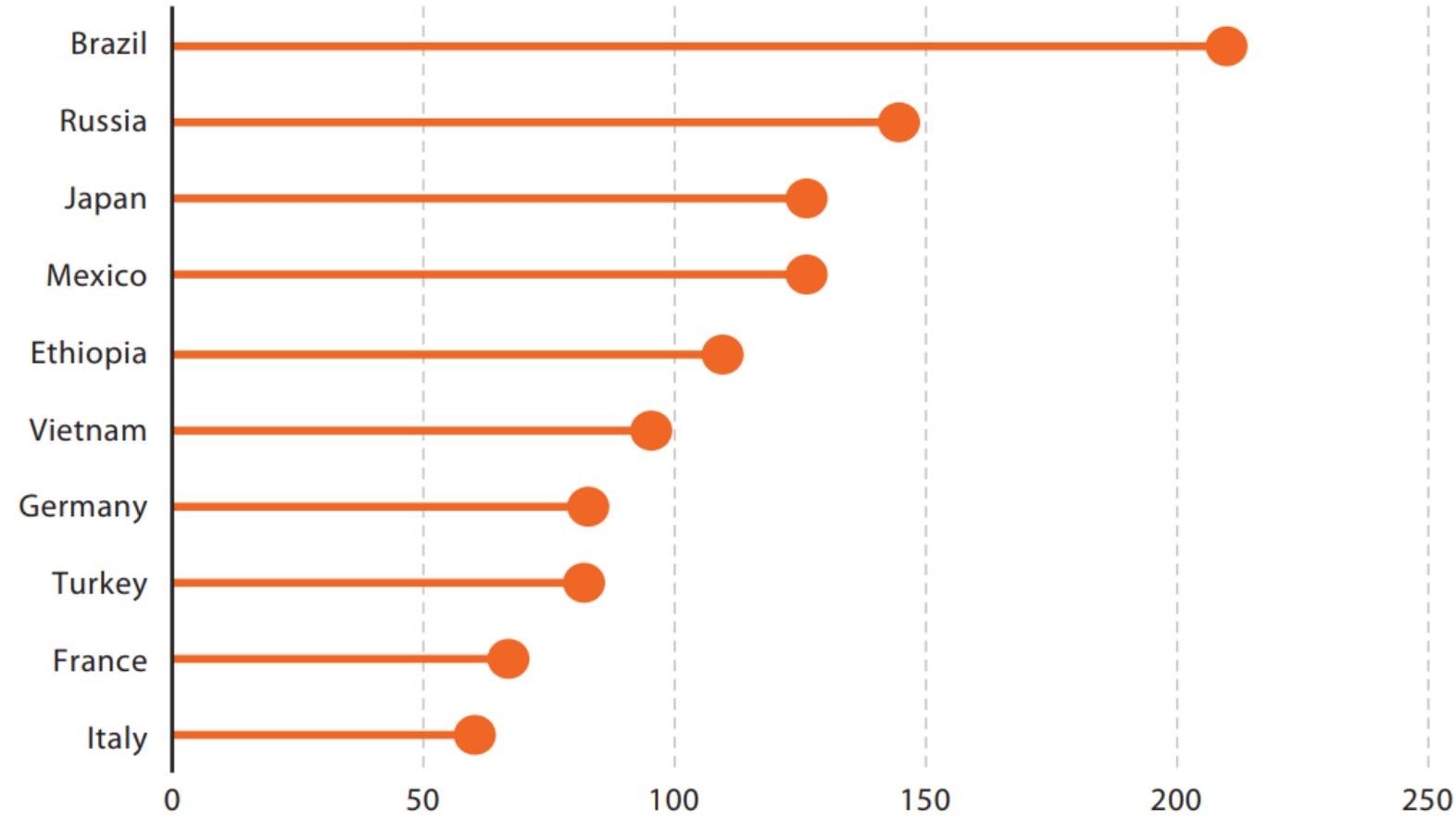


# *Variations on the bar chart*

- There are countless ways to modify the standard bar chart.
- One simple variation is to use other shapes in lieu of bars.

# *Variations on the bar chart*

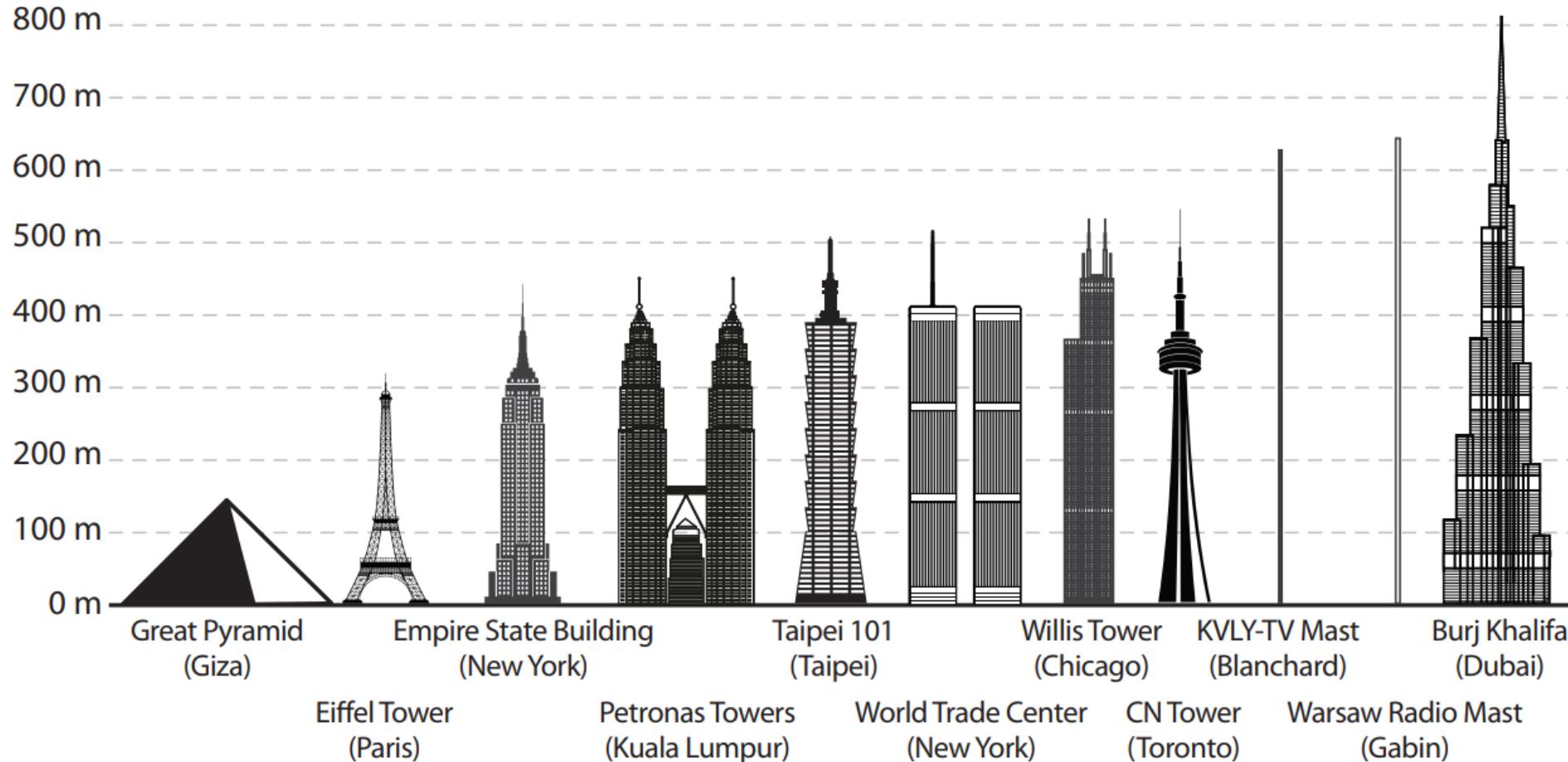
**The total population in Brazil exceeds that of other countries**  
(Millions of people)



The lollipop chart replaces bars with a shape (usually a dot) and a line.

Data Source: The World Bank.

# *Variations on the bar chart*



Alternative shapes, like buildings or people, can be used in lieu of the basic bar shape.

Source: Based on Wikimedia user BurjKhalifaHeight Petronas Towers

# *Bars in a circle*

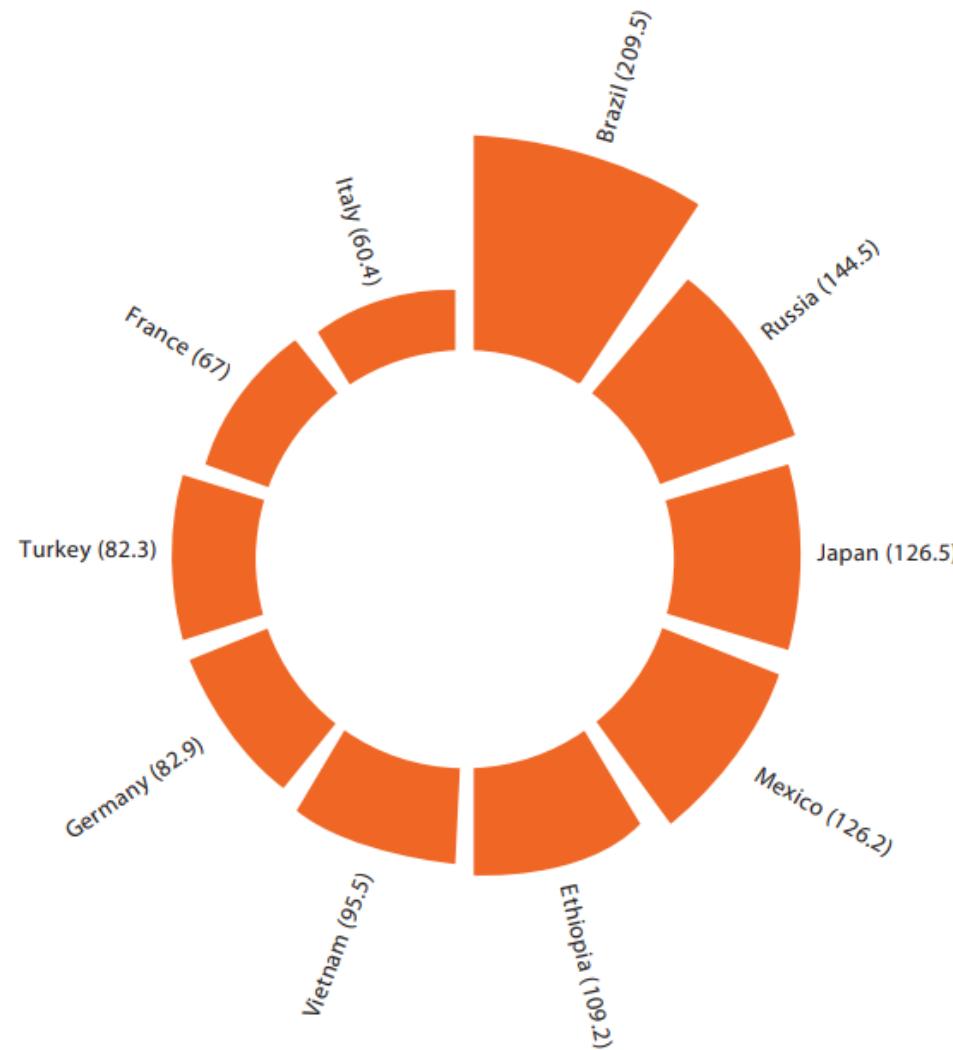
- Another approach to the basic bar chart is to abandon the usual grid and instead place the bars in a circle, called a radial layout.
- There are two common ways to do this:
  - the radial bar chart and
  - the circular bar chart.

# *Radial bar chart*

- Also called the **polar bar chart**, radial bar chart arranges the bars to radiate outward from the center of a circle.
- This graph *lies lower on the perceptual ranking* list because it is *harder to compare* the heights of the bars arranged around a circle than when they are arranged along a single flat axis.
- But this *layout allows you to fit more values in a compact space*, and makes *it well-suited for showing more data, frequent changes* (such as monthly or daily), or changes over a long period of time.

## The total population in Brazil exceeds that of other countries

(Millions of people)



## Change in Brazil's population from 2008 to 2018

(Millions of people)



A radial bar chart wraps the standard bar chart around a circle. This chart type moves down the perceptual ranking list because it is harder to compare the heights of the bars.

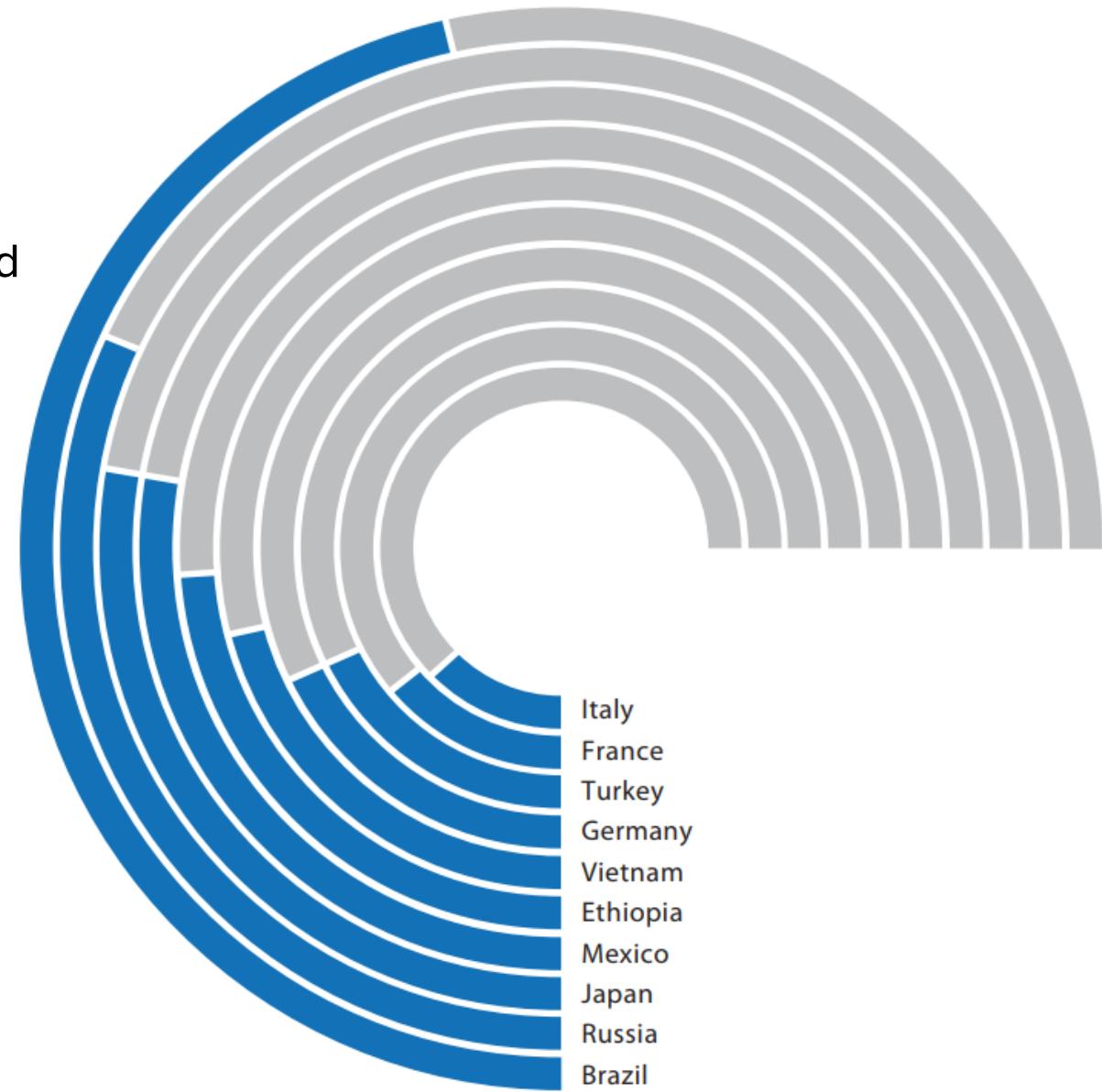
- W. E. B. Du Bois used a circular bar chart in his famous *Exposition des Negres d'Amerique* at the 1900 Paris Exposition.
- It shows the dollar value of household and kitchen furniture held by African Americans in Georgia in six years (1875, 1880, 1885, 1890, 1895, and 1899)
- Perceptually speaking, the circular bar graph is problematic because it distorts our perception of the data.



### Change in Brazil's population from 2008 to 2018

(Millions of people)

- In this case, the lengths of the bars don't correspond to their actual value.
- Like an Olympic footrace, runners start at staggered positions on the track, but they all end up running the same distance because the runner on the outside lane has more distance to cover.
- Here, the visualization doesn't move down the perceptual ranking, but off of it altogether because it distorts the data and for that reason,
- Thus, it is recommended avoiding them altogether.



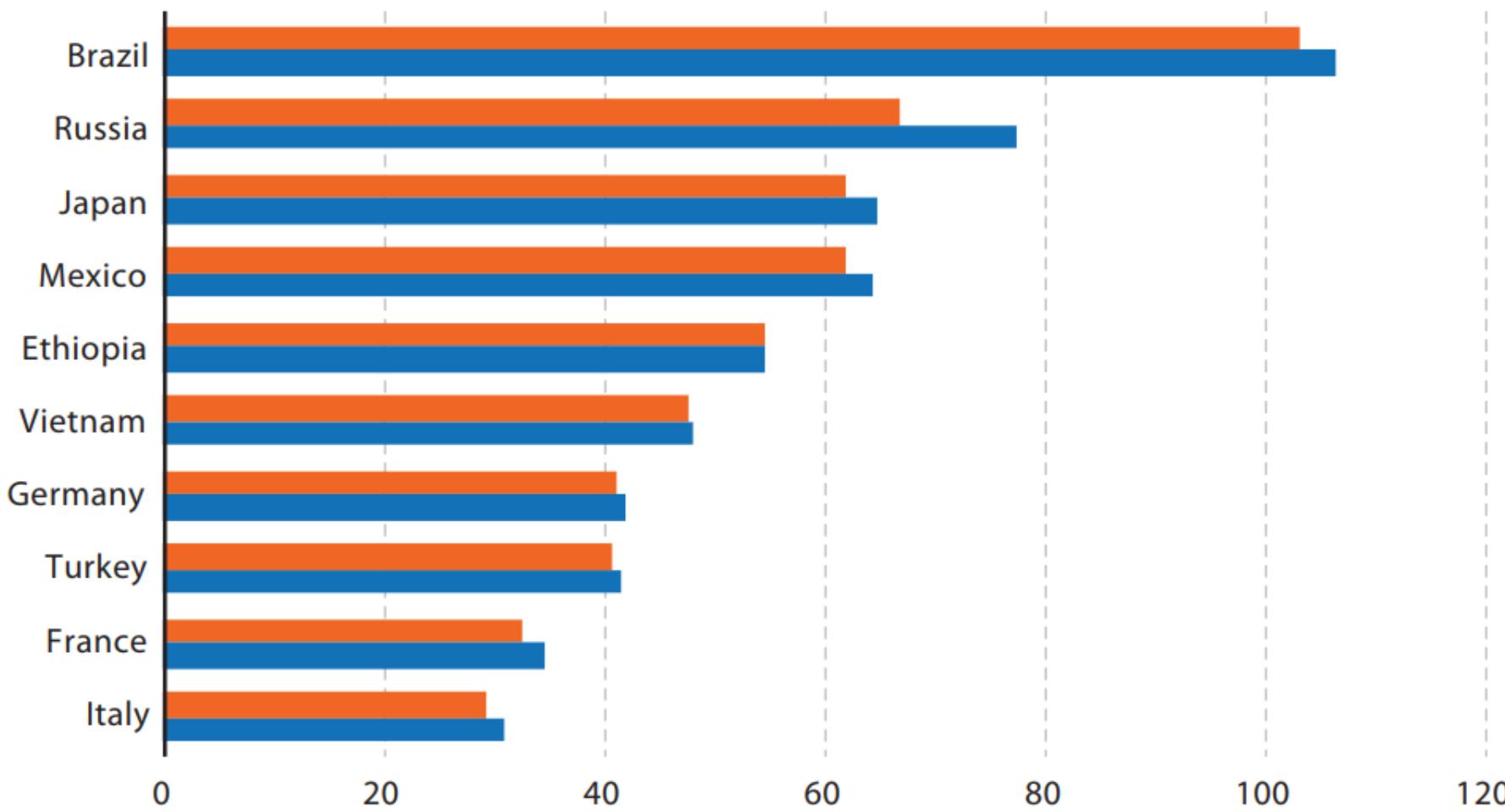
# *Paired bar*

- A simple bar chart is perfect for making comparisons across categories, like comparing populations across countries.
- If one wants to show comparisons not just across but also within countries, the paired bar chart is a good option.
- The paired bar chart will be familiar to most readers and is easy to read, and the shared baseline makes it easy to make comparisons.
- Although called paired, there can be more than one data points in each comparison.

## **There are more women than men in each country except for Ethiopia**

(Millions of people)

Men   Women



A simple paired bar chart is familiar to most readers and easy to read.

Data Source: The World Bank.

## Change in population from 2014 to 2018

(Millions of people)

■ 2014 ■ 2015 ■ 2016 ■ 2017 ■ 2018

250

200

150

100

50

0

Brazil

Russia

Japan

Mexico

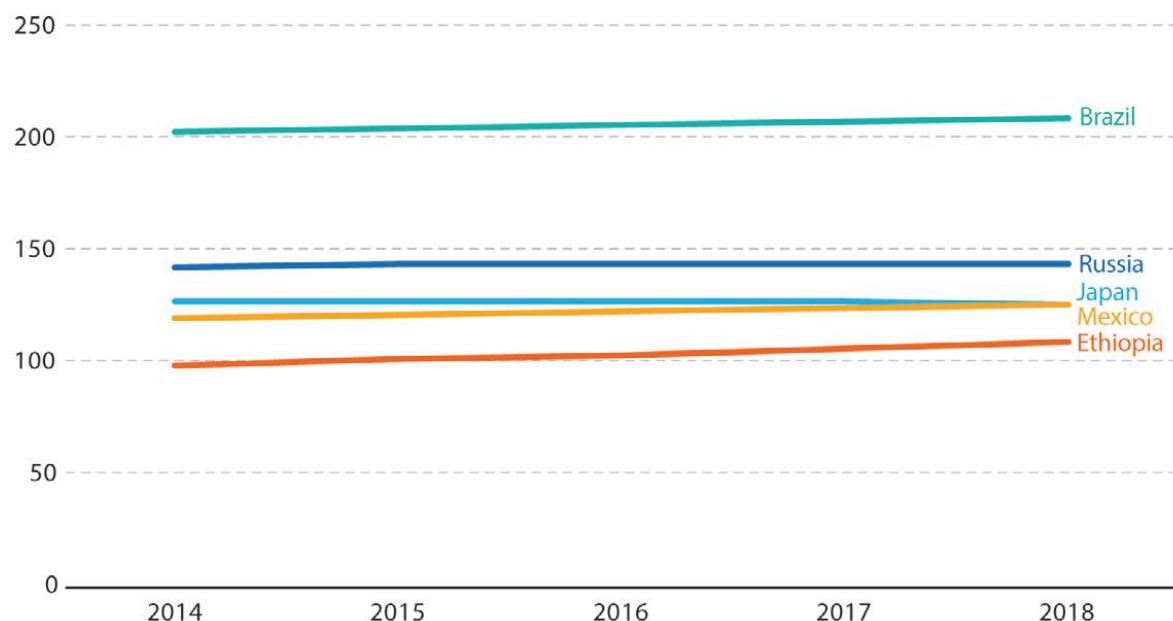
Ethiopia

The paired bar chart can be used to show changes over time and can be used to examine changes within and between countries.

# *Changes over time*

**Change in population from 2014 to 2018**

(Millions of people)

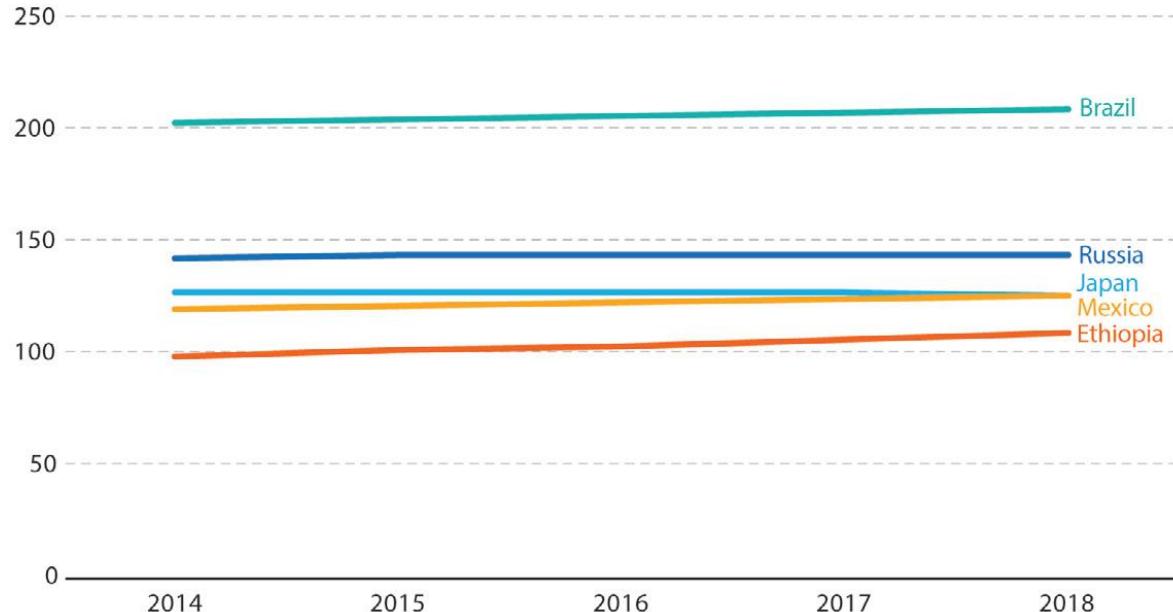


- Another use for the paired bar chart is to show changes over time.
- However, a line chart is a more familiar way to show changes over time.
- As compared to previous chart, this allows the reader to examine the population change within countries and the differences across countries.

# *Changes over time*

**Change in population from 2014 to 2018**

(Millions of people)



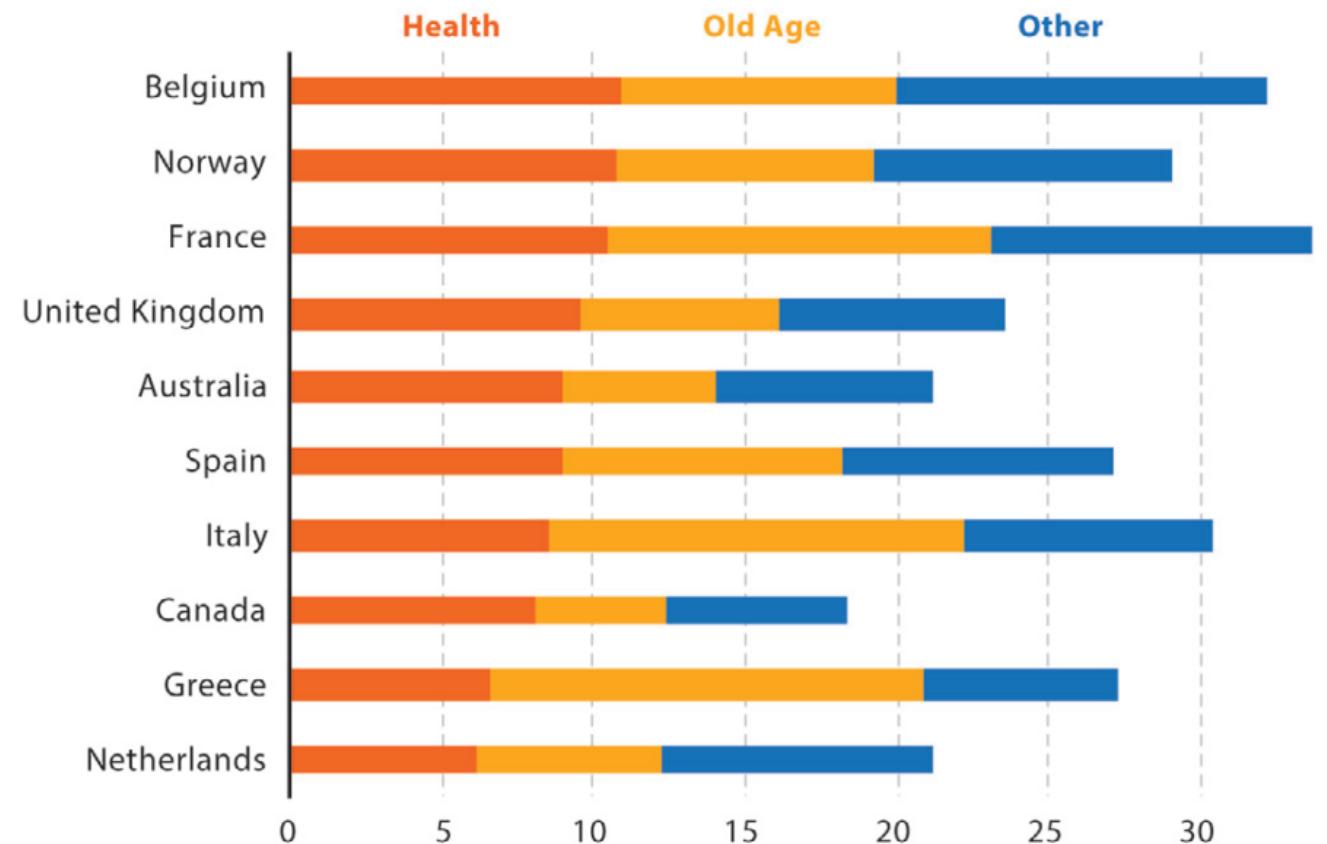
- Clutter is the main issue to keep in mind when assessing whether a paired bar chart is the right approach.
- With too many bars, it can be difficult for the reader to see the patterns and determine whether the most important comparison is between or within the different categories.

# *Stacked bar*

- Another variation on the bar chart is the stacked bar chart.
- While the paired bar chart shows two or more data values for each category, this chart subdivides the data within each category.
- The categories could sum to the same total, say, 100 percent, so that the total length of the bar is the same for every group.

## Social expenditures for 10 OECD countries

(Percent of GDP)



Source: Organisation for Economic Co-Operation and Development

- The stacked bar charts shows how different categories sum to a total.
- The interior series in the chart, however, are harder to compare with one another because they do not sit on the same baseline.

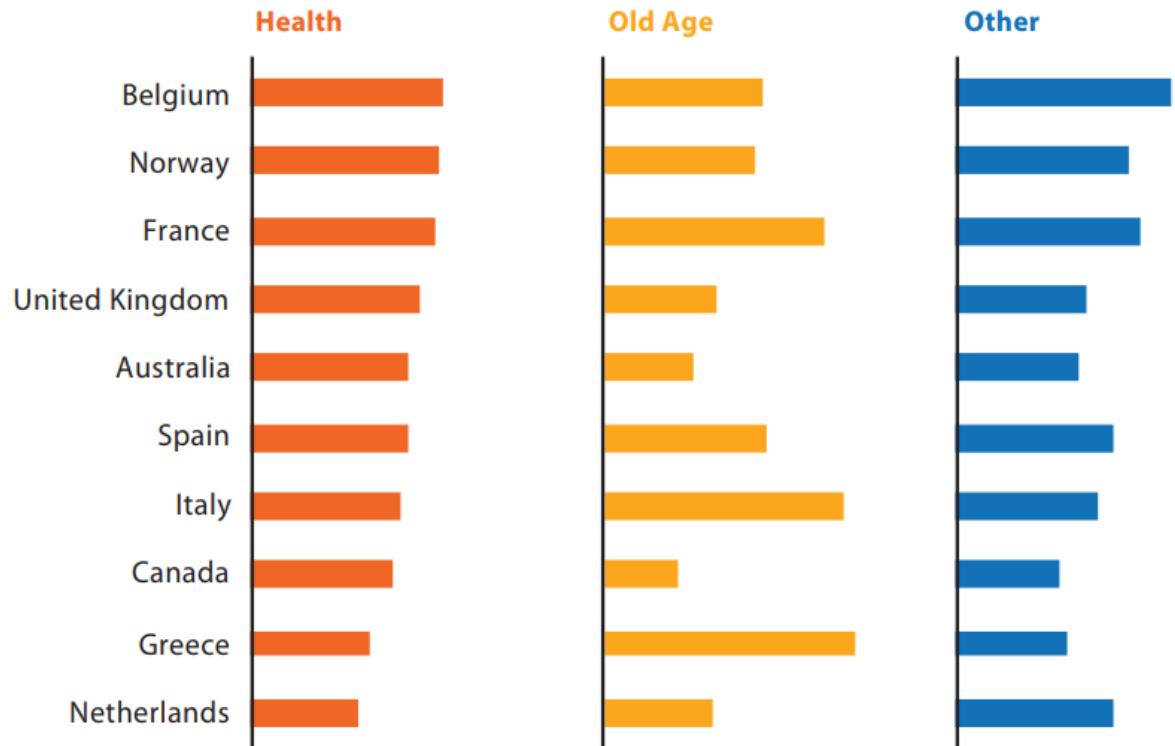
**Which country spends more on old-age programs, Italy or Greece?**

# *Stacked bar*

- One way to address the changing baseline is to break the graph apart so that each series sits on its own vertical baseline.
- This is a *small multiples graph*, arranged side by side.
- It's now easier to see that *Greece spends more on Old Age programs* than Italy.
- The tradeoff is that it is *harder to see the total* values.
- But that too can be overcome: You can still break up the stacked graph and add a final segment that represents the total amount

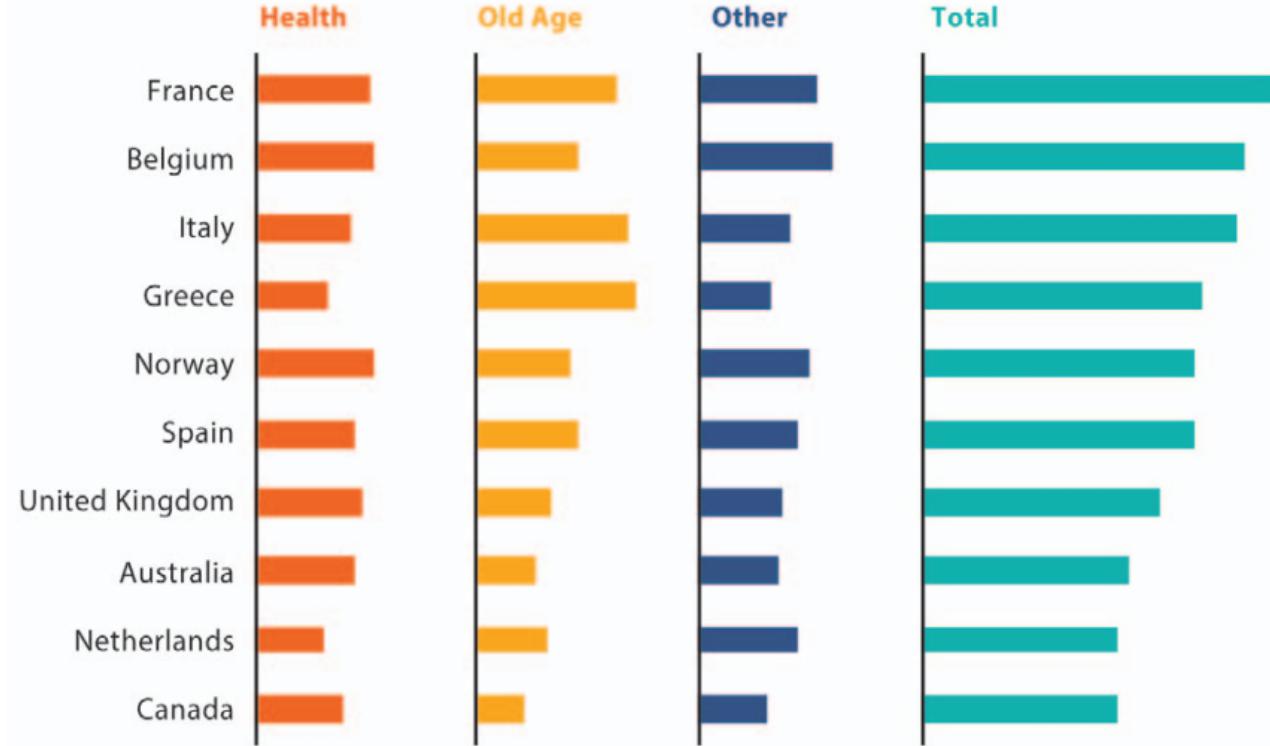
## Social expenditures for 10 OECD countries

(Percent of GDP)



## Social expenditures for 10 OECD countries

(Percent of GDP)



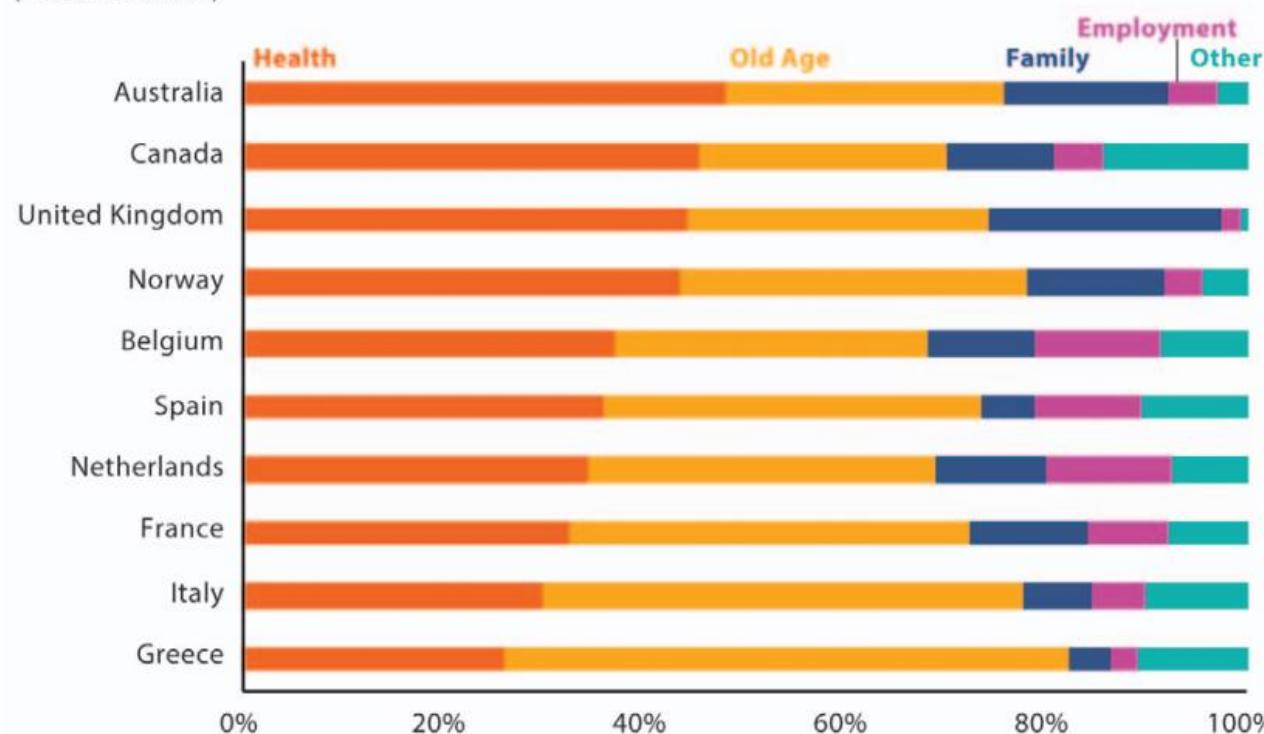
Instead of stacking all the data series together, we can break them up (either with or without the totals) to create a sort of small-multiples approach. Here, we move up to the top of the perceptual ranking list because each series sits on its own baseline.

# *Stacked bar*

- Even though different baselines in the standard stacked bar chart can make it more difficult to compare values, there are cases when the stacked bar chart is preferable.
- In following stacked bar chart, more spending categories are included and divided into shares of the total so the graph highlights the distribution.
- In this view, it becomes clear that around three-quarters of total government spending in these countries goes to programs for health care and old age programs.
- That observation is harder to see in the version on the right, where each category is placed on its own vertical baseline.
- Even though it is easier to compare differences in each category across countries, you don't see large differences between them.

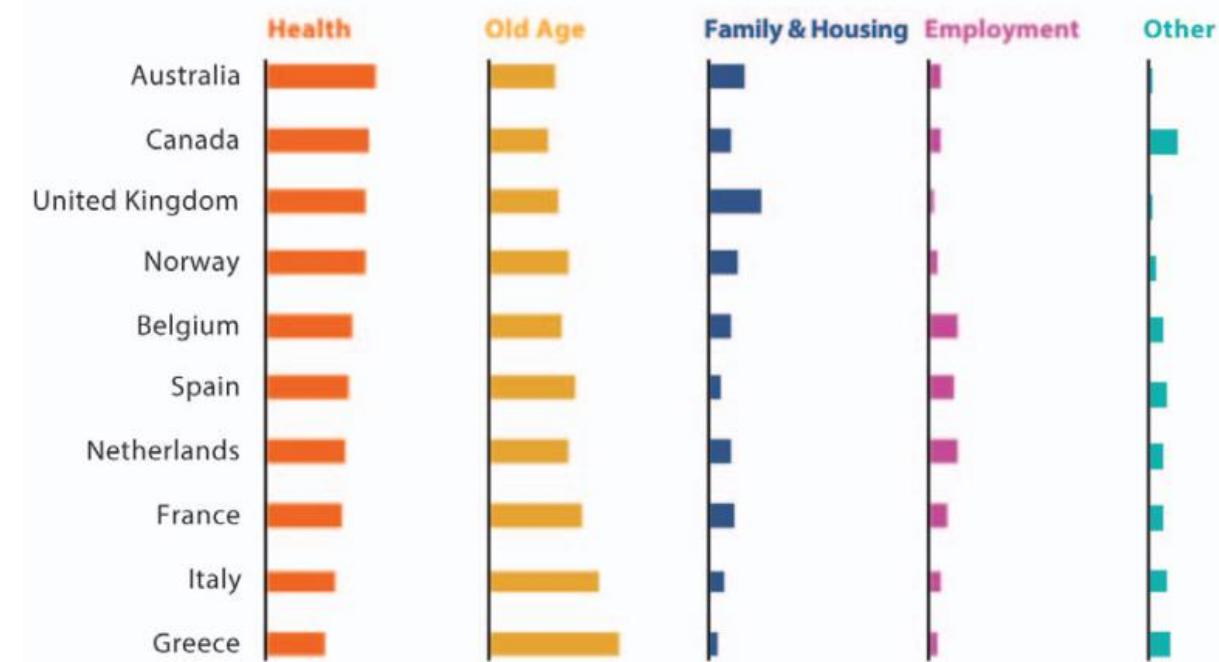
## Social expenditures for 10 OECD countries

(Percent of GDP)



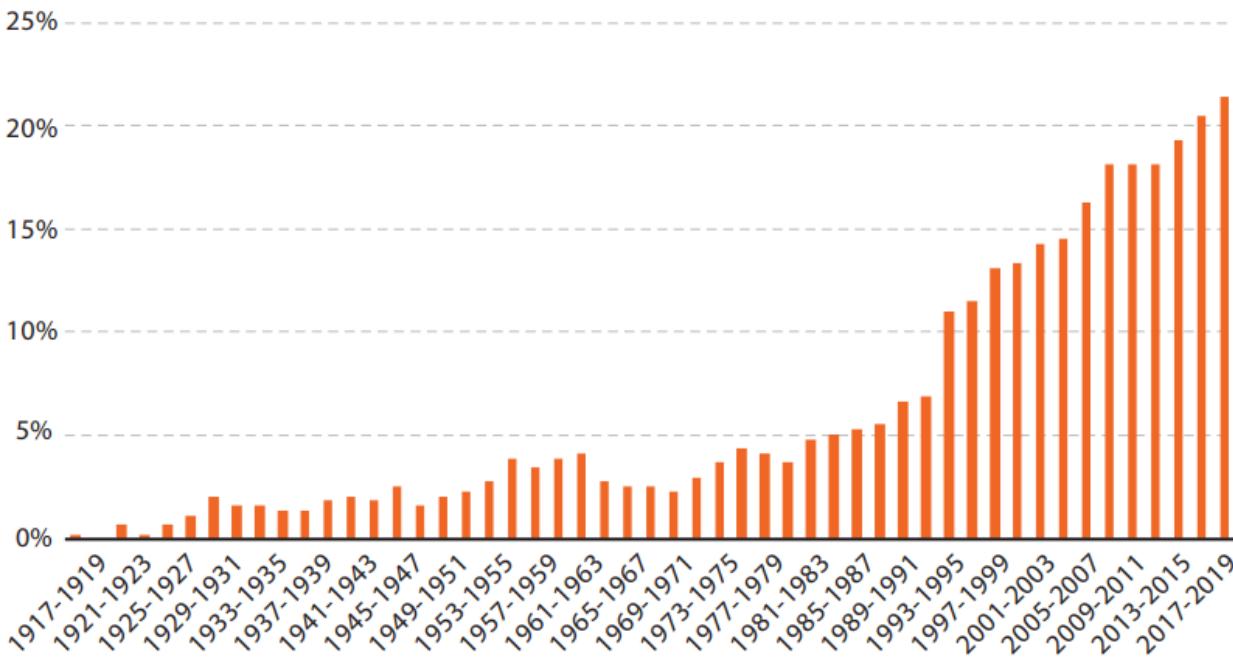
## Social expenditures for 10 OECD countries

(Percent of GDP)



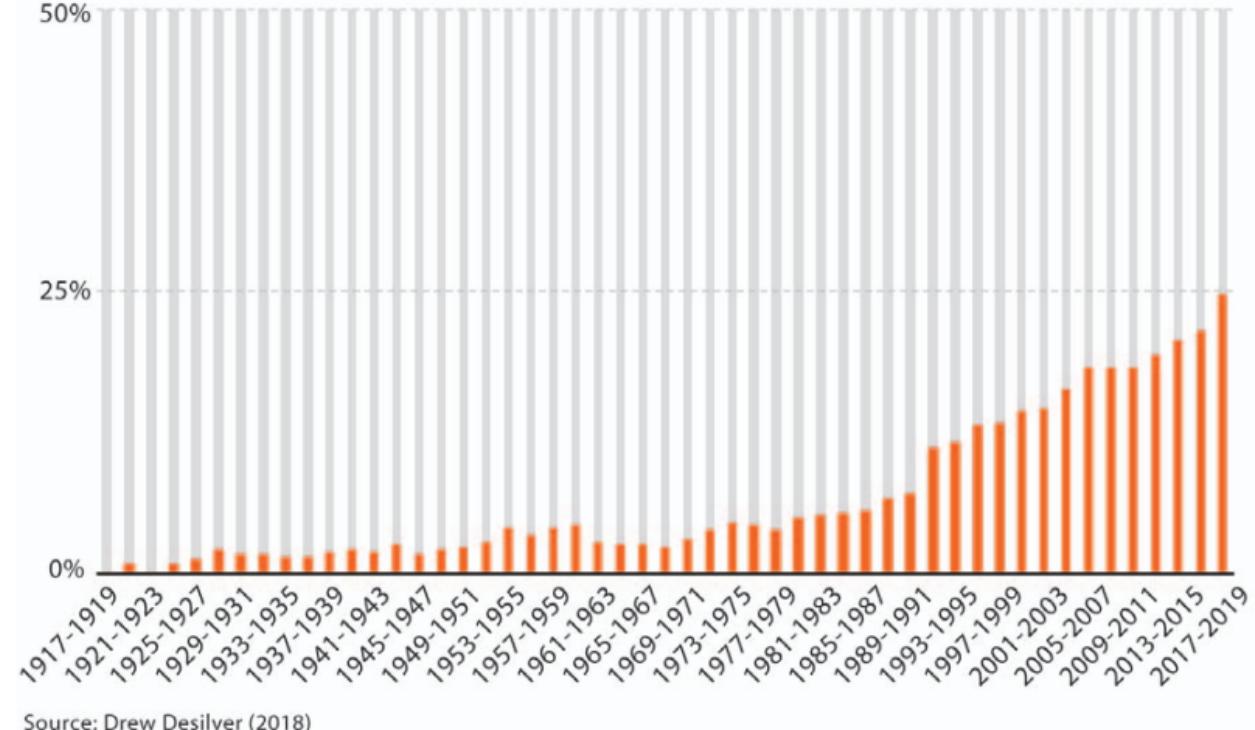
In these examples, we can see how our ability to compare different values within and across countries varies between these two views.

## The 116th Congress represents the biggest jump in women members since the 1990s



Source: Drew Desilver (2018)

## The 116th Congress represents the biggest jump in women members since the 1990s



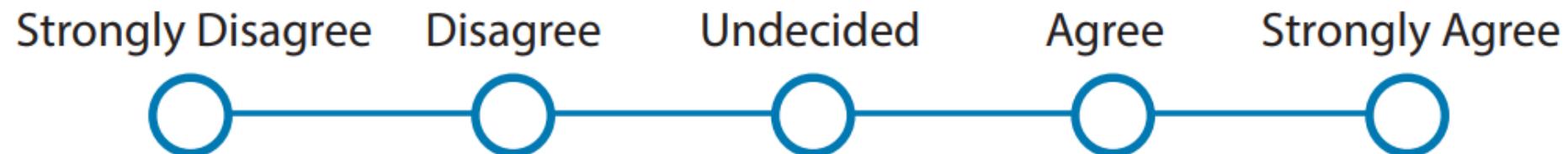
Source: Drew Desilver (2018)

- The version on the left shows the raw percentages; the vertical axis ranges from 0 to 25 percent.
- The version on the right shows the same data, but stacks a gray series on top of the data values to 50 percent.
- In this version, we can emphasize that although there is a dramatic increase in the share of women it is still small.

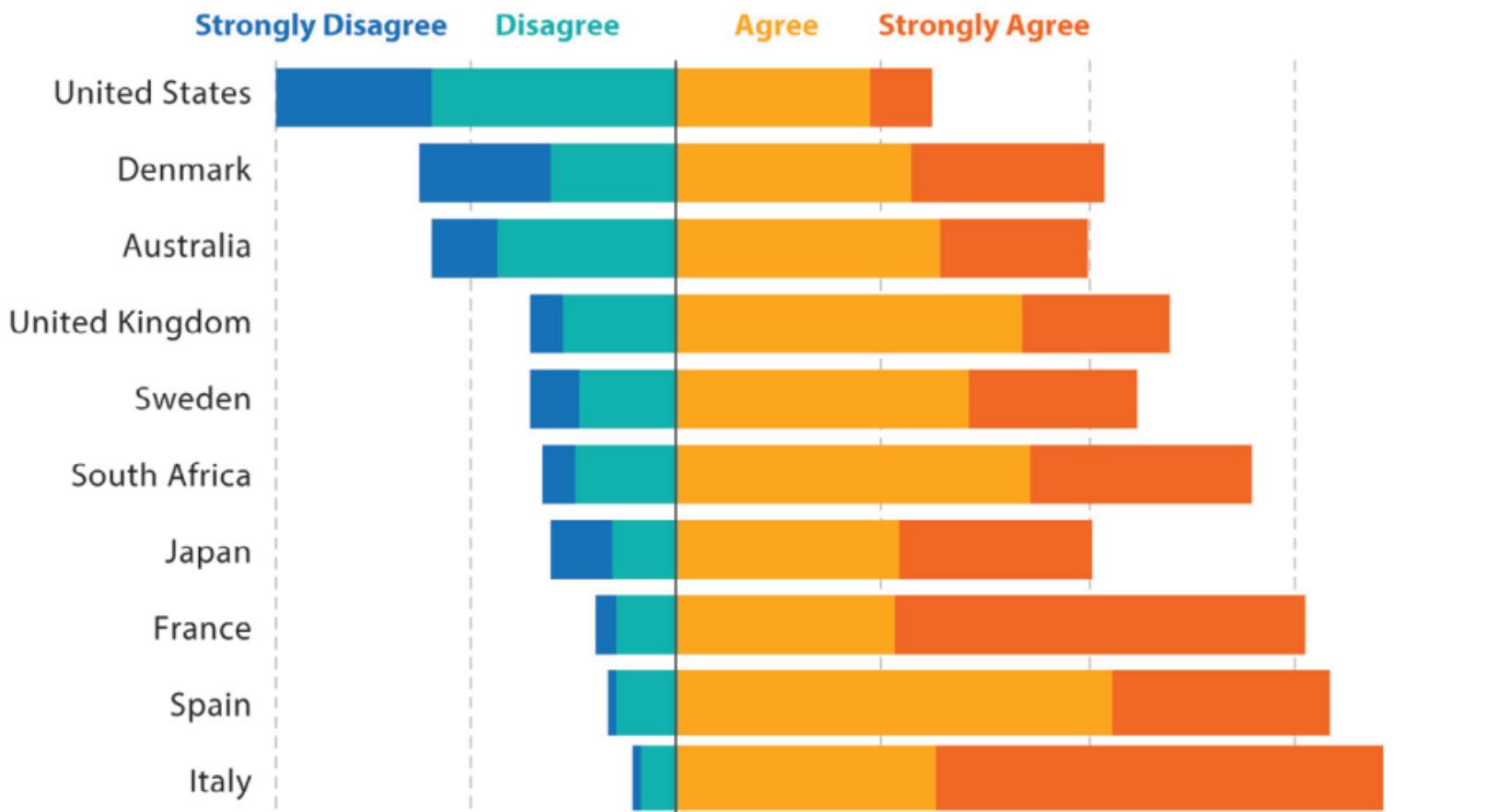
# *Diverging bar*

- A variation on the stacked bar chart is one in which the stacks diverge from a central baseline in opposite directions.
- These are often found in surveys where the responses are arrayed in ranges from, for instance, strongly disagree to strongly agree.
- These are often called “Likert Scales,” named after the psychologist Rensis Likert, who invented the scales in the early 1930s.

**This book is fun to read.**



**It is the responsibility of government to reduce differences in income  
between people with high & low incomes**  
(Percent)

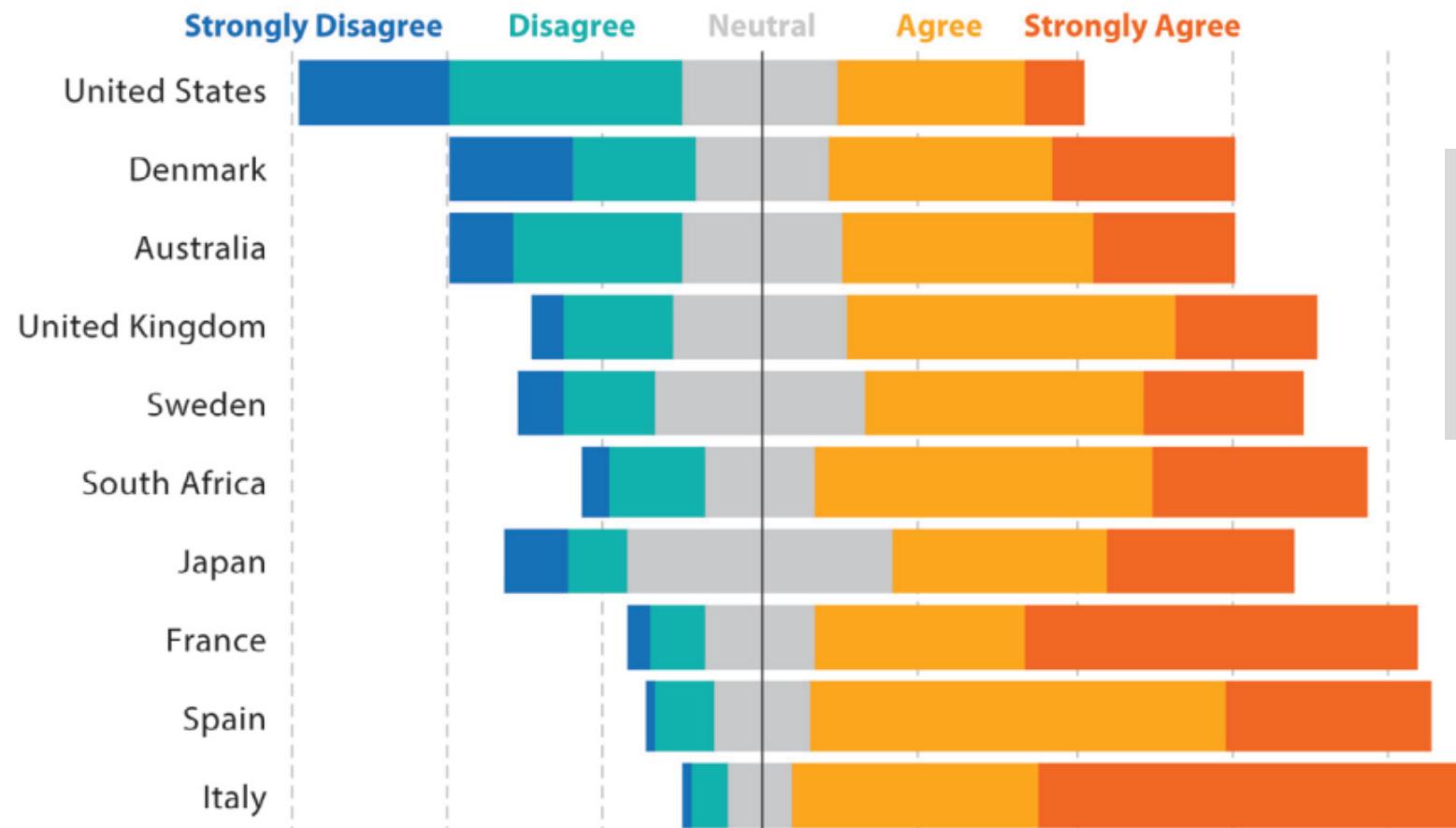


Source: International Social Survey Programme, 2009

The diverging bar chart can show differences in opposing sentiments or groups, such as  
“agree/disagree” or “true/false.”

## It is the responsibility of government to reduce differences in income between people with high & low incomes

(Percent)



By definition, the neutral survey response is neither agree nor disagree, and should therefore be grouped with neither category.

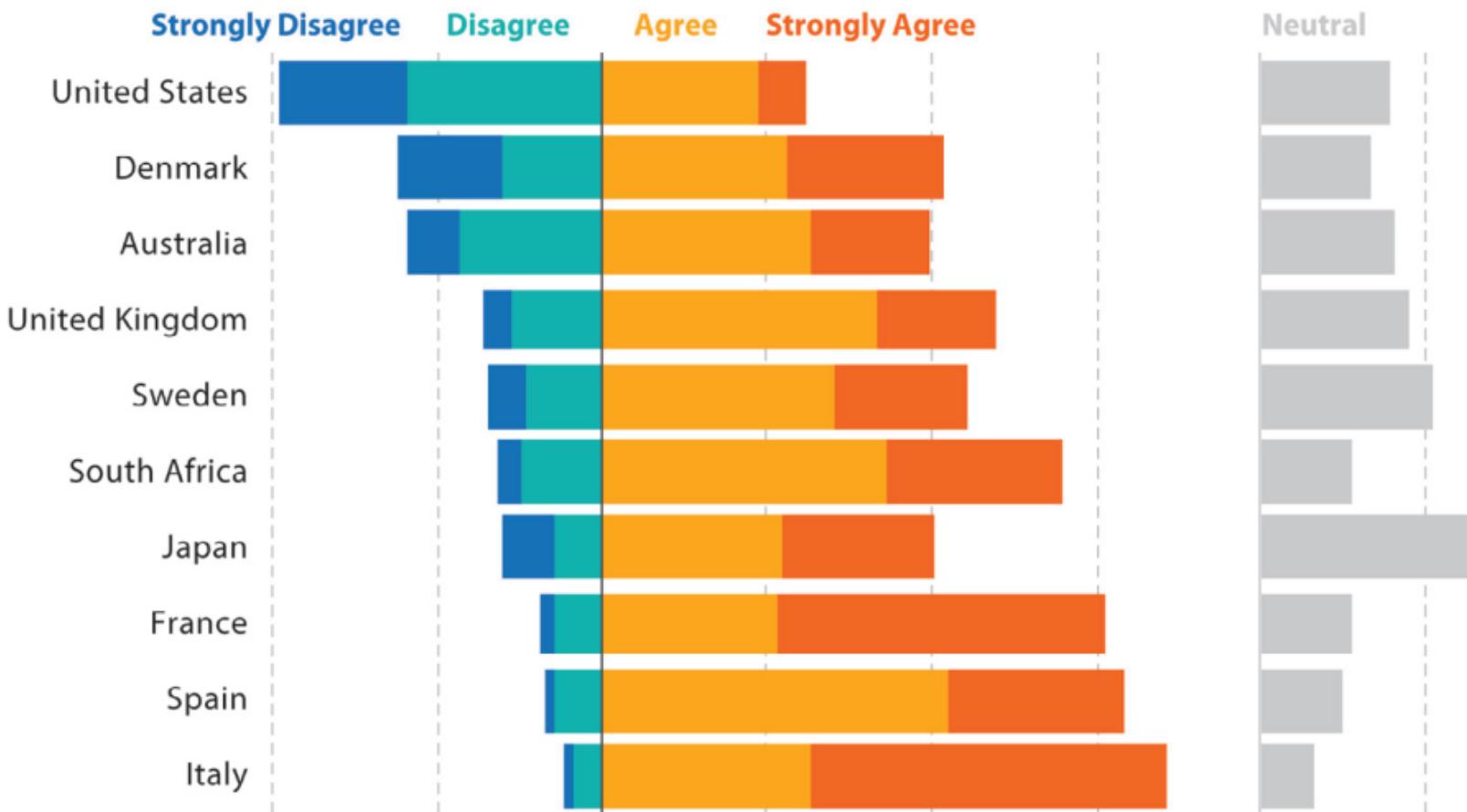
Source: International Social Survey Programme, 2009

Placing the *Neutral* category of a diverging bar chart in the middle wrongly implies that the neutral responses are split between the two sentiments.

# *Diverging bar*

- Placing the neutral category in the middle of the chart along the vertical baseline creates a misalignment between the two groups and implies the neutral responses are split between the two sentiments.
- None of the segments sit on a vertical baseline.
- Placing it to the side of the chart is a better strategy because the disagree, agree, and neutral categories now all sit on their own vertical axes, even though the neutral category is somewhat emphasized as it sits to the side.

**It is the responsibility of government to reduce differences in income  
between people with high & low incomes**  
(Percent)



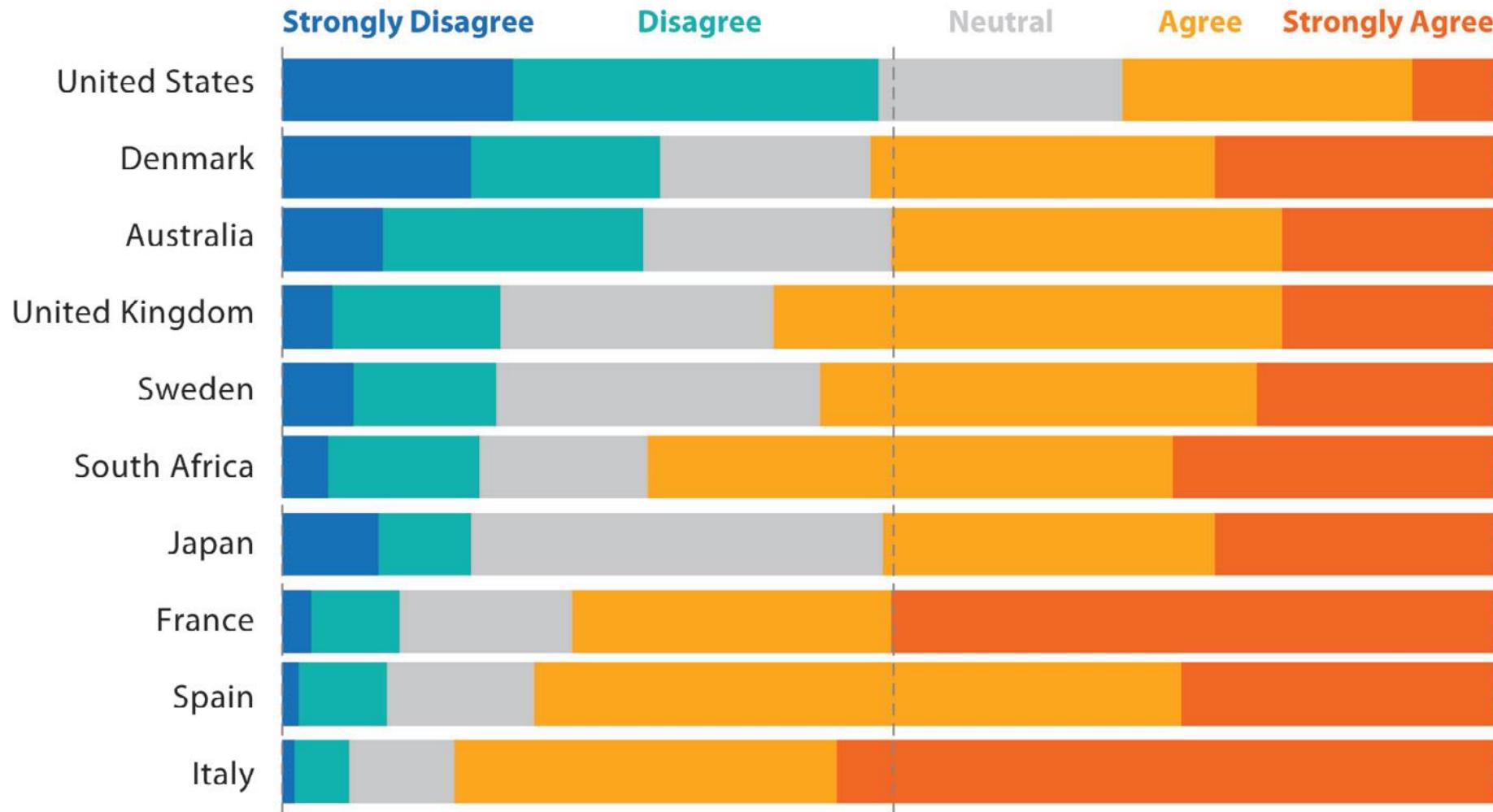
Source: International Social Survey Programme, 2009

A better approach is to place the *Neutral* category off to the side of the graph.

# *Diverging bar*

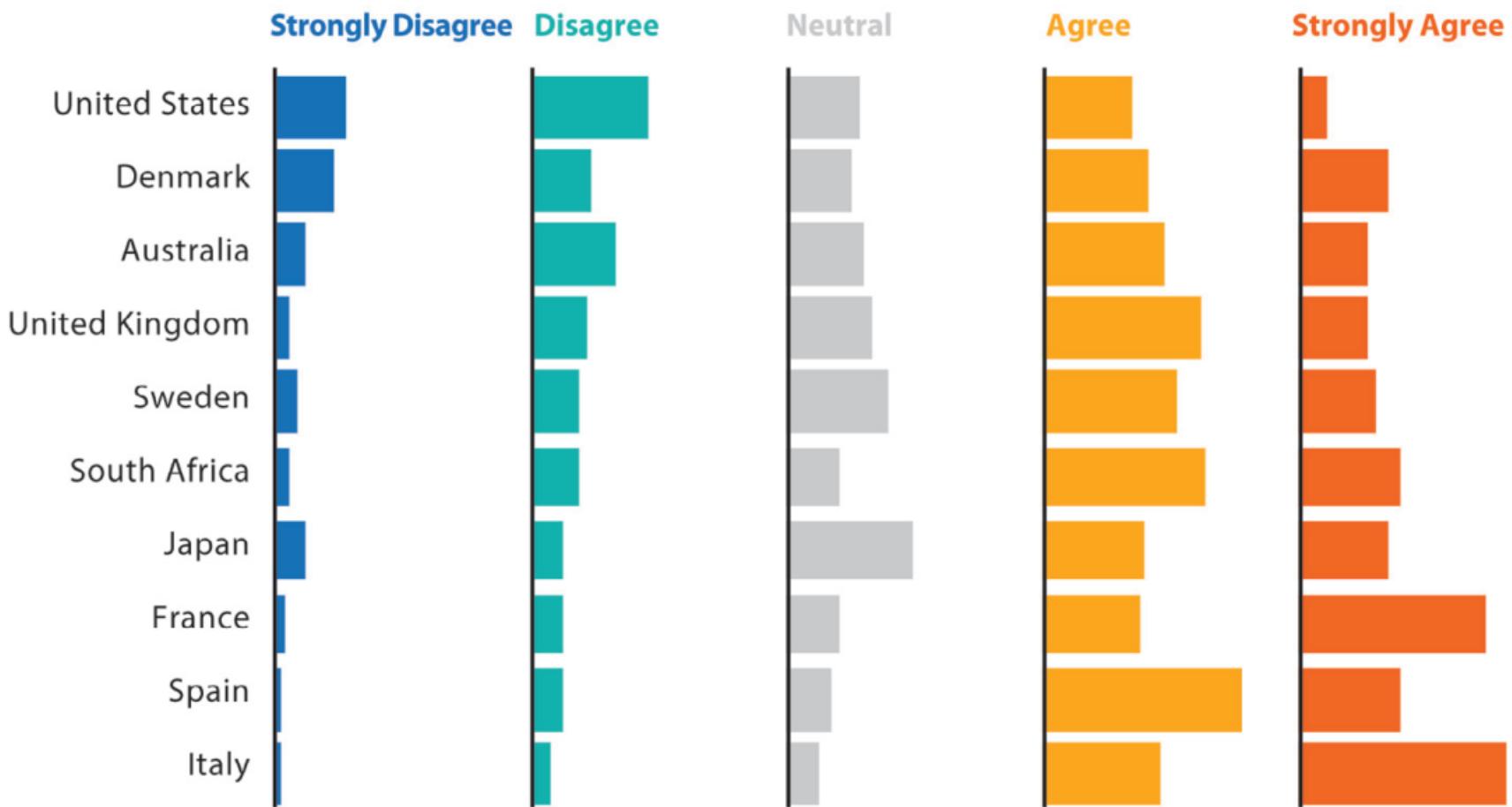
- Another alternative—regardless of whether you have a neutral category—is the stacked bar chart.
- In this view, the different categories sum to 100 percent, and one can more easily compare the totals between the countries.
- A good strategy is to mark specific aggregate values to guide the reader.
- Here, for example, the 50-percent position is marked to make it clear for which countries the total “agree” and “disagree” sentiments are at least half of the total.

**It is the responsibility of government to reduce differences in income between people with high & low incomes**  
(Percent)



A stacked bar chart can be used to show these kinds of Likert scales.

**It is the responsibility of government to reduce differences in income  
between people with high & low incomes**  
(Percent)



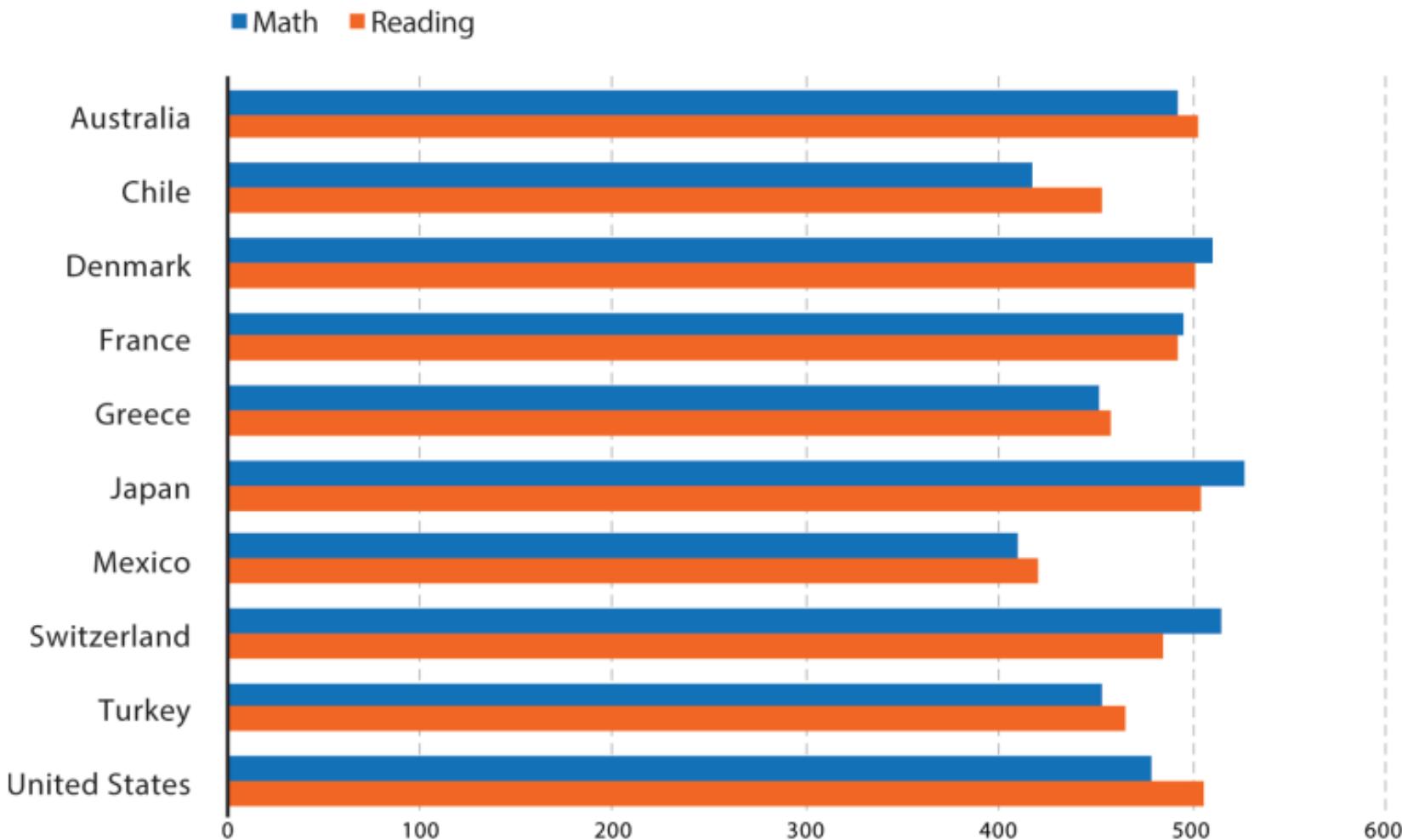
Source: International Social Survey Programme, 2009

The small multiples bar chart is yet another way to visualize these kinds of data.

# *Dot plot*

- The **dot plot** also called a **dumbbell chart**, **barbell chart**, or **gap chart** is one of the preferred alternatives to a paired or stacked bar chart.
- Developed by William Cleveland, one of the early pioneers in data visualization research, the **dot plot uses a symbol**—often but not always a circle—**corresponding to the data value, connected by a line or arrow.**
- The **data values correspond to one axis** and **the groups to the other**, which do not necessarily need to be ordered in a specific way, though sorting can help.
- The dot plot is **an easy way to compare categories**—especially many categories

## PISA scores for math and reading among 10 OECD countries



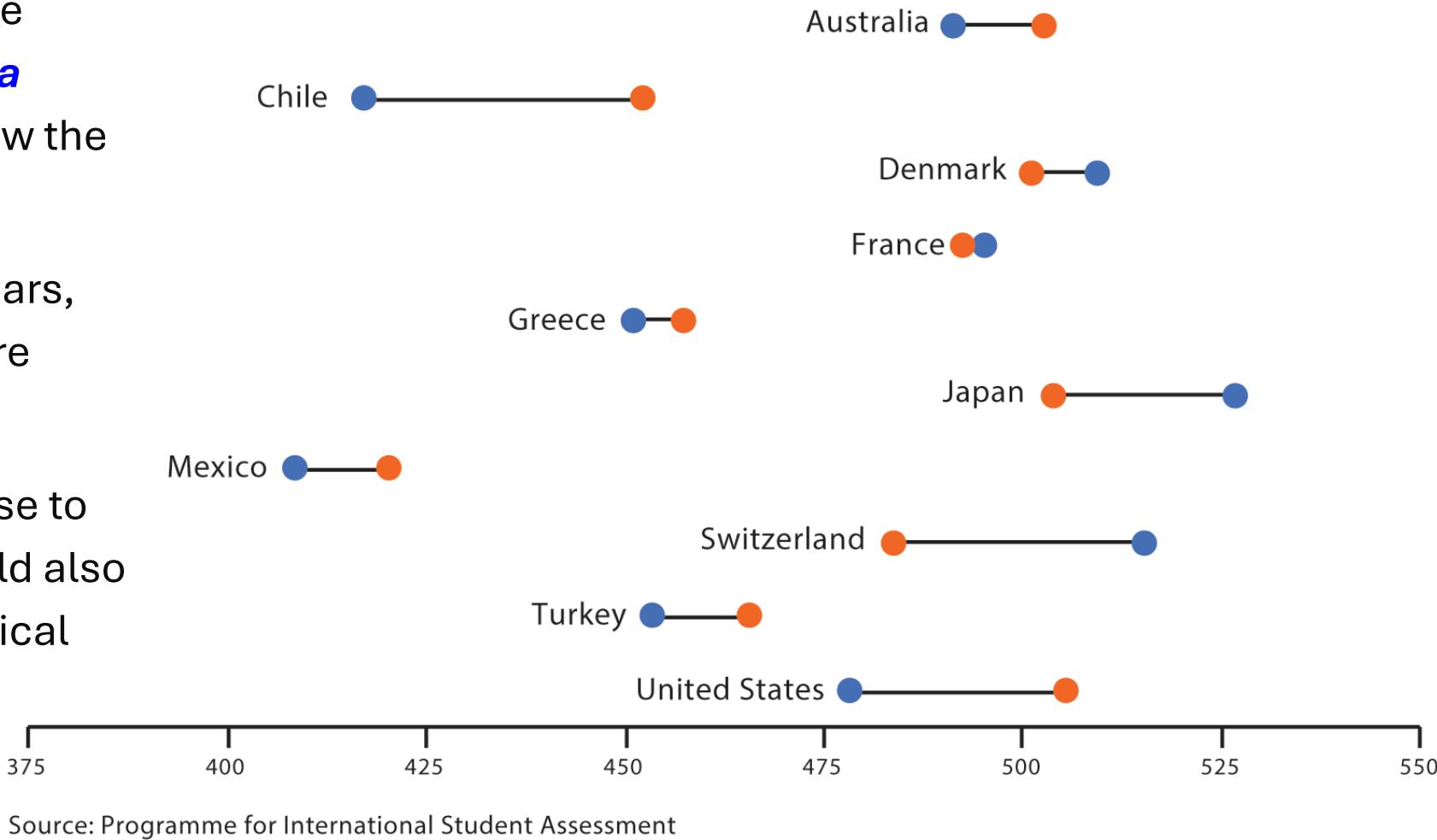
Source: Programme for International Student Assessment

This simple bar chart shows math and reading scores across multiple countries. Bar charts are often the default visual for these kind of data, but it looks heavy and dense.

## PISA scores for math and reading among 10 OECD countries

● Math ● Reading

- By contrast, the dot plot shows the same data with a **dot at each data value connected by a line** to show the range or difference.
- The circles use less ink than the bars, which lightens the visual with more empty space.
- The country labels are placed close to the leftmost dot, though they could also be set off to the left along the vertical axis.



Notice how more white space lightens the visualization.

# *Dot plot*

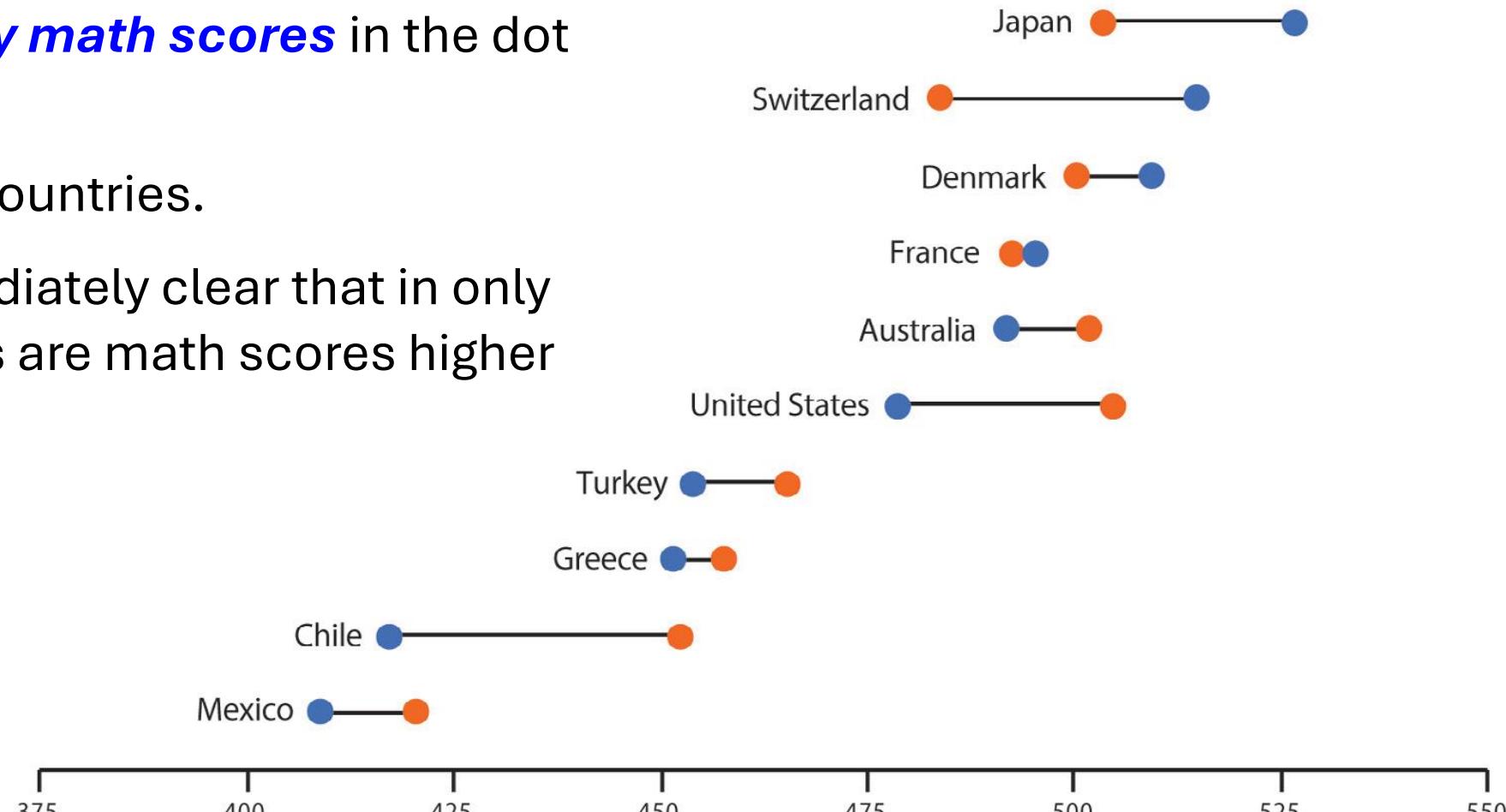
- Dot plots are not restricted to two dots and a connecting line, nor are they restricted to simply comparing different categories.
- You can use dot plots to show a change between two years, for example. You could use different shapes or icons or arrows instead of lines to denote direction.
- You can also use more than two objects.
  - E.g., we could add science test scores to this plot, but we need to be sure to add sufficient labeling so our reader knows what each object on the graph represents.
  - Axes and gridlines can be included or not, depending on how important it is for the reader to determine the exact values

# *Dot plot - caution*

- It's not entirely obvious when the direction of the values change, as in the last chart.
- Did you notice that math scores were higher than reading scores in four of the countries in the dot plot above?
- That difference is not immediately evident unless the reader carefully examines the points and their coloring.
- In this and other cases, we should consider how sufficient annotation, clear labeling, and highlighting colors can help clarify different directions

## PISA scores for math and reading among 10 OECD countries

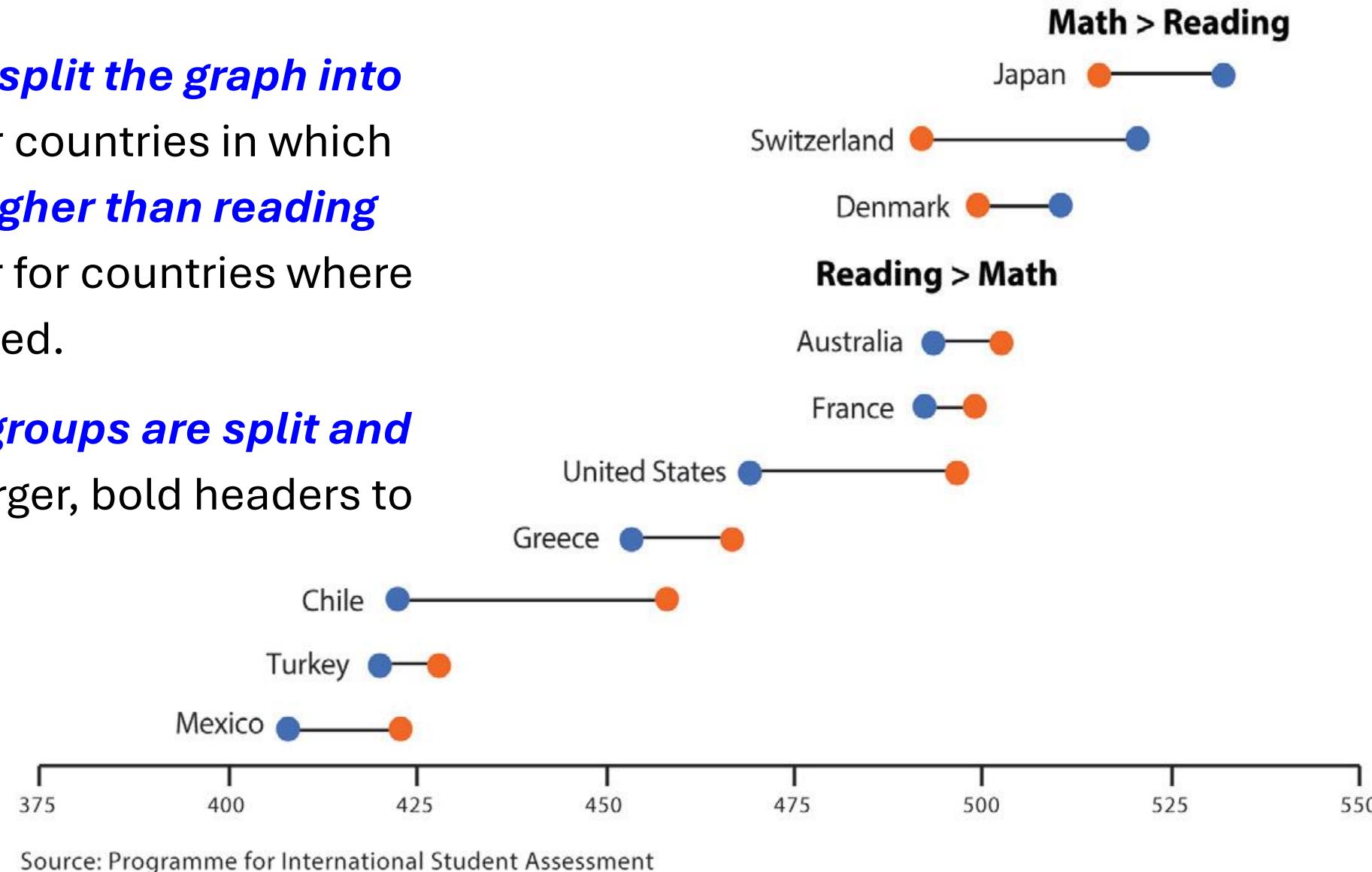
- The data are **sorted by math scores** in the dot plot
- It helps organize the countries.
- But it is still not immediately clear that in only the first four countries are math scores higher than reading scores.



Source: Programme for International Student Assessment

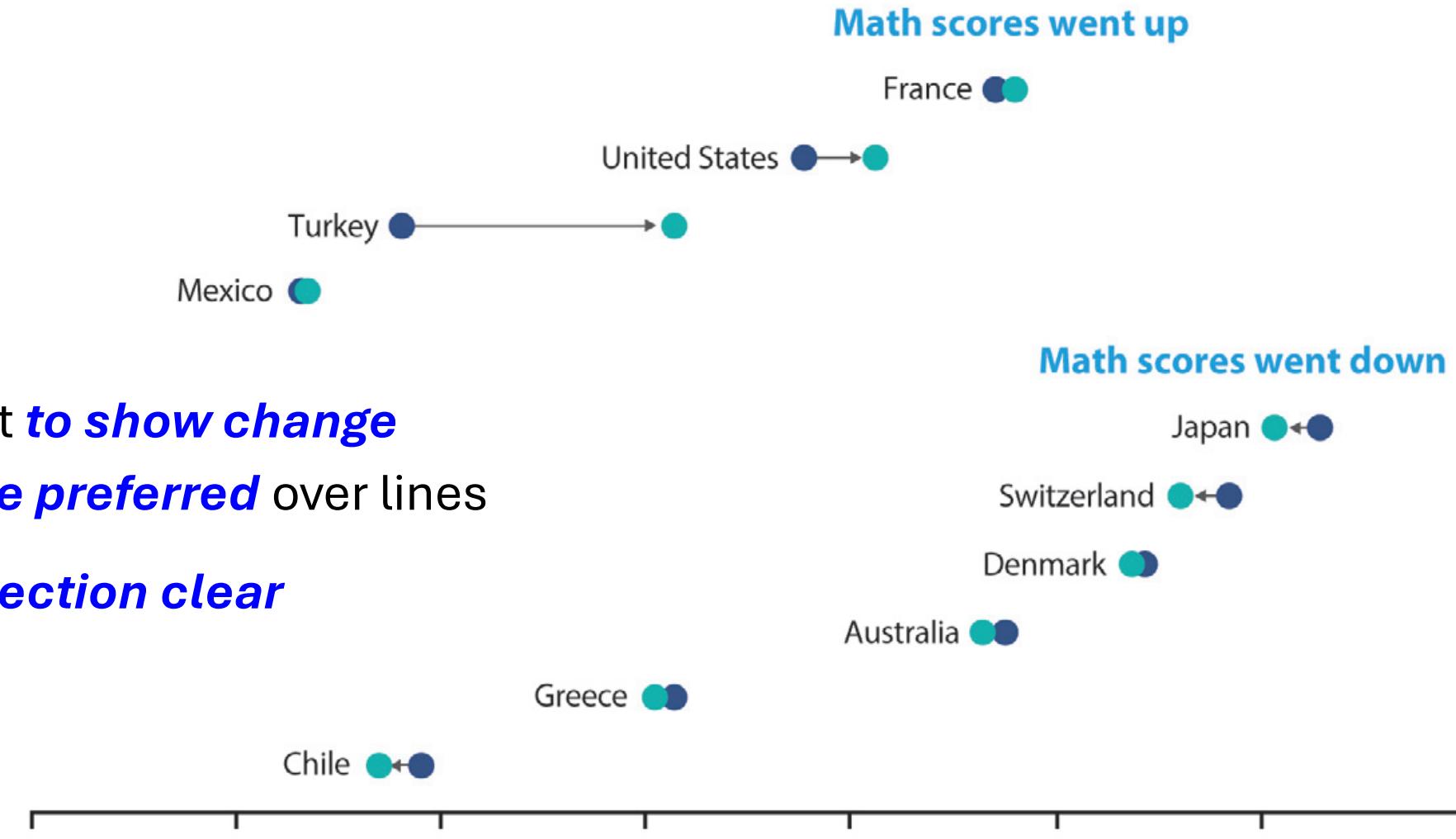
## PISA scores for **math** and **reading** among 10 OECD countries

- One approach is to **split the graph into two groups**, one for countries in which **math scores are higher than reading scores** and another for countries where the opposite occurred.
- In this version, the **groups are split and then sorted** with larger, bold headers to distinguish them.



## PISA math scores rose for 4 of 10 OECD countries between 2015 and 2018

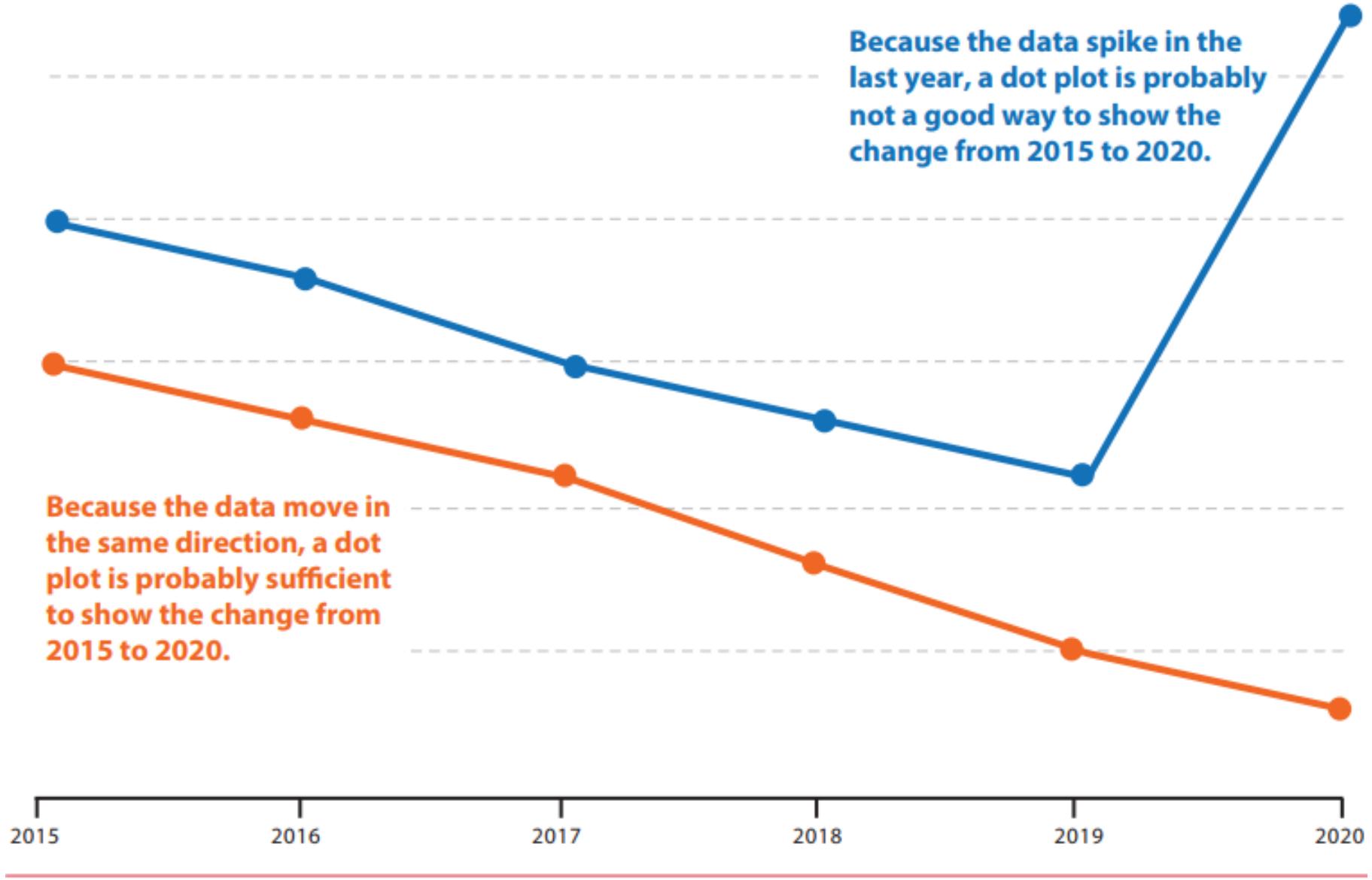
- When using a dot plot **to show change over time, arrows are preferred** over lines
- It helps **make the direction clear**



Source: Programme for International Student Assessment

# *Dot plot - caution*

- Another word of caution for dot plots that show changes over time.
- The dot plot is, by definition, a summary chart.
- It does not show all of the data in the intervening years.
- If the data between the two dots generally move in the same direction, a dot plot is sufficient.
- But if the data contain sharp variations year by year, a dot plot will obscure that pattern (as it also does for bar charts).
  - E.g., if test scores decreased between 2015 and 2019 and then increased sharply between 2019 and 2020, the dot plot would only show an overall increase, masking the change in the intervening years.
  - In some cases, you may not have a choice—if you are using data from the decennial U.S. Census, by definition you will only have data for every ten years.
  - That's something you can't help, but if you are familiar enough with your content, you'll know whether showing only those points is enough to clearly and accurately make your point.



Because dot plots are essentially a summary plot, be wary of using highly variable data with dot plots.

# *Marimekko and mosaic charts*

- Marimekko charts may look odd at first, but they are just an extension of the bar chart.
- This type of chart is useful to compare two variables:
  - one comparing categories and
  - one showing how they sum to a total
- The name of the chart comes from the [Finnish design firm Marimekko](#), founded in 1951 by Armi Ratia and her husband, Viljo.
- Early Marimekko style featured ***straight, oversized, geometric patterns*** and ***bright colors***.

# Marimekko

- In the standard vertical bar (or column) chart, the data are measured along the height of the vertical axis and the widths of the columns are identical.
- The Marimekko chart takes that standard column chart and expands the width of each bar according to another data value.
- The Marimekko chart is an easy way to add a second variable to your standard column or bar chart



---

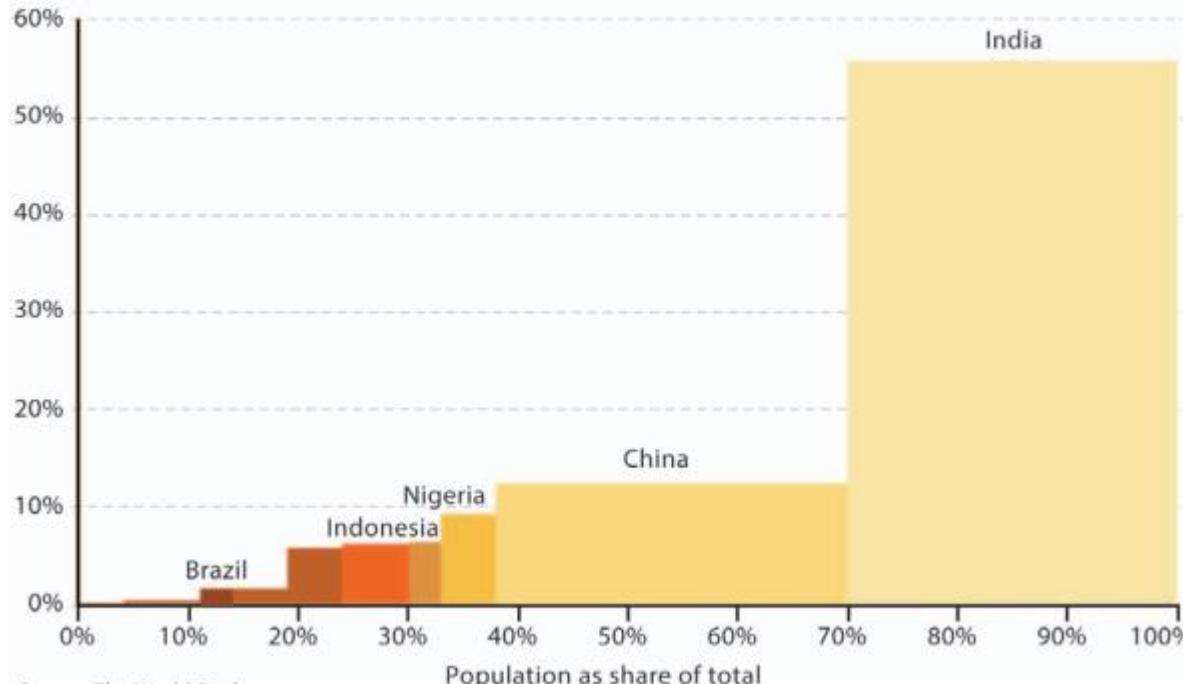
Early Marimekko fabric styles like this one featured straight, oversized, geometric patterns and bright colors.

# *ten most populous countries*

- Share of people with less than \$5.20 per day and the share of the total population.
- Percent of people with less than \$5.20 per day is plotted along the vertical axis
- Widths of each bar scaled according to the share of each country's population summing to 100 percent

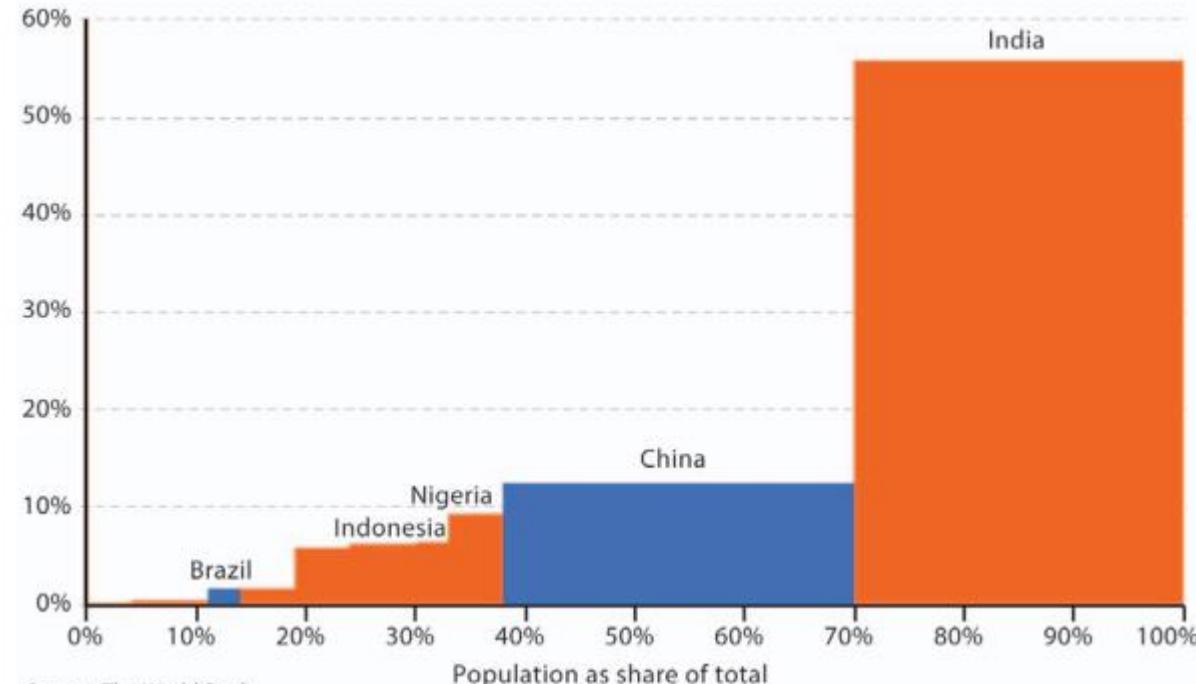
**High population, high poverty**

(Percent of people with less than \$5.20/day)

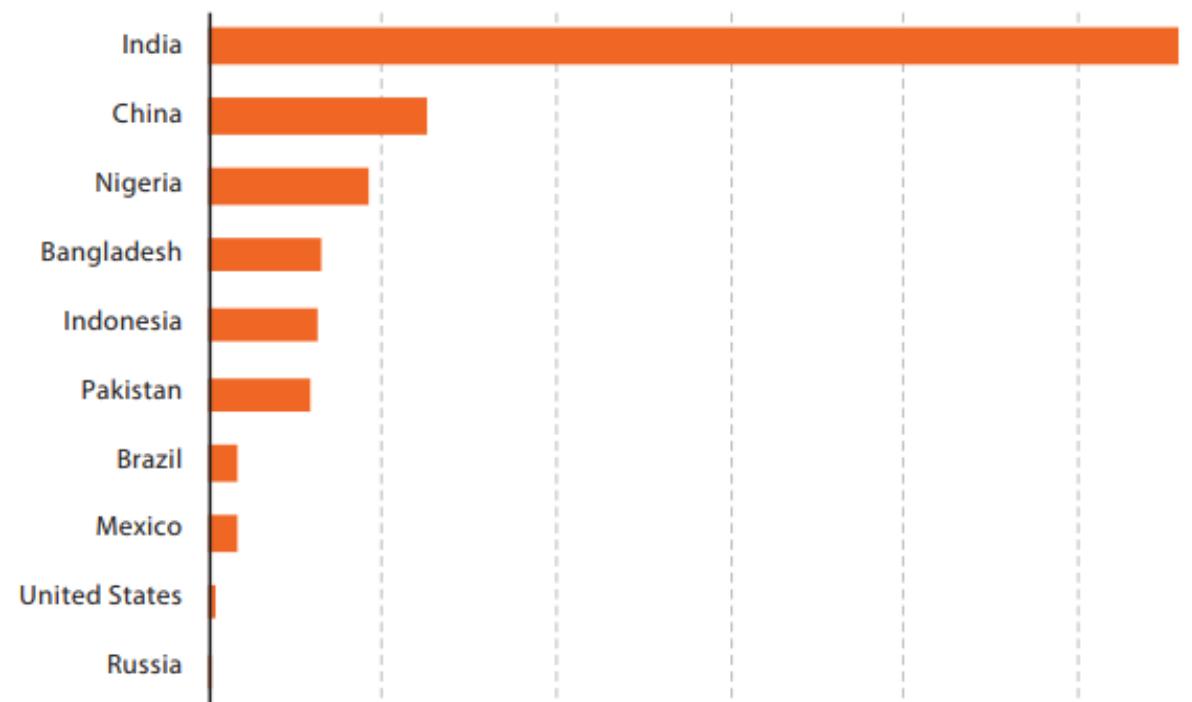


**High population, high poverty**

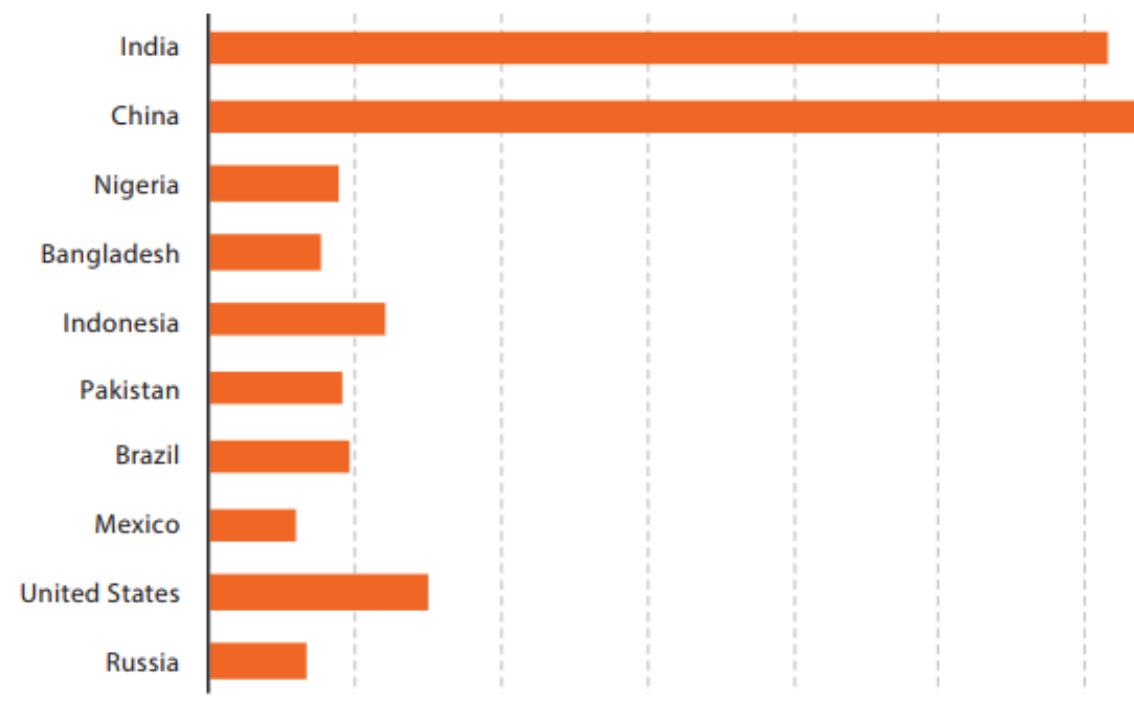
(Percent of people with less than \$5.20/day)



**Percent of people with less than \$5.20/day**



**Population (as percent of all countries)**



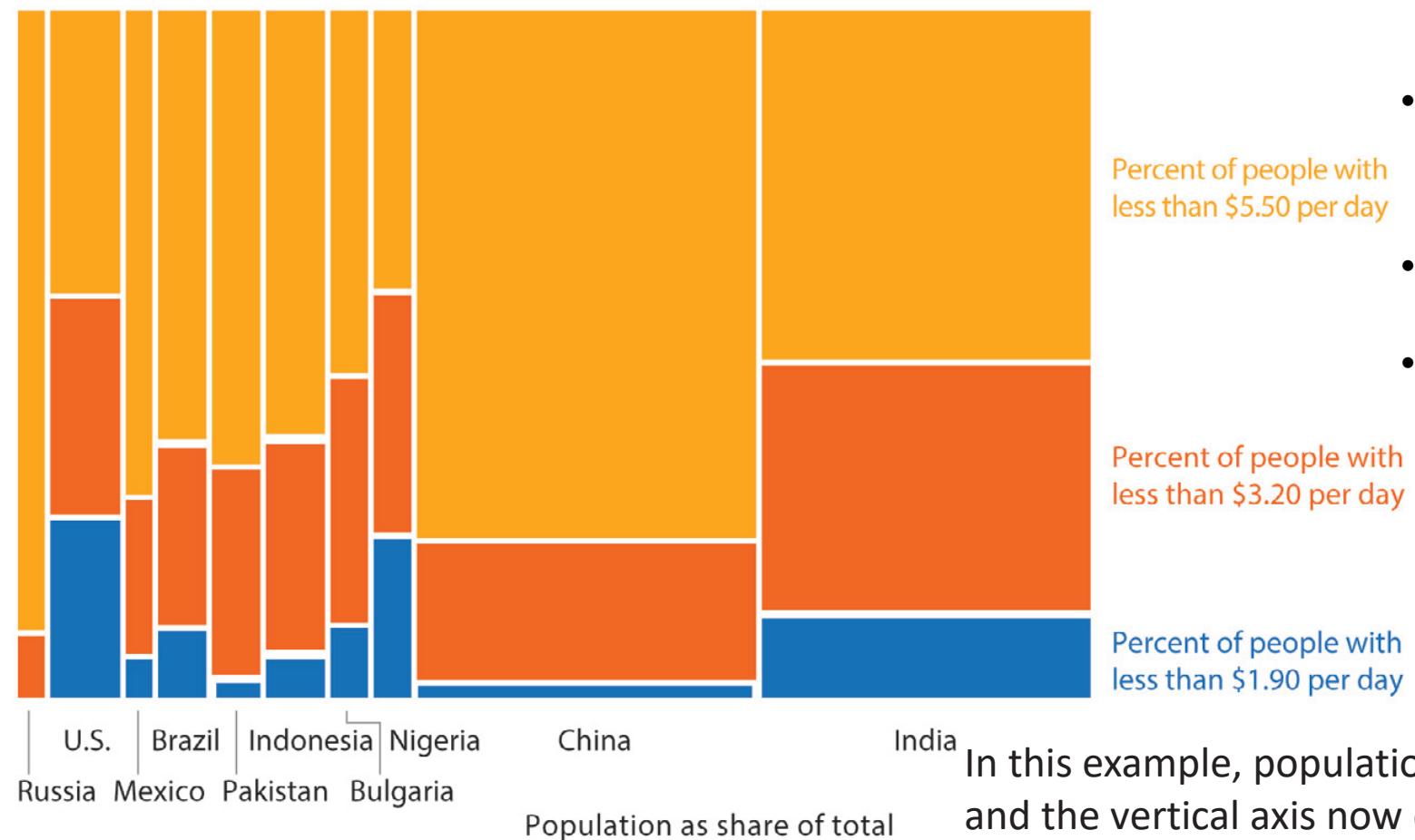
Source: The World Bank

Source: The World Bank

The two variables could also be plotted separately in two bar charts, and while these graphs are familiar and easy to read, they do not communicate the relationship between the two variables as well.

# Mosaic chart

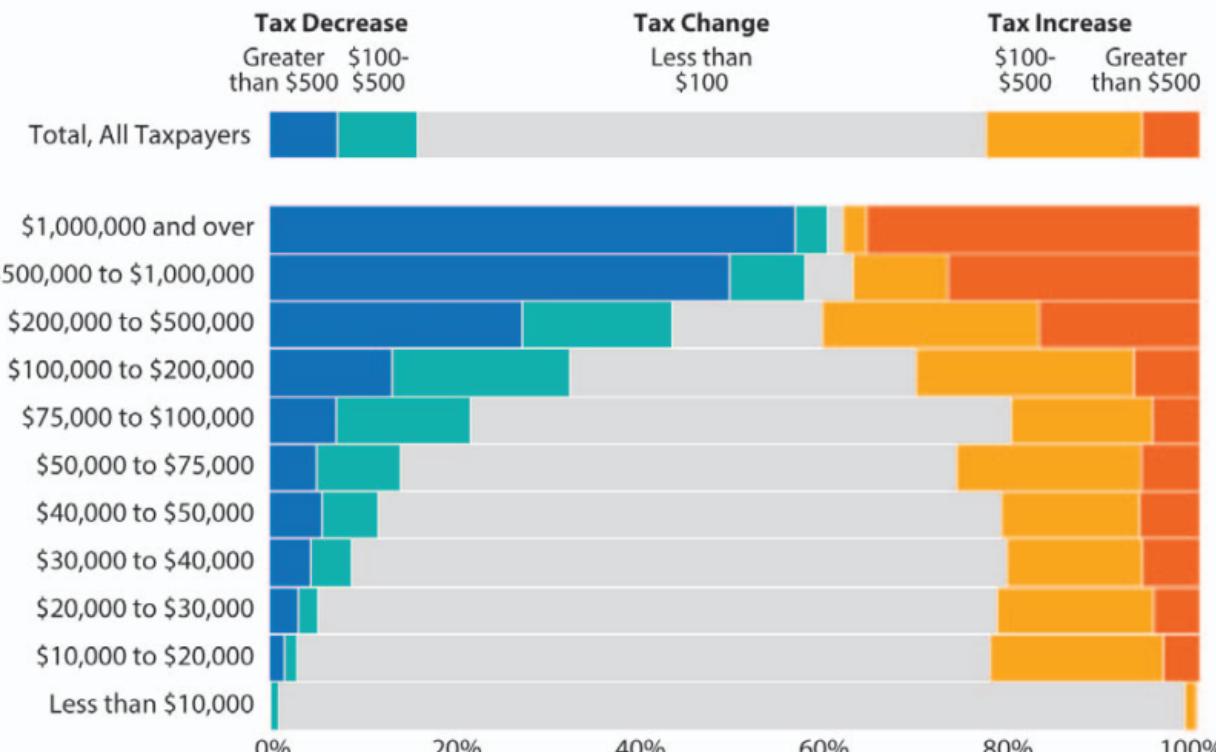
High population, high poverty



- A variation on the Marimekko is to have ***both the heights and the widths of the bars sum to 100 percent***.
- This is sometimes called a ***mosaic chart***.
  - You fill the entire graph space
  - Provides a part-to-whole perspective of the data along both dimensions.

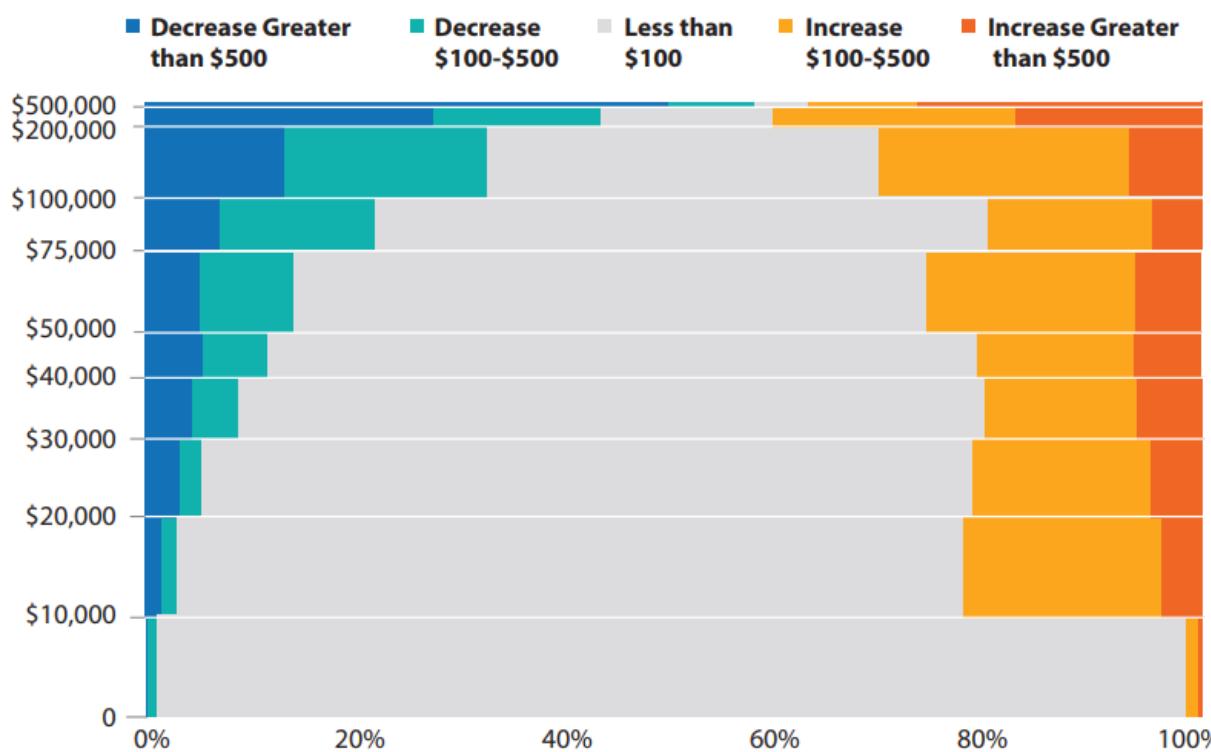
In this example, population is still plotted along the horizontal axis, and the vertical axis now contains three categories for people with low levels of income: share of people with less than \$1.90 per day, \$3.20 per day, and \$5.20 per day.

### Distribution of tax returns by size of tax change for the "Tax Cuts and Jobs Act"



Source: Joint Committee on Taxation

### Distribution of tax returns by size of tax change for the "Tax Cuts and Jobs Act"



Source: Joint Committee on Taxation

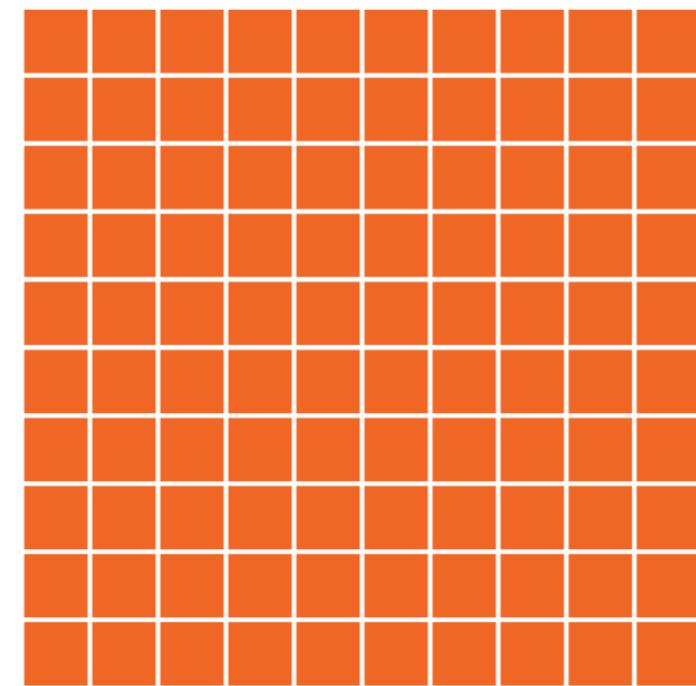
- Notice the difference between the stacked bar chart on the left (where all the bars are the same width) and the mosaic chart on the right.
- While the mosaic chart adds another variable, it is harder to see the details in the top category.

# *Unit charts*

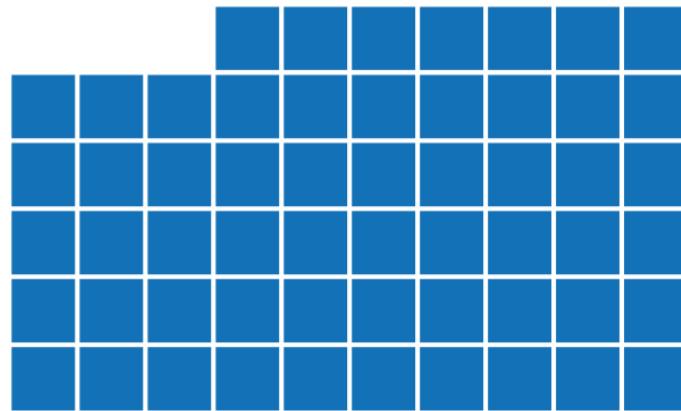
- Unit charts show counts of a variable.
- Each symbol can represent an observation or a number of units.
- For example, if one symbol represents ten cars and there are ten car icons, the reader mentally multiplies the two for the total of one hundred cars.
- You can use unit charts to show percentages, dollars, or any other discrete amount.
- You can arrange them in different directions or break them down into subcategories by using colors or outlines.

2000  
**100 million**

2015  
**57 million**



Unit charts use symbols to show counts of a variable



**Global out-of-school children of primary school age**

**100 million**

2000

**57 million**

2015

Source: The World Bank

BANs—or Big-Ass Numbers—are a way to just show the values

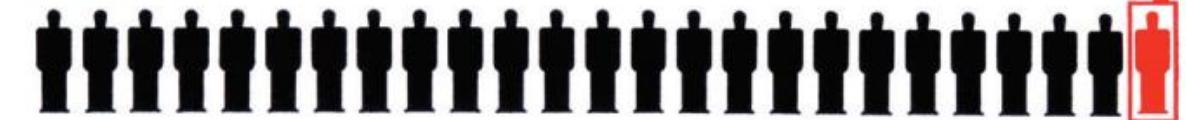
# *Isotype Charts*

- Isotype charts are a subclass of unit charts that use images or icons instead of simple shapes.
- The term Isotype—International System of Typographic Picture Education—was coined by German philosopher and political economist Otto Neurath, his wife Marie Neurath, and their collaborator Gerd Arntz in the 1920s.

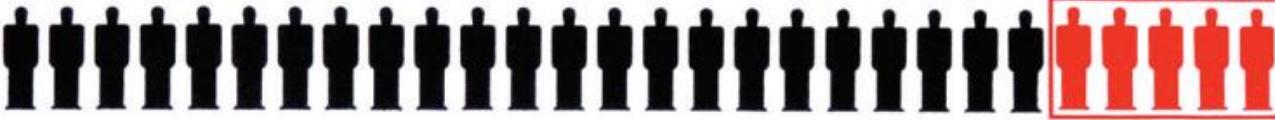
# *Isotype Charts*

## Home and Factory Weaving in England

1820



1830



1845



1860



1880



Each blue symbol represents 50 million pounds total production

Each black man symbol represents 10,000 home weavers

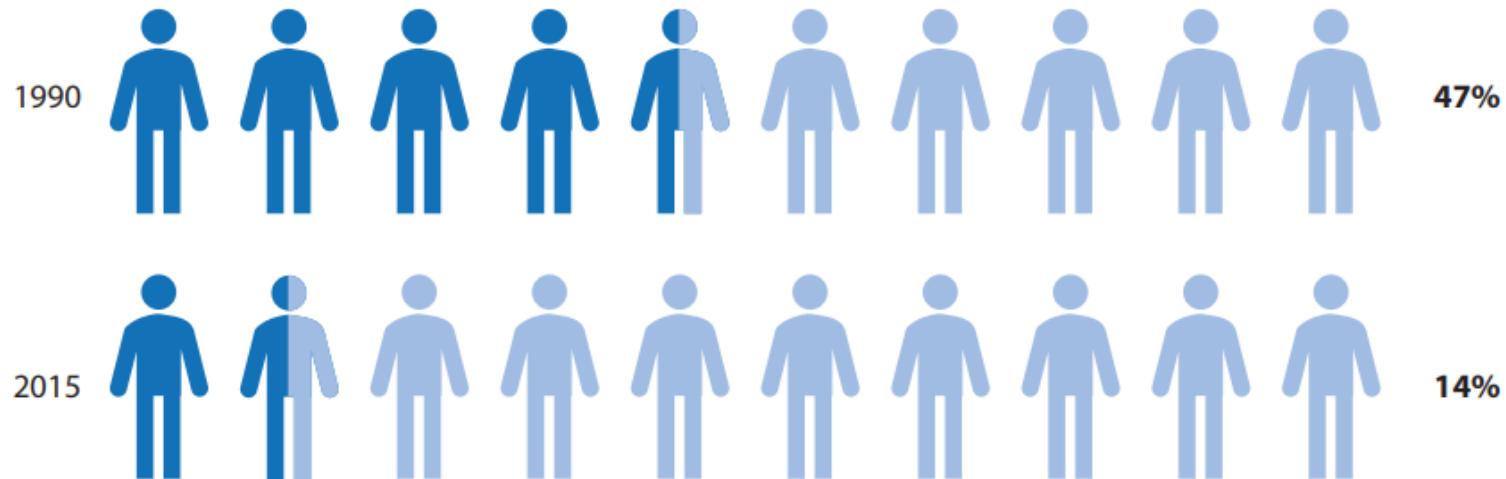
Each red man symbol represents 10,000 factory weavers

# Isotype Charts

We can take the same approach with the poverty data

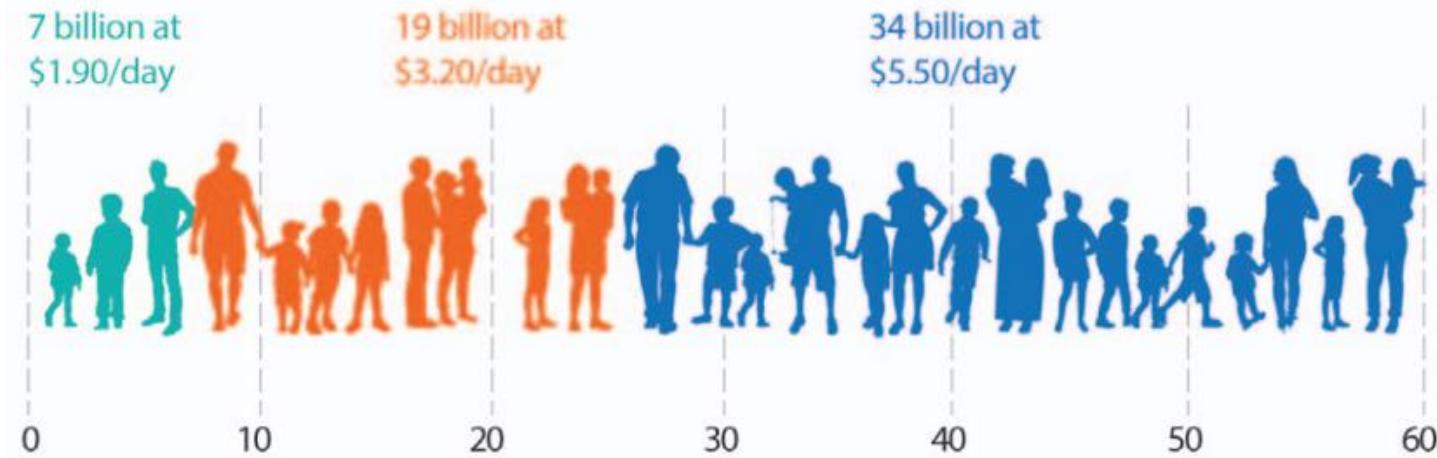
Two different ways to use icons in your data visualizations.

**Extreme poverty rate in developing countries**



Source: The World Bank

**Billions of people in poverty**

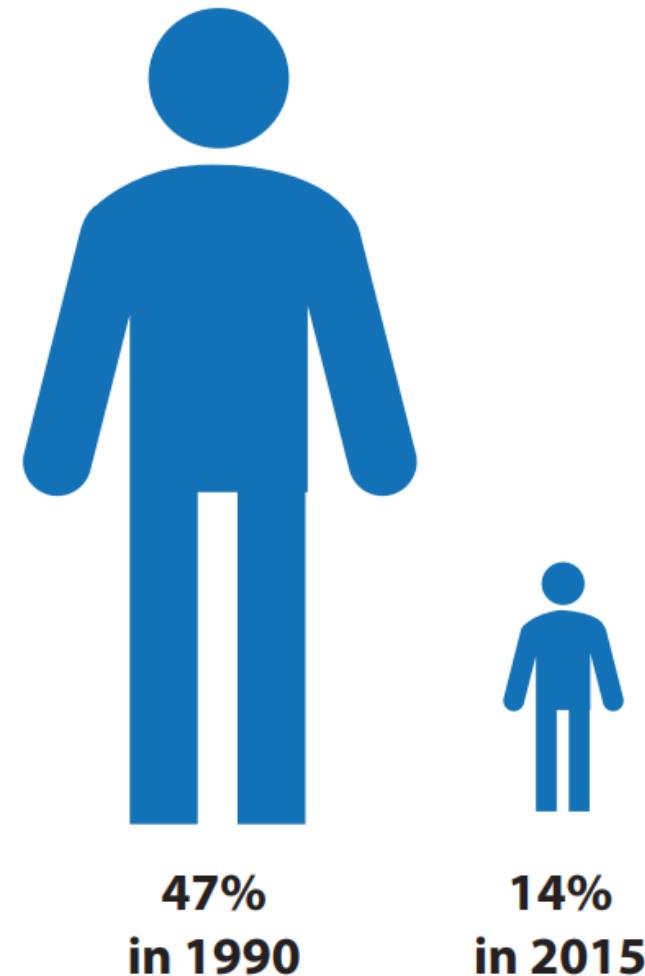


Source: The World Bank; silhouettes ("weepeople") from ProPublica and Alberto Cairo

# *Isotype Charts*

- Icons can also be scaled according to their data values.
- However it's hard to know whether they are scaled according to the height, width, or area.
- Here, the vertical distance represents the data values, but the area of the icon on the left is about 10 times the size as the icon on the right.
- Such charts are misleading because the area of the pictured man increases more rapidly than his height

## **Extreme poverty rate in developing countries**



Source: The World Bank

# *Waffle charts*

- Waffle charts are another subclass of unit charts. They are especially good for visualizing part-to-whole relationships.
- Waffle charts are arranged in a  $10 \times 10$  grid in which each colored cell represents one percentage point.
- You can use multiple waffle charts to show separate percentages—so the graph both shows part-to-whole relationships and lets your reader compare across the categories.

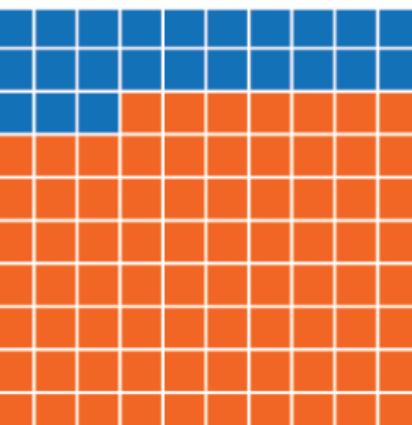
# *Waffle charts*

## **Overall poverty rate in five countries**

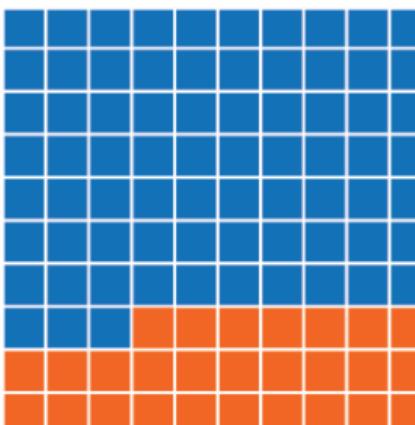
(Percent of people at \$1.90 per day)

Democratic Republic

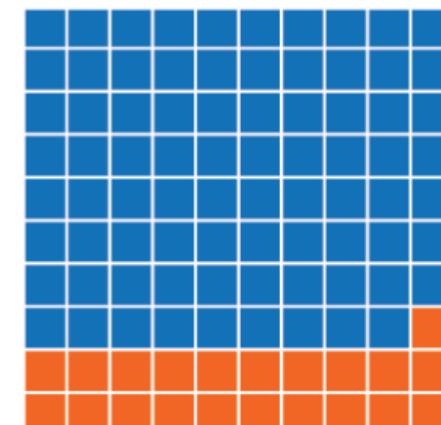
of the Congo | 77%    Nigeria | 54%



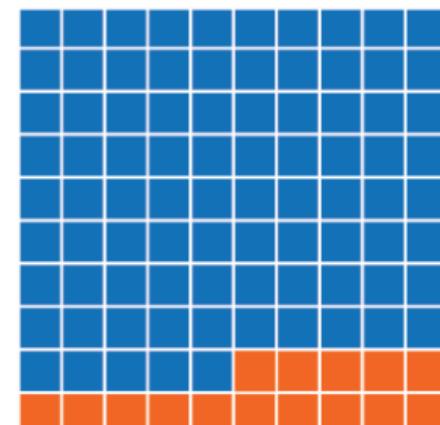
Ethiopia | 27%



India | 21%



Bangladesh | 15%



Source: The World Bank

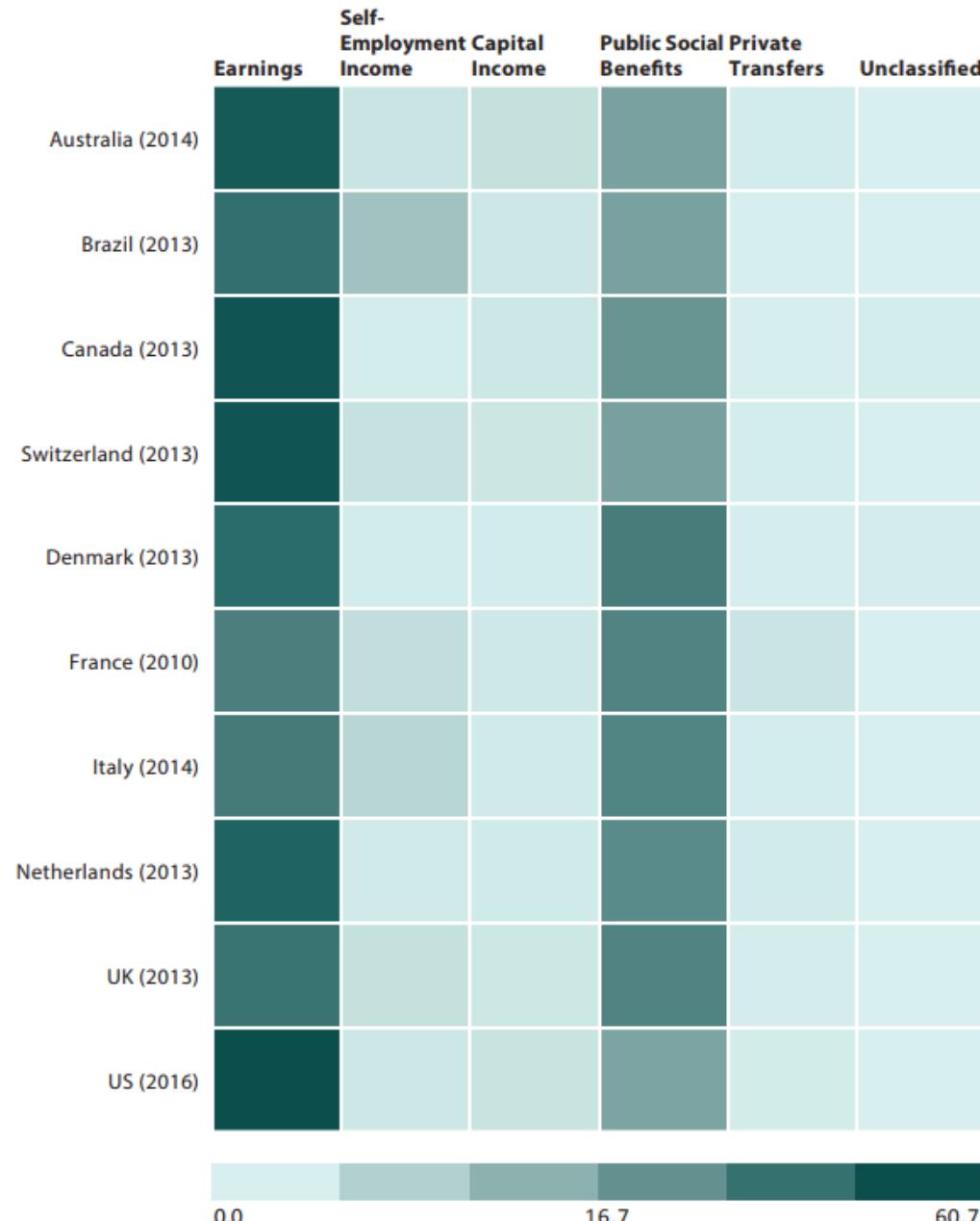
Waffle charts are a subclass of unit charts and, in this case, arrange the squares in a  $10 \times 10$  grid.

# *Heatmap*

- Heatmaps use colors and color saturations to represent data values.
- Simply put, a heatmap is a table with color-coded cells.
- They are often used to visualize high-frequency data or when seeing general patterns is more important than exact values.

## Composition of total income

(Percent of total income)

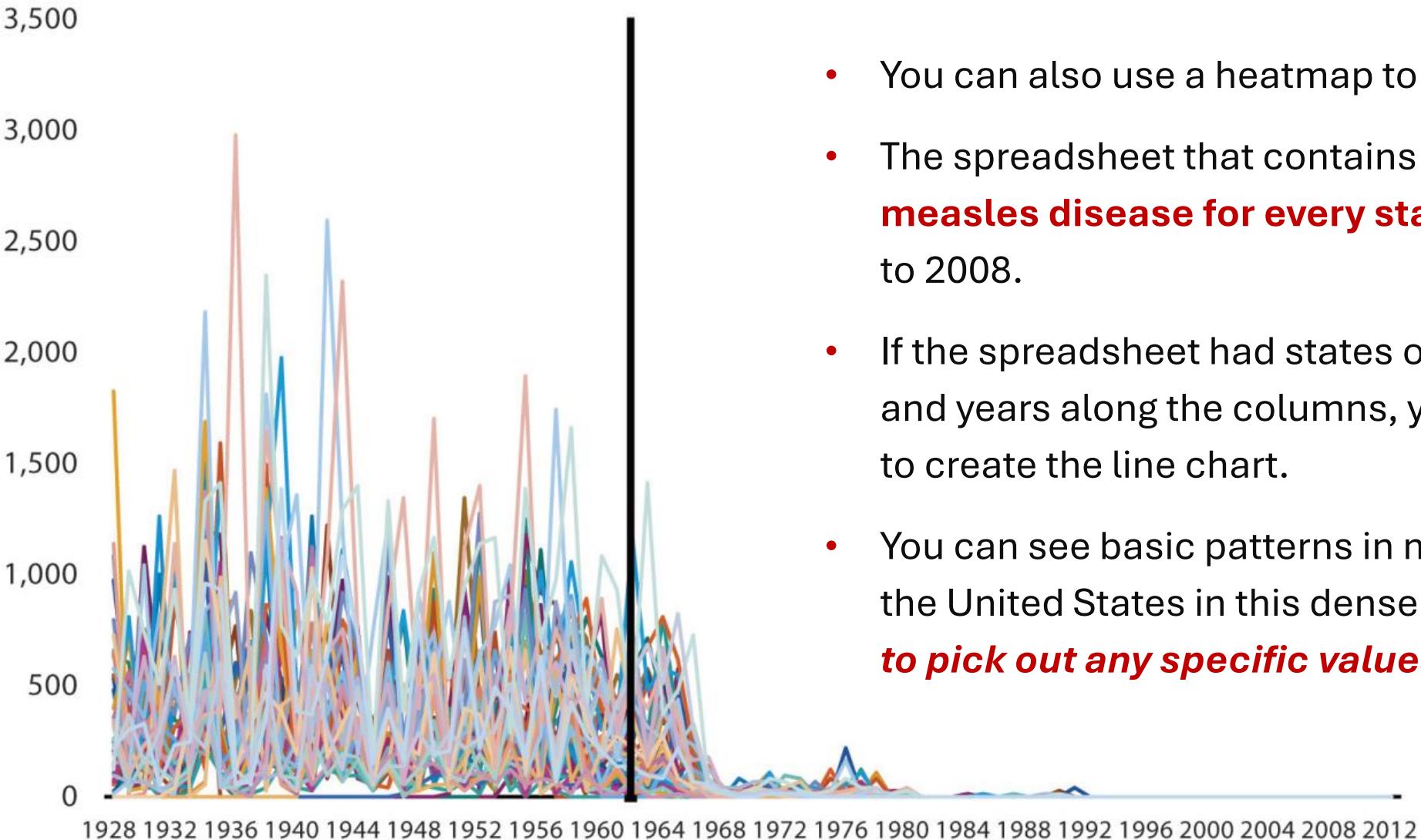


- Heatmaps use colors and color saturations to represent data values and can focus the reader's attention along the columns or across the rows.
- In this view, you can see that people's earnings account for the greatest share of their total income, and, in most countries, public social benefits appear to be the second-largest share.

- In this heatmap, each category is assigned its own color scale.
- Here, you can more clearly see that public social benefits (in the fourth column) account for a smaller share of total income in Australia, Brazil, Switzerland, and the United States.
- Which one is better depends on your goals.
- Do you want your reader to compare across all of the values or within each category?



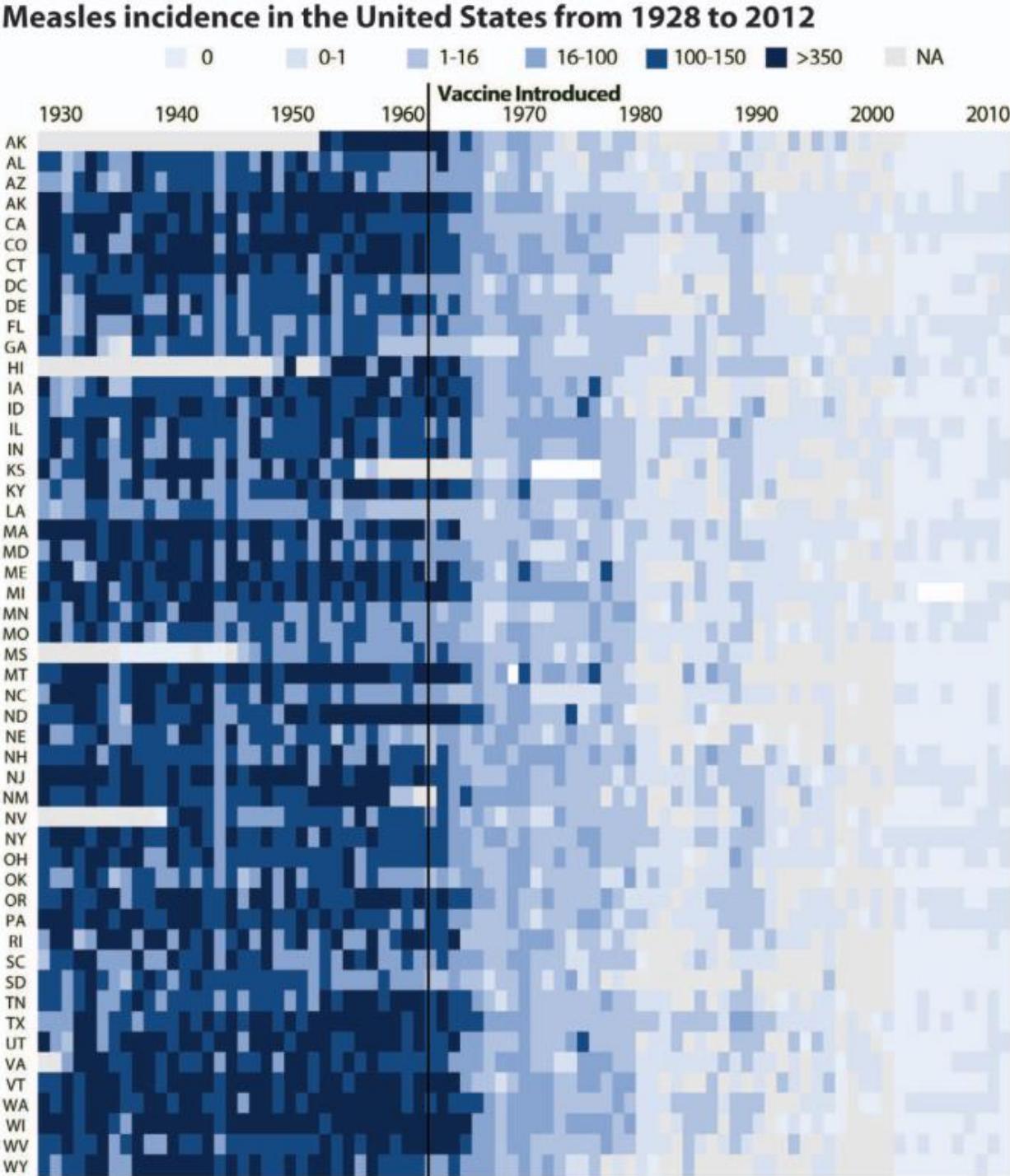
## Measles incidence in the United States from 1928 to 2012



- You can also use a heatmap to show changes over time.
- The spreadsheet that contains **infection rates from the measles disease for every state** in America from 1928 to 2008.
- If the spreadsheet had states ordered along the rows and years along the columns, your first instinct might be to create the line chart.
- You can see basic patterns in measles infections across the United States in this dense line chart, but **it's hard to pick out any specific values**.

Source: Data from Project Tycho, <https://www.tycho.pitt.edu/data>

- Instead of creating a dense line chart, let's create a heatmap.
- This heatmap may not be inherently better than the line chart at showing measles infections rates, but **it does let the reader more easily examine each state or year.**
- The heatmap, with its different look and color, can draw readers in.



- Another way to use the heatmap is to modify the layout, for example, applying it to a calendar.
- In this example, vehicle fatalities in 2015 are plotted on heatmaps of months.
- Notice how easy it is to see the higher fatality rate on Fridays and Saturdays along the right edge of each column

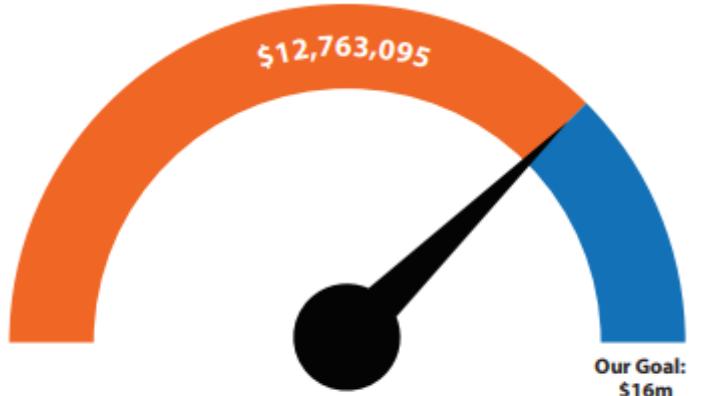
**Vehicle fatalities in 2015**



# *Gauge and bullet charts*

- The gauge chart (or gauge diagram) **looks like the speedometer** in your car's dashboard.
- Typically set up somewhere between **a half-circle and a circle, it uses a pointer or needle to indicate where your data fall** within a particular range.
- Sections of the gauge are shaded to illustrate sections such as poor, good, and excellent.
- We see **gauge diagrams most often in financial planning tools** because they give an easy, familiar way to visualize targets or progress towards a goal.
- They also **frequently show up in fundraising campaigns** where the entire semicircle represents the goal, and the needle and filled area represent money raised so far.

# Gauge charts

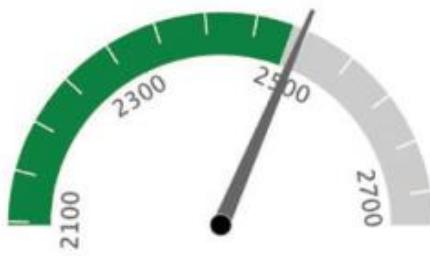


Gauge charts are familiar and easy to read.

This series of gauge charts shows four real estate trends in my Northern Virginia neighborhood.

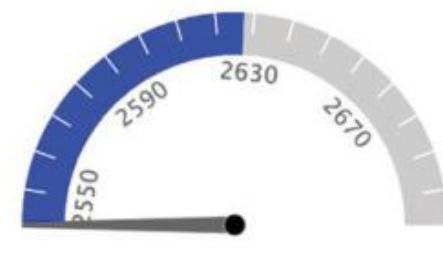
## NORTHERN VIRGINIA MARKET STATISTICS

Northern Virginia  
Closed Sales



© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia  
New Pensions



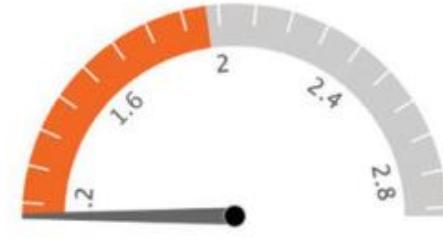
© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia  
Active Listings



© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

Northern Virginia  
Months of Supply



© 2019 MarketStats by ShowingTime. Data provided by Bright MLS as of Nov 6, 2019

# *Gauge charts*

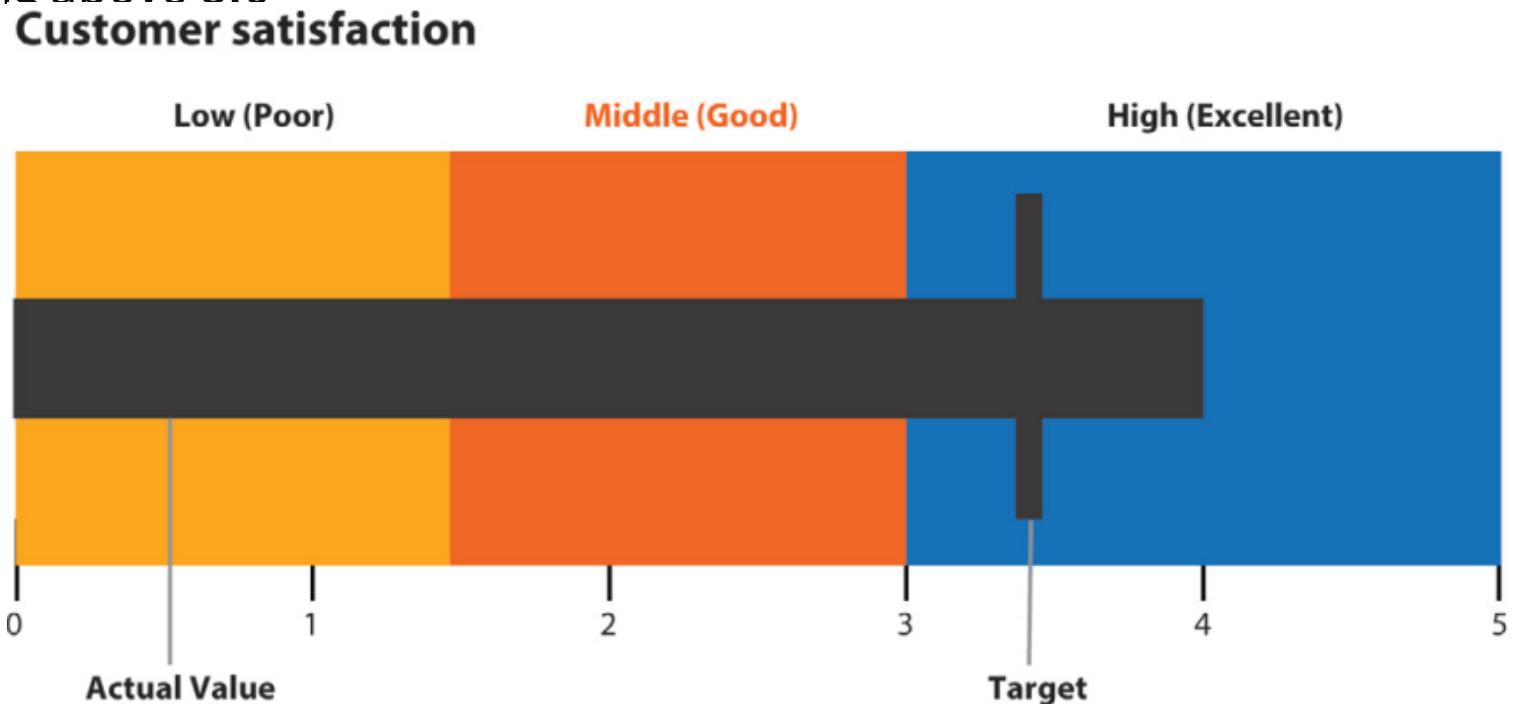
- Gauge diagrams do, however, introduce perceptual challenges because, again, people are not very good at measuring and comparing angles.
- If you want to give your reader a general sense of the values, the gauge chart is a decent choice.
- But if enabling your reader to discern the specific values and compare those values to the ranges is of utmost importance, then it is not.

# *Bullet charts*

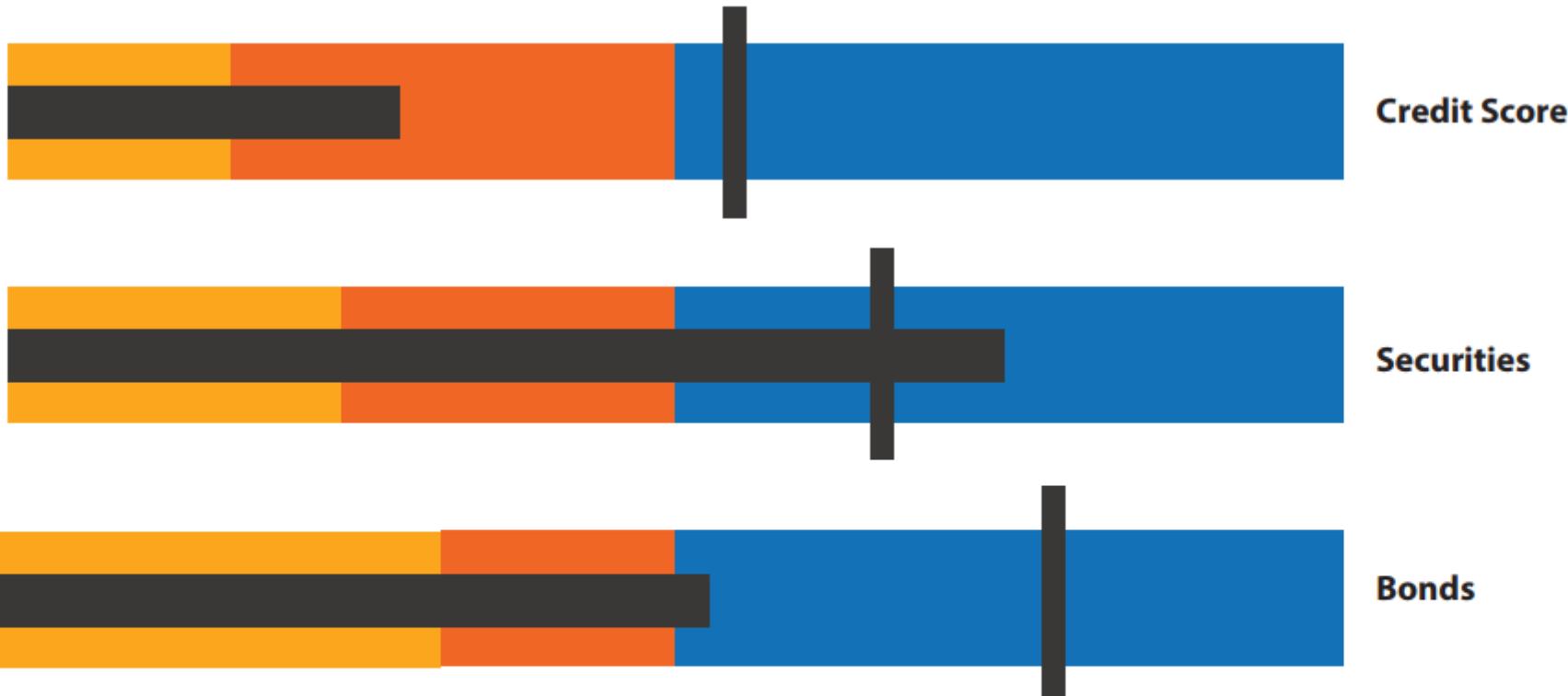
- Because of these perceptual challenges with the gauge chart, author Stephen Few invented the bullet chart, which is a linear, more compact way to show similar kinds of data.
- The basic bullet chart contains three different data elements:
  1. First, there is the actual or observed value, shown here as the black horizontal bar.
  2. Second, there is a target value, shown here as the black vertical line.
  3. Finally, there is the background range, which shows grades or bands of success, such as poor, good, and excellent.

# Bullet charts

1. **Actual:** In this illustration, the bar represents an average customer satisfaction score of 4.0.
2. **Target:** Here, we were aiming for a satisfaction score of 3.5.
3. **Background Range:** These sit behind the other two series so the reader can compare the actual and target values. Here, poor scores are 1.5 and below, good scores are from 1.5 to 3.0, and excellent scores are anything above 3.0



# *Bullet charts*



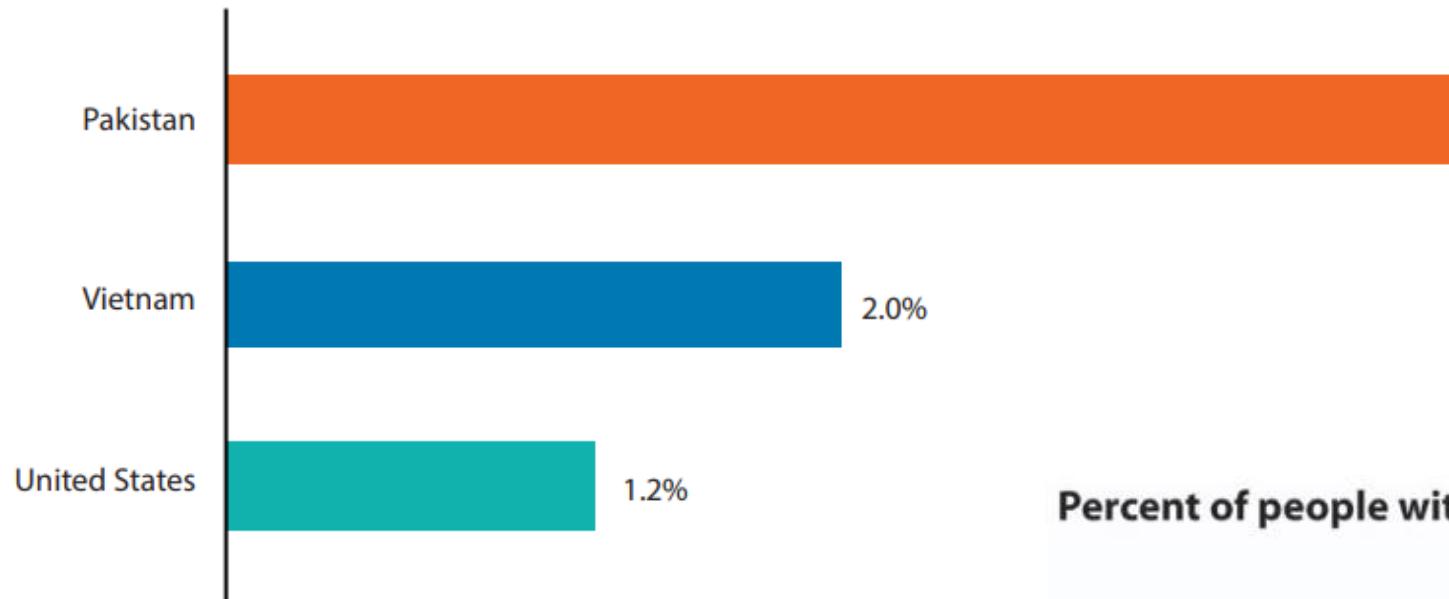
---

Combining different bullet charts is a compact way to let the reader make a series of different comparisons.

# *Bubble comparison and Nested Bubbles*

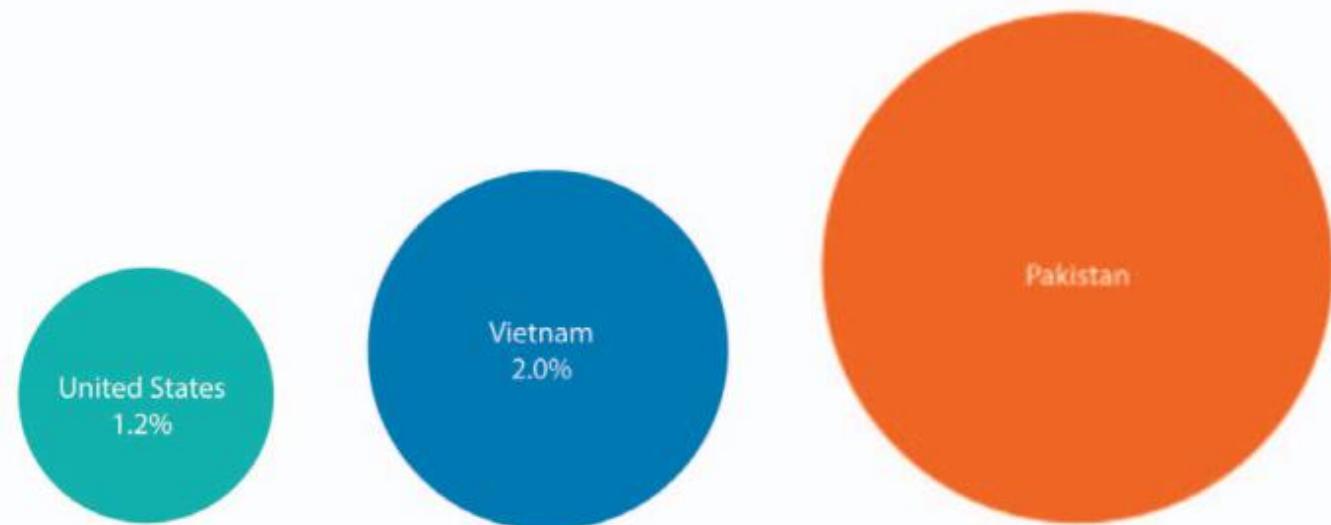
- In the basic bubble graph, circles represent values.
- Unlike the bar chart, humans are not very good at accurately comparing the sizes of circles.
- Still, circles may be more visually interesting, they can reinforce a visual or metaphor, and they are a good choice when discerning exact quantities is not paramount.
- Another drawback of proportionally sized circles is that you cannot visualize negative values.
- We are not very good at making accurate estimates from the circles even when they are sized by area.

## Percent of people with less than \$1.90 a day in 2011



Source: The World Bank

## Percent of people with less than \$1.90 a day in 2011



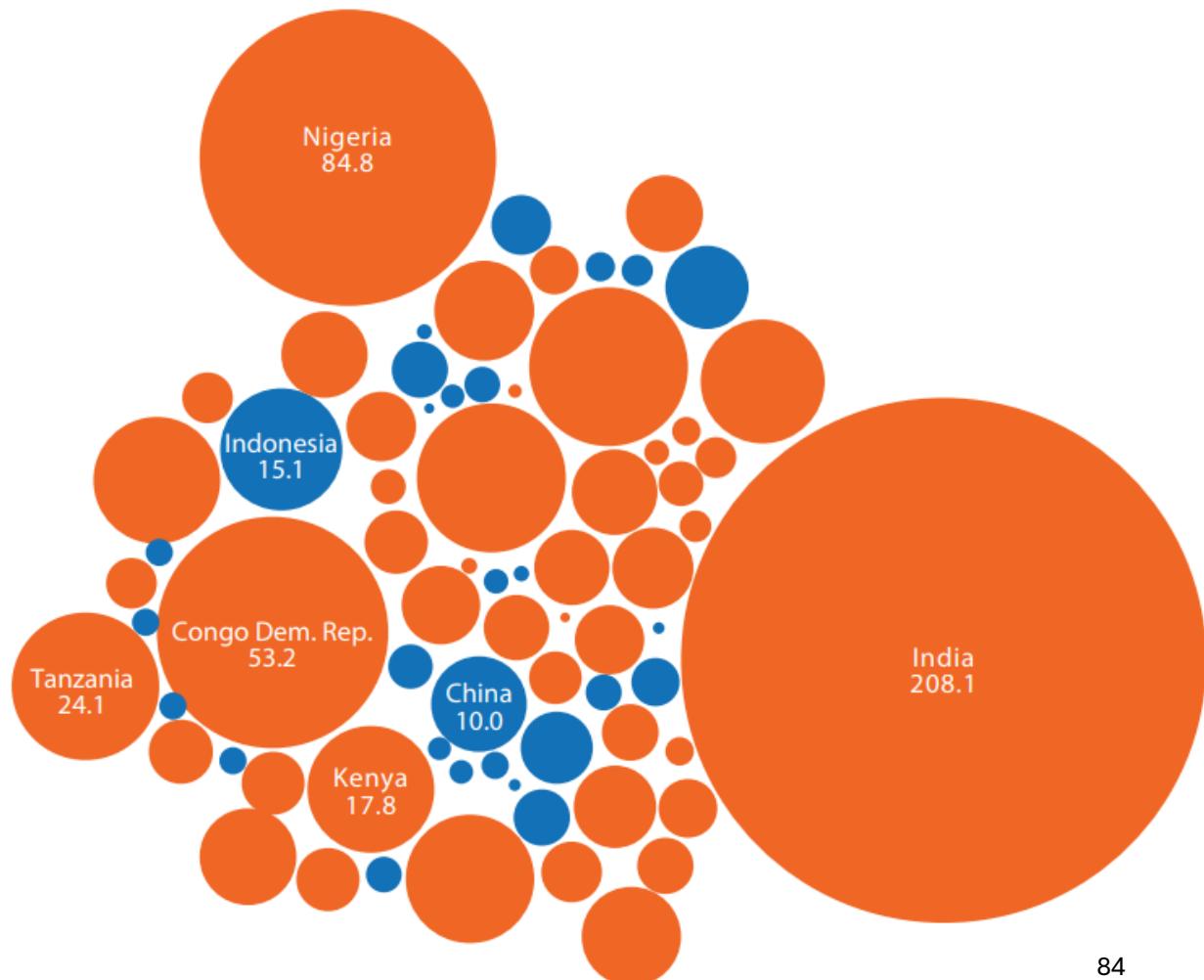
Source: The World Bank

# *Bubble comparison charts*

- We can see that India, the Democratic Republic of the Congo, and Nigeria have the largest number of people in poverty,
- It's difficult to quickly assess how different they are or the numbers of the next set of countries.

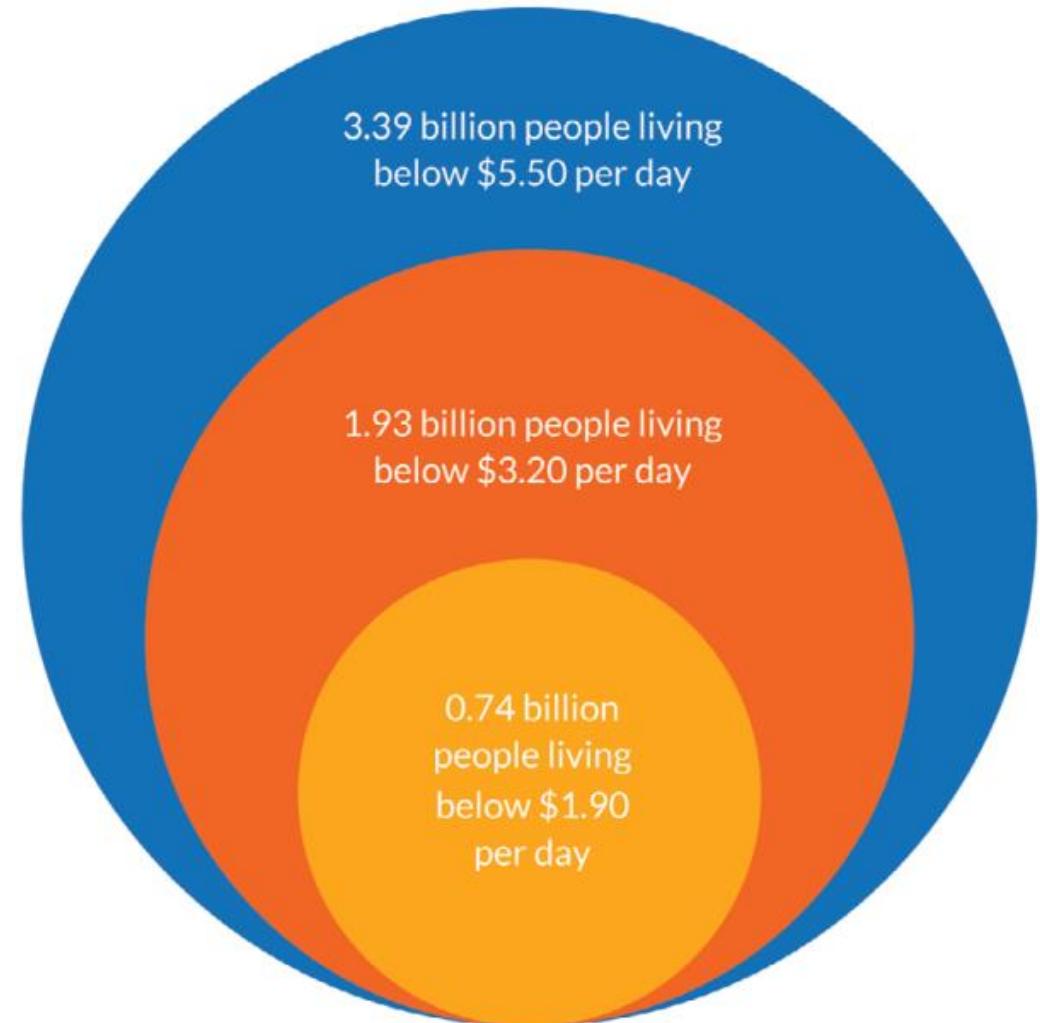
**Number of people in poverty**

(Orange circles: poverty rate > 14.5 percent; Blue circles: poverty rate < 14.5 percent)



# *Nested bubble*

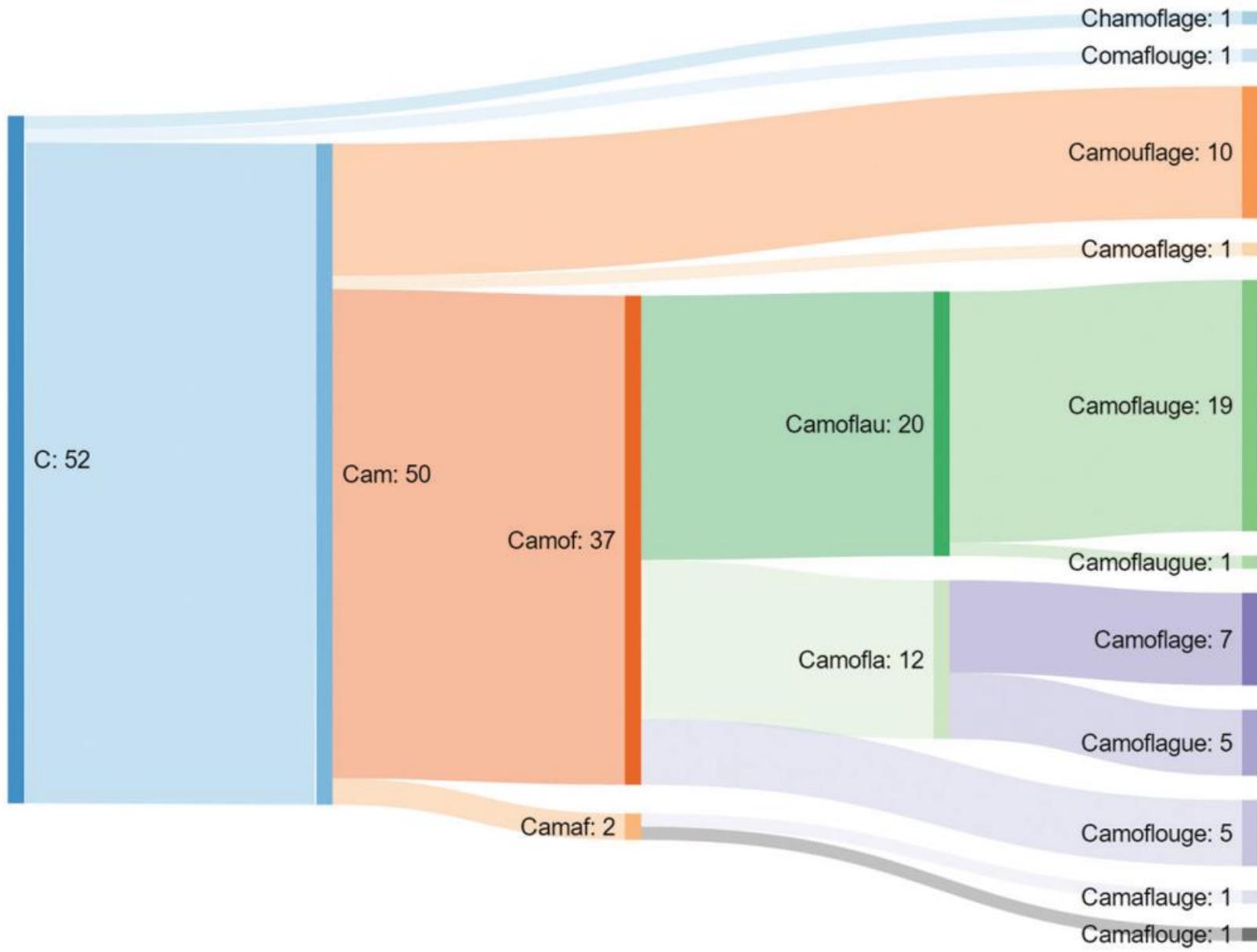
- Layering circles on top of each other, are often called nested bubble charts.
- The nested bubble chart can sometimes mask circles in the back, but it can also make for easier comparisons.



# *Sankey diagram*

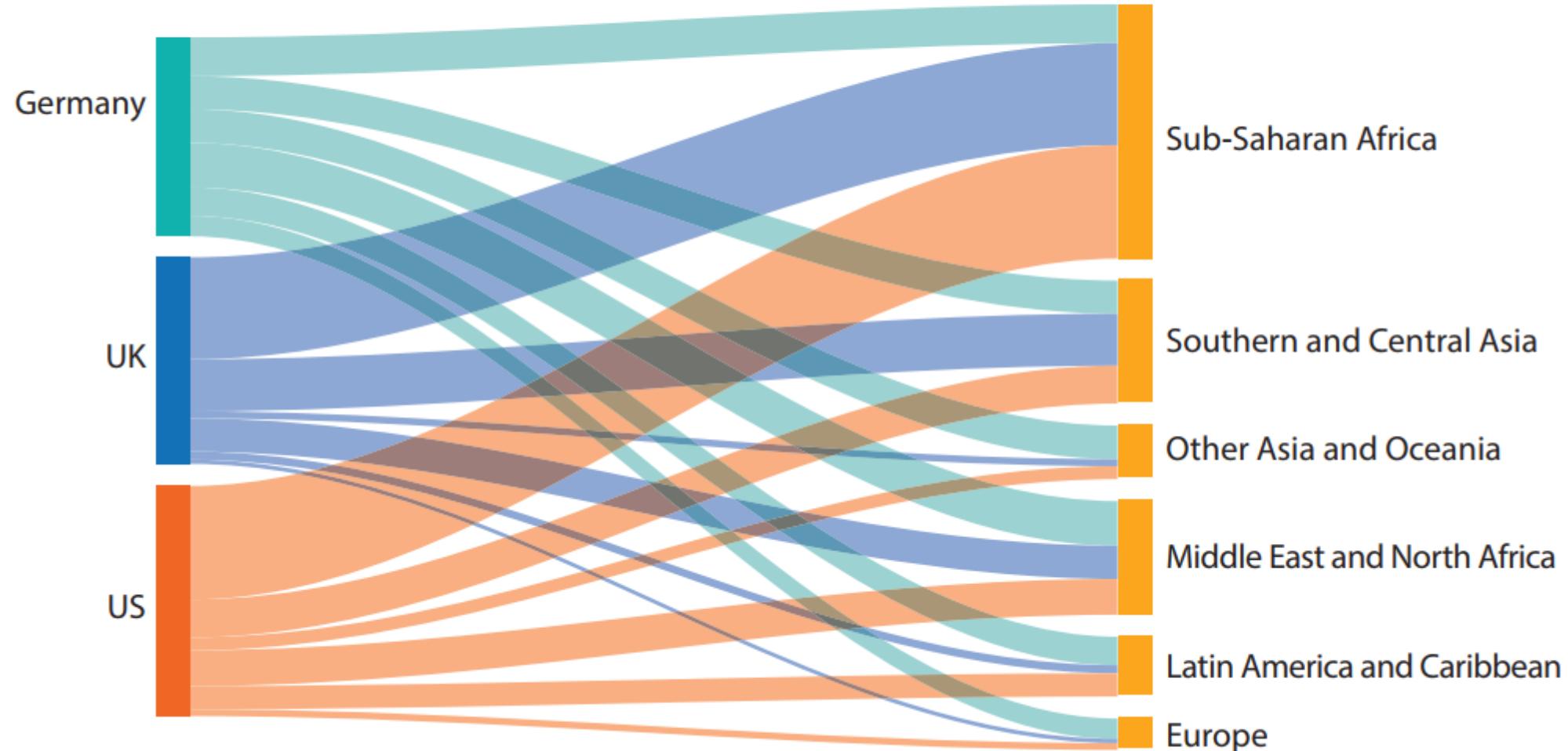
- Sankey diagrams—named for their creator Matthew Henry Phineas Riall Sankey in 1898—are especially useful for showing how categories compare to one another and flow into other states or categories.
- Arrows or lines display the transition from one state to another, and the width of the lines denote the magnitude of each transition.
- Changes can occur over time or as comparisons between categories.

- It shows how fifty-two students tried to spell the word camouflage.
- The first blue segment shows that all fifty-two students started with the letter “C”,
- Fifty then went to “Cam”, followed by thirty-seven with “Camof ”, and so on.
- Ten students spelled the word correctly, shown in the orange segment near the top of the graph.



# Financial support flows from Germany, the United States, and the United Kingdom to different areas of the world

(Percent of total support)



# *Sankey diagram vs Paired Bars*

**Financial support flows from Germany, the United States, and the United Kingdom to different areas of the world**

(Percent of total support)

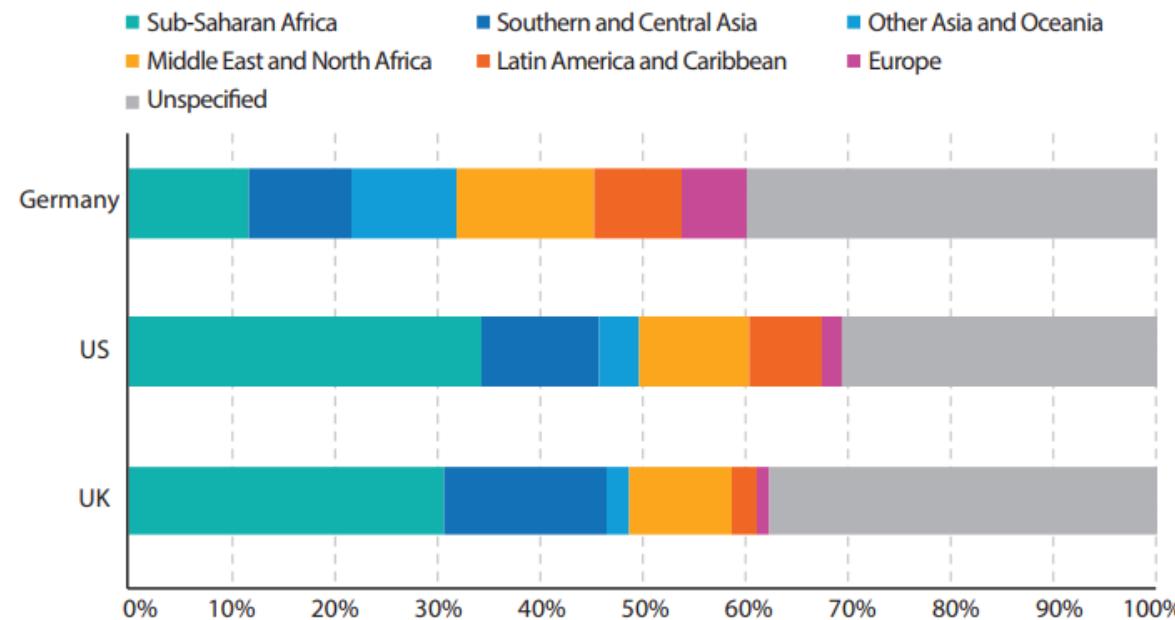


- Presenting these data as paired or stacked bar charts give us a different perspective.
- In the paired bar chart on the left, the obvious first comparison is across the funder countries for each region of the world.

# *Sankey diagram vs Stacked bars*

**Financial support flows from Germany, the United States, and the United Kingdom to different areas of the world**

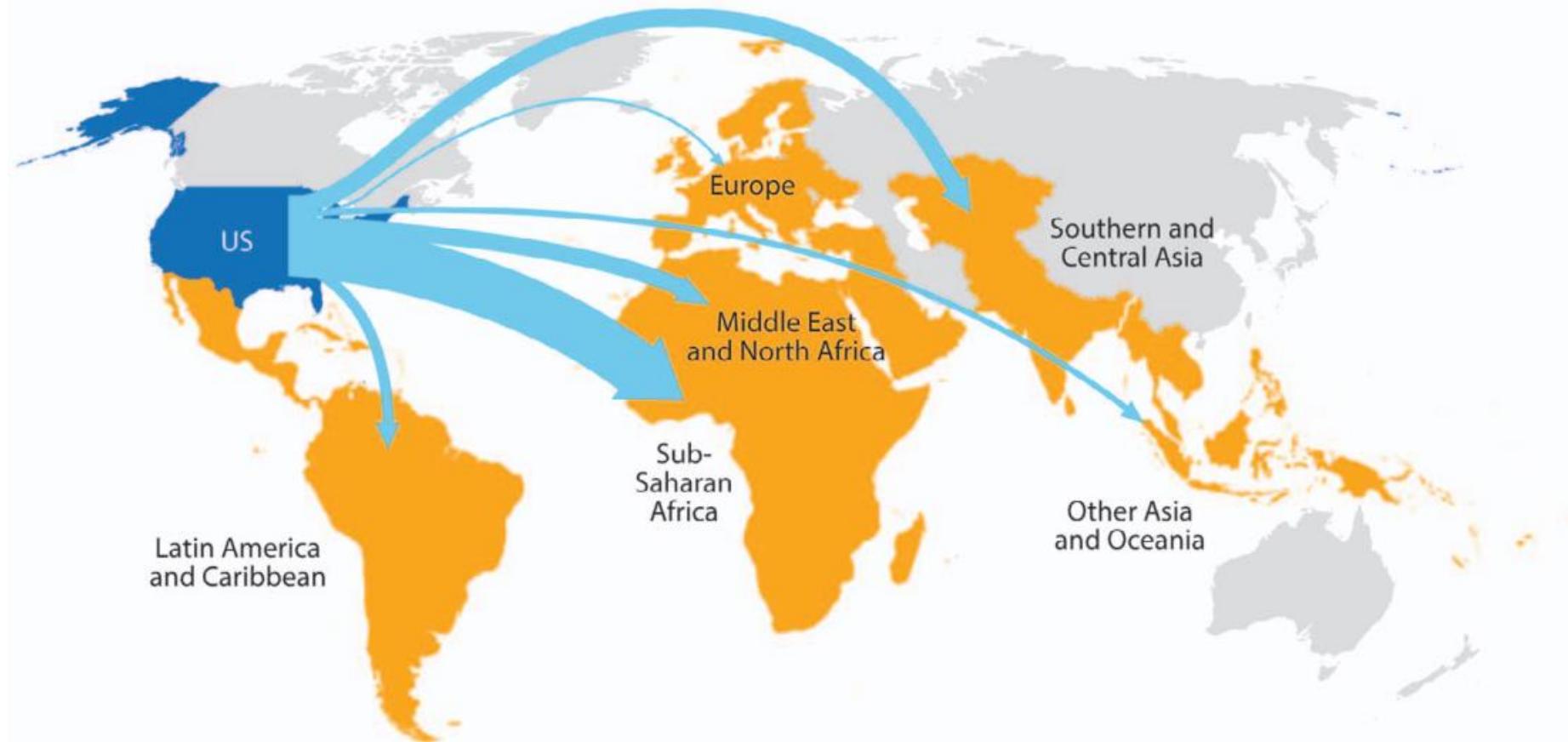
(Percent of total support)



- In the stacked bar chart, by comparison you're more likely to compare funding across the recipient regions—that, for example, the greatest share of spending on Sub-Saharan African countries is from the United States.

# **Financial support flows from Germany, the United States, and the United Kingdom to different areas of the world**

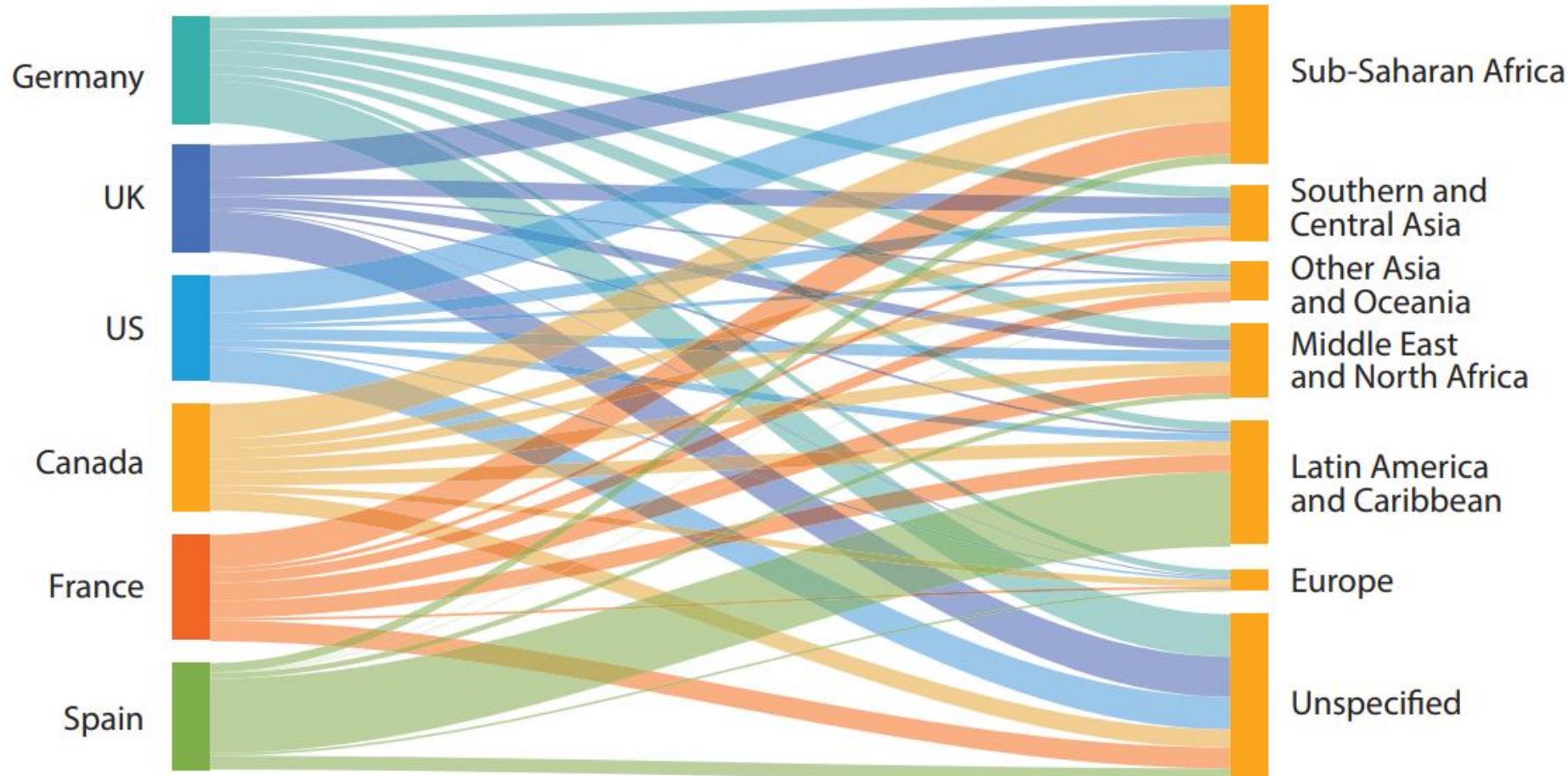
(Percent of total support)



This flow map provides a geographic view of the financial flow data, but it also simplifies things by only showing flows from the United States.

## Financial support flows from Germany, the United States, and the United Kingdom to different areas of the world

(Percent of total support)



The biggest problem with Sankey diagrams is plotting too many series makes it difficult to identify any patterns or trends.

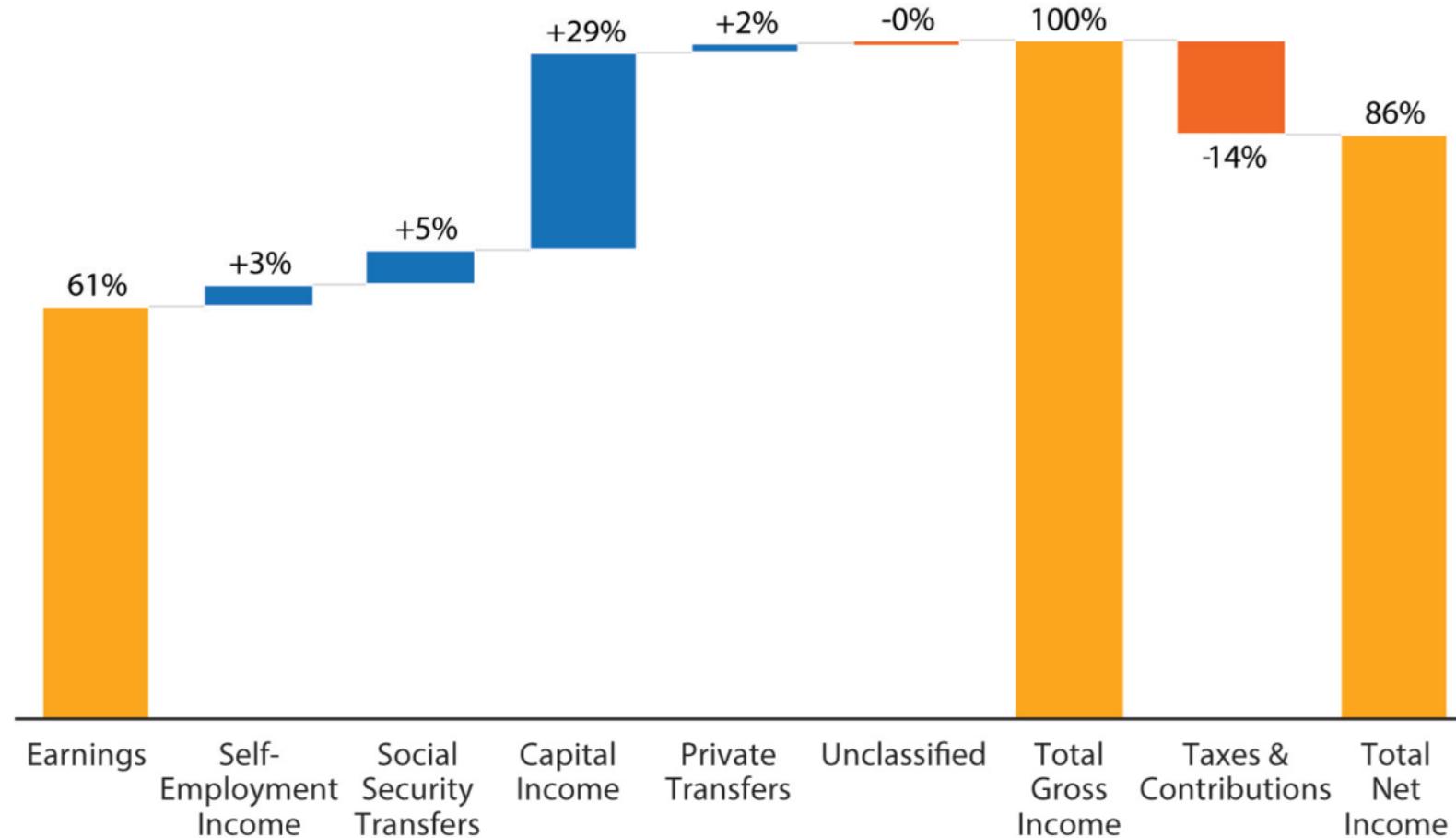
# *Waterfall chart*

- A waterfall chart shows a basic mathematical equation: adding or subtracting values from some initial value to produce a final amount.
- It is essentially a bar chart, but each subsequent bar starts where the previous one left off, showing how they accumulate across the graph.
- Typically, negative values are given a different color than positive values, and so are the totals at the beginning and end. Including lines that connect the bars can guide the reader through the visualization.
- Because the lines are guides and not actual data, they should be lighter and thinner than the other elements.

# Waterfall chart

**Income composition in Australia in 2016**

(Percent of total gross income)



A waterfall chart shows a basic mathematical equation: adding or subtracting values from some initial value to produce a final amount.

# *Conclusion*

---

# *Making Comparison - Conclusion*

- From single bars to groups of bars to stacks of bars, the bar chart is one of the most familiar data visualizations for showing categorical comparisons.
- It also ranks at the top of our perceptual ranking scale from earlier.
- But the bar chart also poses certain challenges: too many bars can make the visual seem overwhelming and cluttered, and stacking the series on top of one another makes it more difficult to compare series that are not aligned on the same axis.

# *Making Comparison - Conclusion*

- While bar charts sit at the top of the perceptual ranking list – They can be very boring. We see bar charts every day.
- There are other ways to let your reader make comparisons.
- Dot plots especially are very appealing because they remove a lot of the heavy ink from a standard bar chart and free up space to add annotation and labels.
- Using icons, squares, or other shapes can engage our audience in ways that standard charts may not, but may be less data dense.

# *Reference*

- Better Data Visualizations A Guide for Scholars, Researchers, and Wonks Jonathan Schwabish - Columbia University Press, 2021