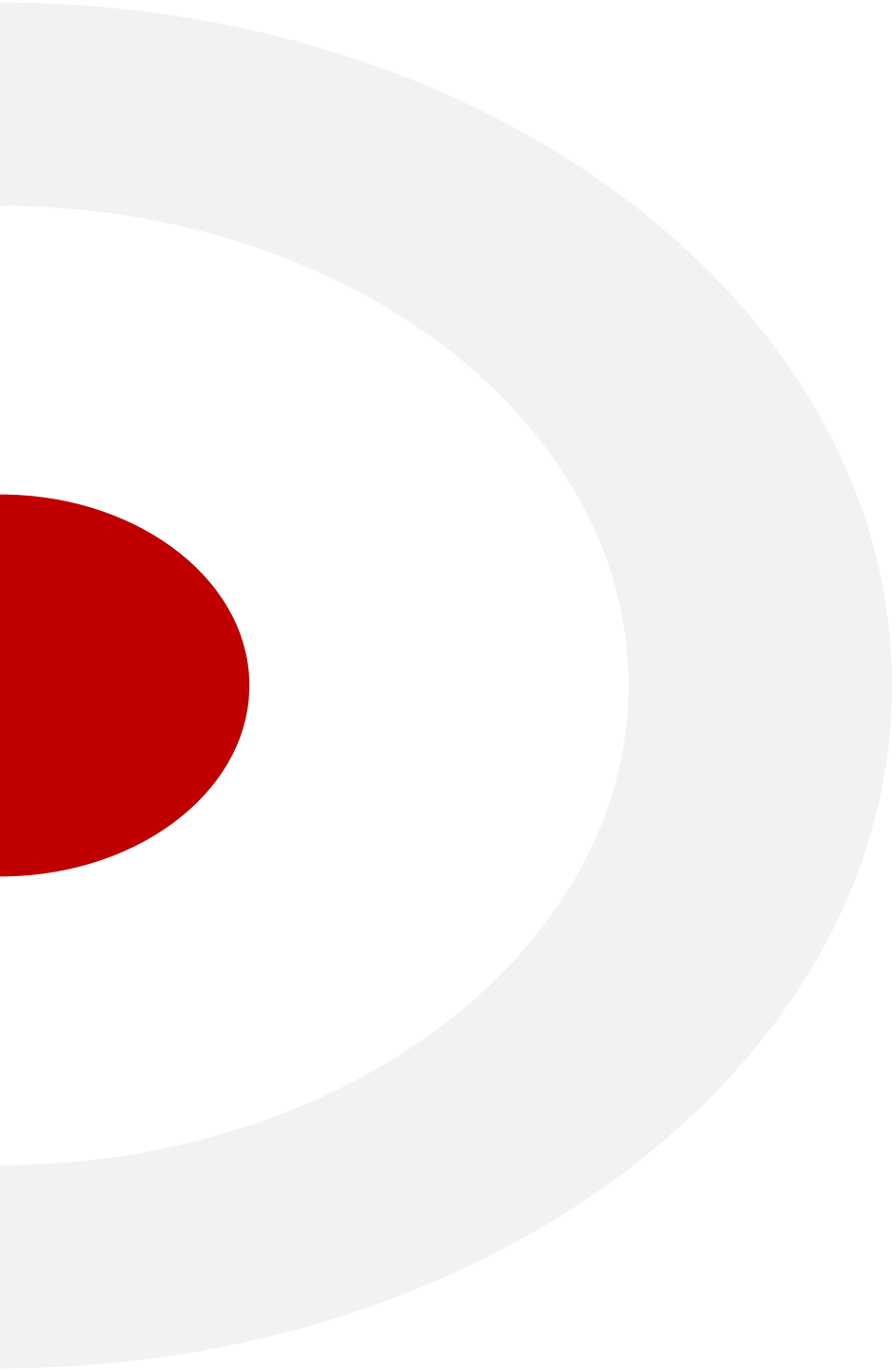




DATA PREPARATION AND VISUALIZATION

Mathematical Economics Faculty

National Economics University
<https://www.neu.edu.vn/>



Chapter 12: Plotting and Visualization

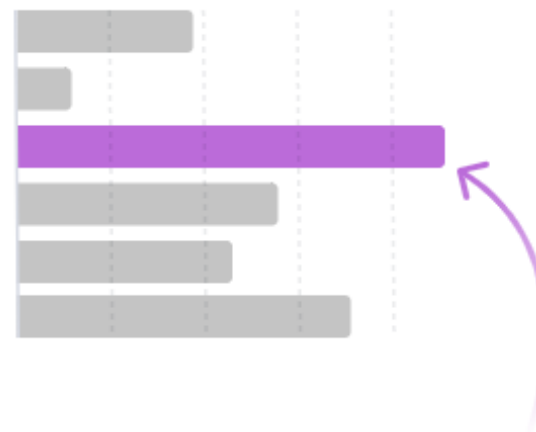
— Introduction

Exploratory
data visualization



We build graphs for ourselves
to explore data and find
patterns.

Explanatory
data visualization



We build graphs for others
to communicate and explain
the patterns we've found
through exploring data.

— The importance of context

When it comes to explanatory analysis, there are a few things to think about and be extremely clear on before visualizing any data

1. To whom are you communicating?
2. What do you want your audience to know or do?
3. How can you use data to help make your point?

— The importance of context

Example

Who, what, and how: illustrated by example

Let's consider a specific example to illustrate these concepts. Imagine you are a fourth grade science teacher. You just wrapped up an experimental pilot summer learning program on science that was aimed at giving kids exposure to the unpopular subject. You surveyed the children at the onset and end of the program to understand whether and how perceptions toward science changed. You believe the data shows a great success story. You would like to continue to offer the summer learning program on science going forward.

— The importance of context

Example

- **Who:** The budget committee that can approve funding for continuation of the summer learning program
- **What:** The summer learning program on science was a success; please approve budget of \$X to continue
- **How:** Illustrate success with data collected through the survey conducted before and after the pilot program

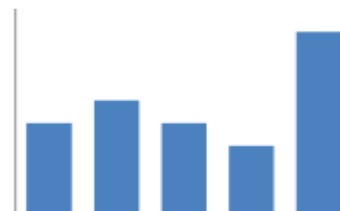
Choosing an effective visual

91%

Simple text



Scatterplot



Vertical bar



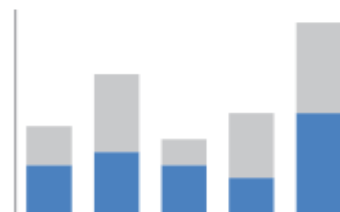
Horizontal bar

| | A | B | C |
|------------|-----|-----|-----|
| Category 1 | 15% | 22% | 42% |
| Category 2 | 40% | 36% | 20% |
| Category 3 | 35% | 17% | 34% |
| Category 4 | 30% | 29% | 26% |
| Category 5 | 55% | 30% | 58% |
| Category 6 | 11% | 25% | 49% |

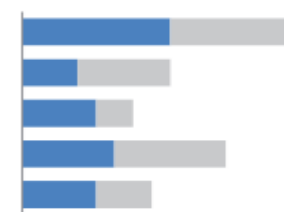
Table



Line



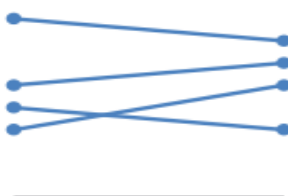
Stacked vertical bar



Stacked horizontal bar

| | A | B | C |
|------------|-----|-----|-----|
| Category 1 | 15% | 22% | 42% |
| Category 2 | 40% | 36% | 20% |
| Category 3 | 35% | 17% | 34% |
| Category 4 | 30% | 29% | 26% |
| Category 5 | 55% | 30% | 58% |
| Category 6 | 11% | 25% | 49% |

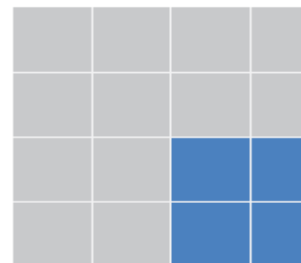
Heatmap



Slopegraph



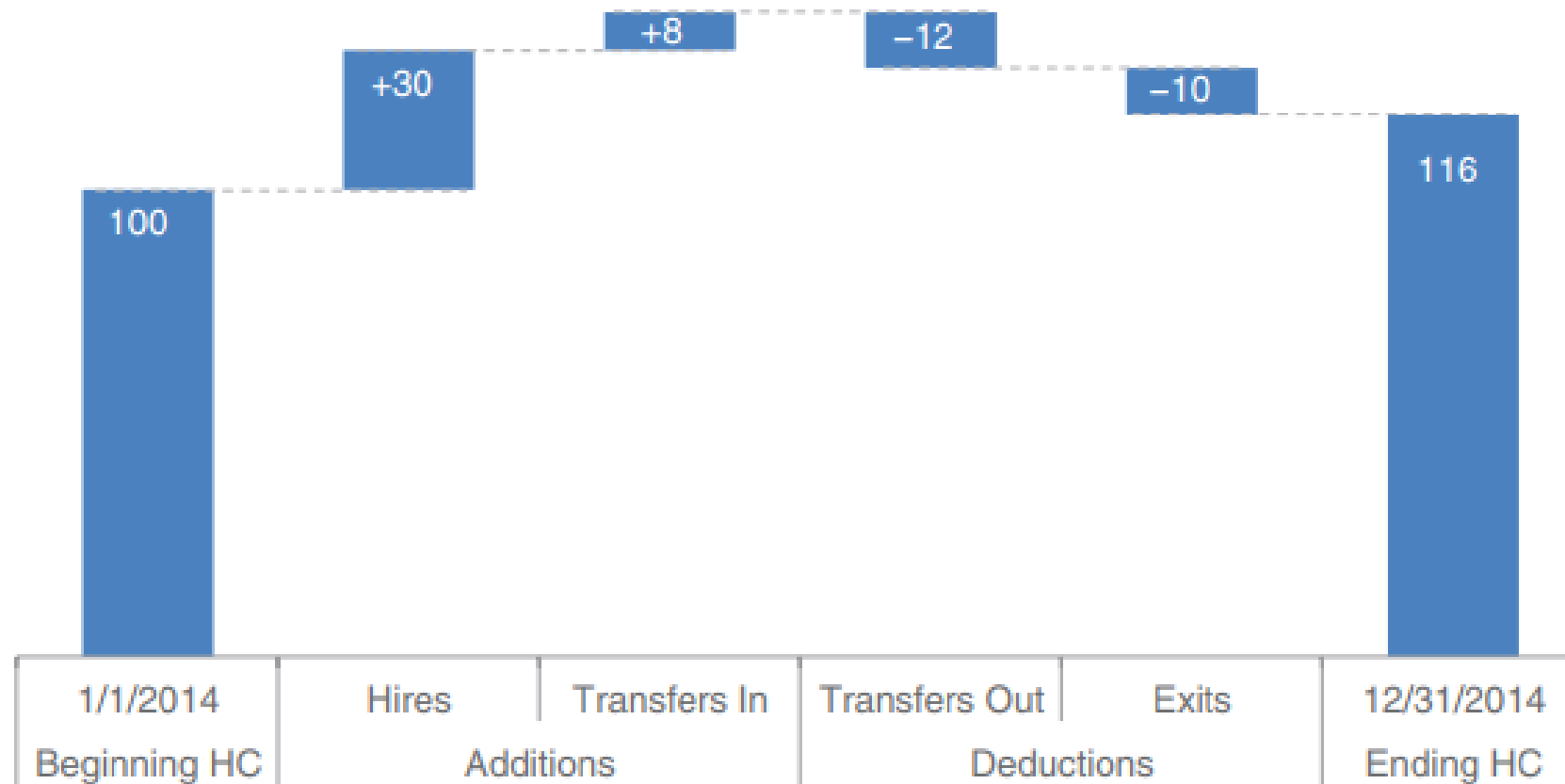
Waterfall



Square area

FIGURE 2.1 The visuals I use most

Waterfall chart - Example

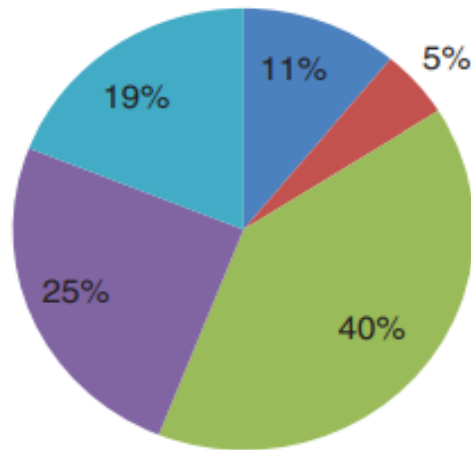


— Choosing an effective visual example

Survey results: summer learning program on science

PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

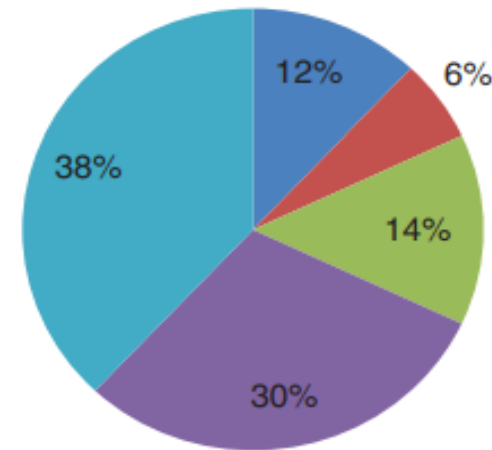


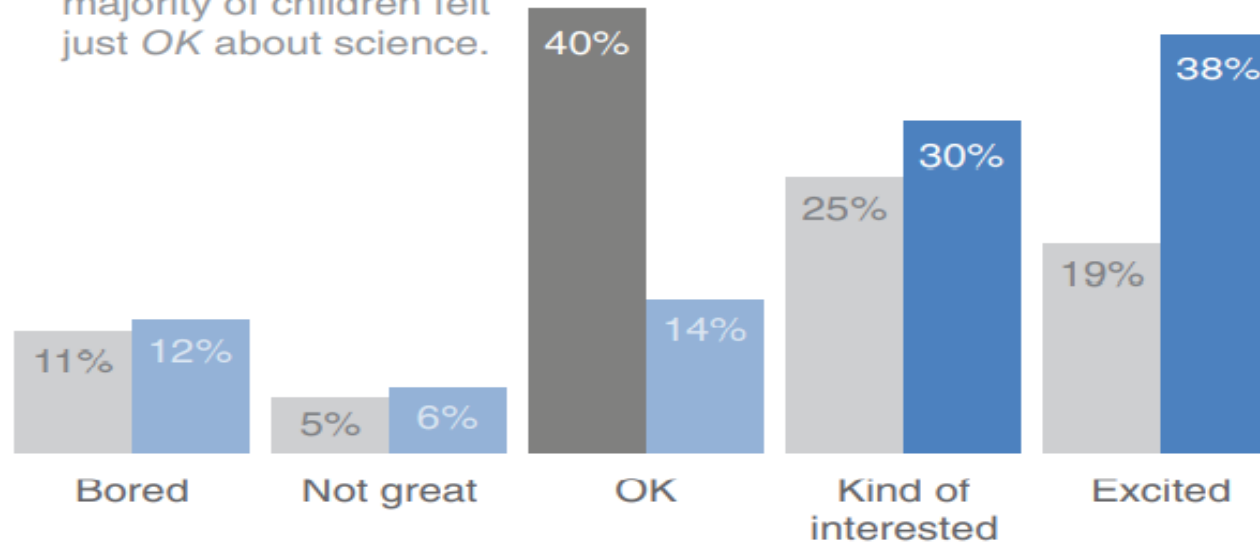
FIGURE 9.28 Original visual

— Choosing an effective visual example

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.



AFTER program, more children were *Kind of interested & Excited* about science.

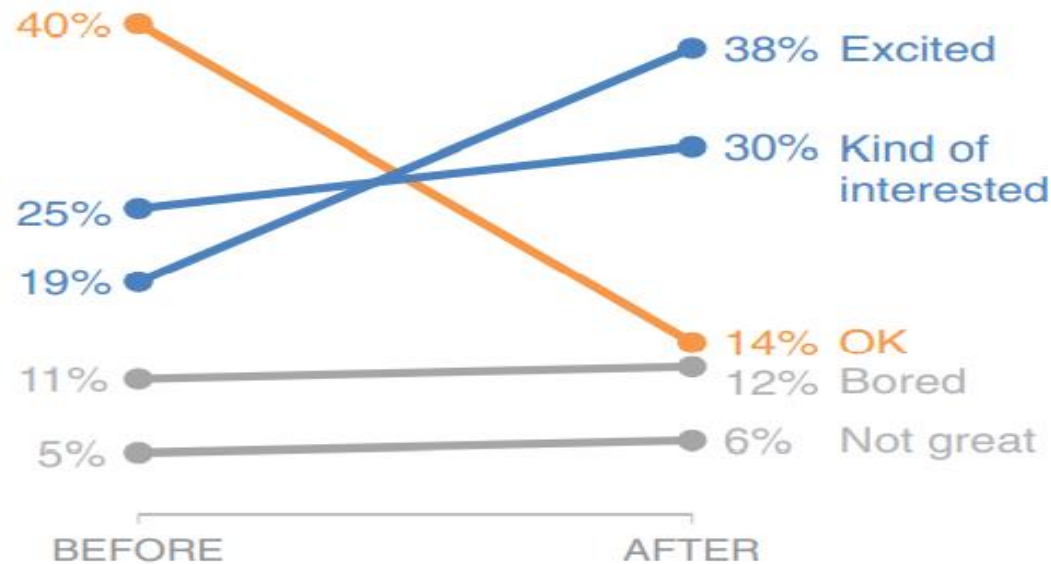
Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.30 Simple bar graph

— Choosing an effective visual example

Pilot program was a success

How do you feel about science?



BEFORE program, the majority of children felt just *OK* about science.

AFTER program, more children were *Kind of interested & Excited* about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.32 Slopegraph

— Focus your audience's attention

Gestalt principles of visual perception

- Gestalt Principles help you identify which elements in our visuals are signals, which might be noise
- The six principles: proximity, similarity, enclosure, closure, continuity and connection



FIGURE 3.1 Gestalt principle of proximity

Gestalt principles of visual perception

Proximity

- We can leverage this way that people see in **table design**
- In figure 3.2, your eyes are drawn either down the columns in the first case or across the rows in the second case



FIGURE 3.2 You see columns and rows, simply due to dot spacing

Gestalt principles of visual perception

Proximity

- We can leverage this way that people see in table design

Table

| | A | B | C |
|------------|-----|-----|-----|
| Category 1 | 15% | 22% | 42% |
| Category 2 | 40% | 36% | 20% |
| Category 3 | 35% | 17% | 34% |
| Category 4 | 30% | 29% | 26% |
| Category 5 | 55% | 30% | 58% |
| Category 6 | 11% | 25% | 49% |

Gestalt principles of visual perception

Similarity

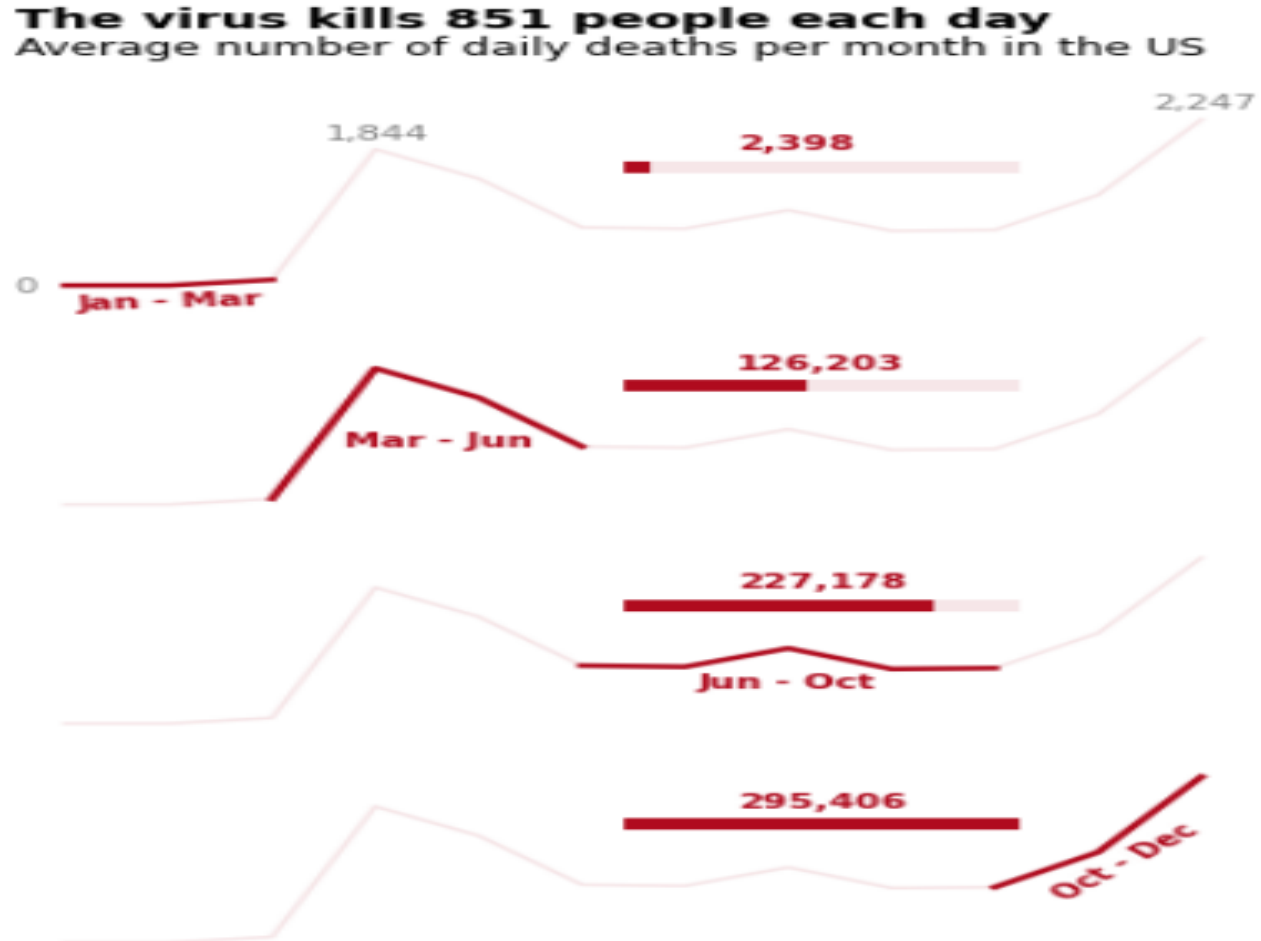
- Objects that are of similar color, shape, size or orientation are perceived as related or belonging to part of a group



FIGURE 3.3 Gestalt principle of similarity

Gestalt principles of visual perception

Similarity



Gestalt principles of visual perception

Enclosure

- We think of objects that are physically enclosed together as belonging to part of a group



FIGURE 3.5 Gestalt principle of enclosure

Gestalt principles of visual perception

Enclosure

- Enclosure comes in handy when we want to **separate or draw attention** to certain portions of a graph
- Ex: figure 3.6

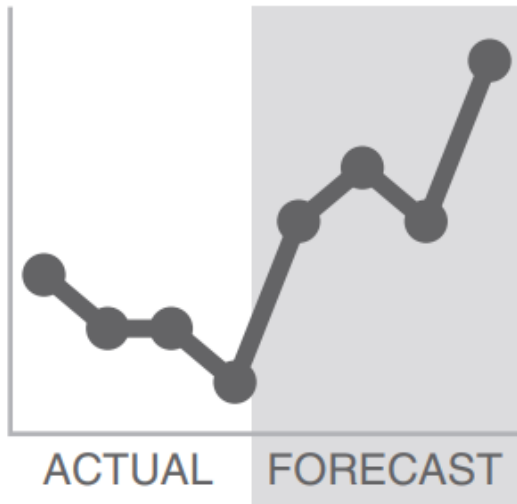


FIGURE 3.6 The shaded area separates the forecast from actual data

Gestalt principles of visual perception

Connection

- The connective property typically has a stronger associative value than similar color, size or shape

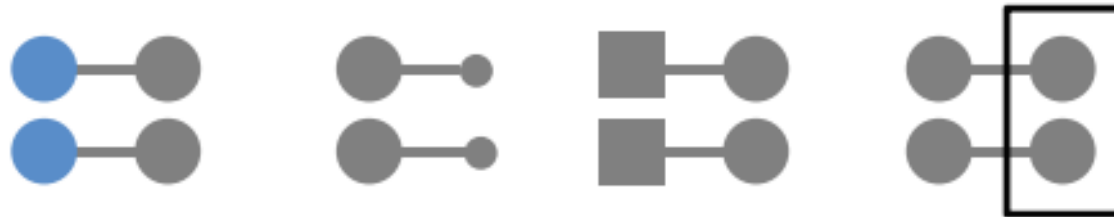


FIGURE 3.11 Gestalt principle of connection

Gestalt principles of visual perception

Connection - example

One way that we frequently leverage the connection principle is in line graphs, to help our eyes see order in the data

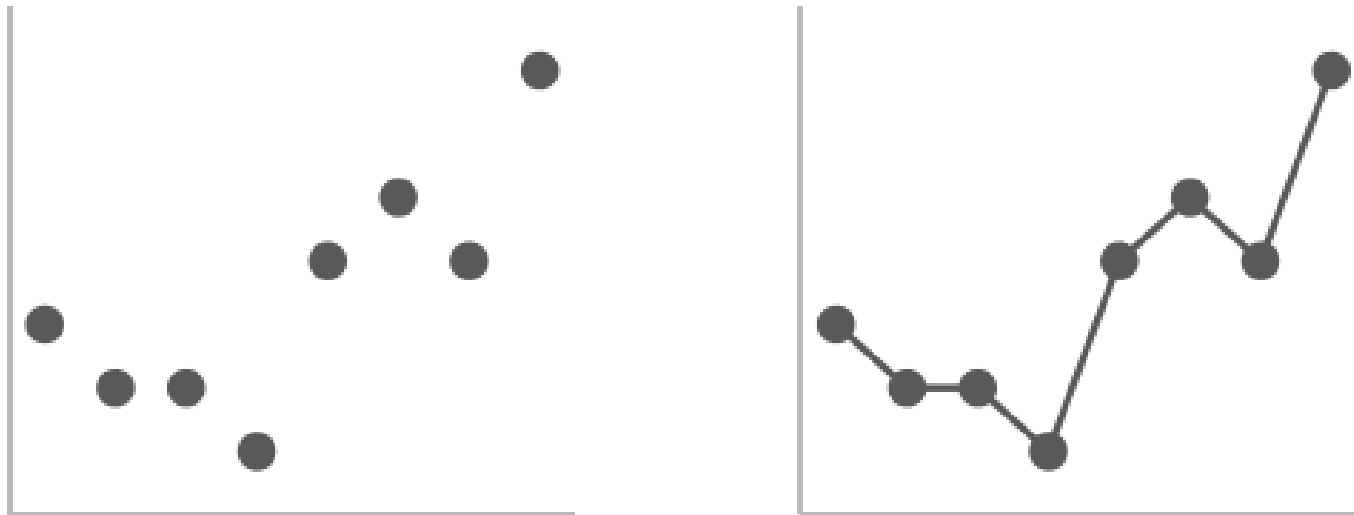


FIGURE 3.12 Lines connect the dots

Gestalt principles of visual perception

Closure

The elements in Figure 3.7 will tend to be perceived as a circle first and only after that as individual elements

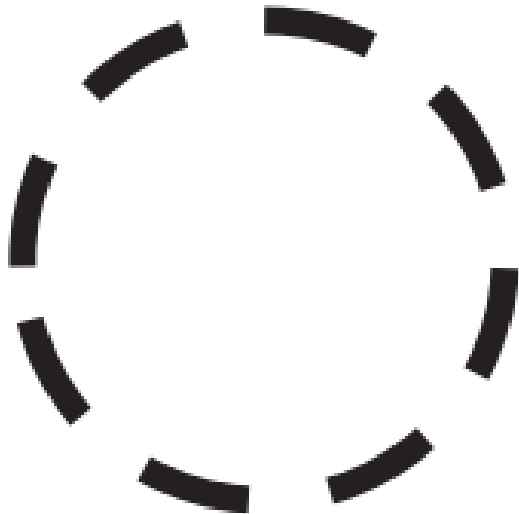


FIGURE 3.7 Gestalt principle of closure

The closure principle tells us that these are unnecessary—we can remove them and our graph still appears as a cohesive entity

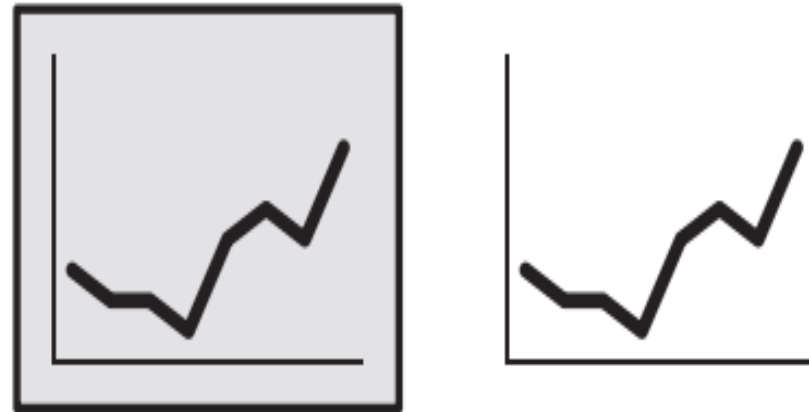


FIGURE 3.8 The graph still appears complete without the border and background shading

Gestalt principles of visual perception

Continuity

- The principle of continuity is similar to closure: when looking at objects, our eyes seek the smoothest path and naturally create continuity in what we see even where it may not explicitly exist

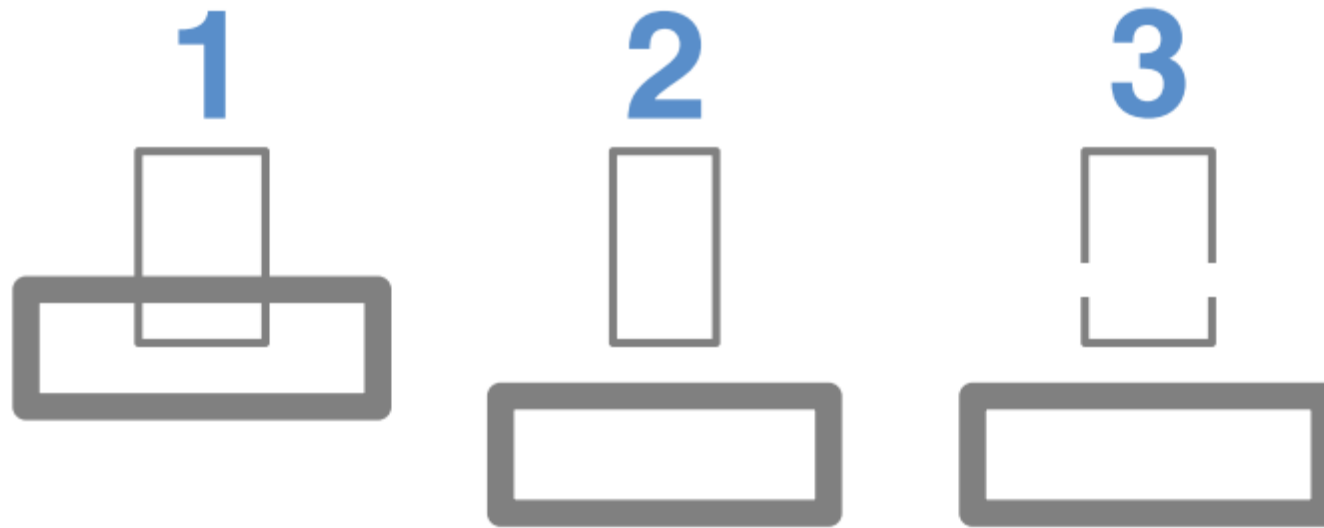


FIGURE 3.9 Gestalt principle of continuity

Gestalt principles of visual perception

Continuity

- In the application of this principle: We can remove the vertical y-axis line from the graph in the Figure 3.10 altogether

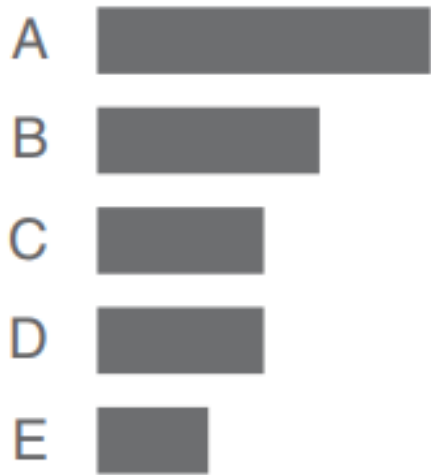
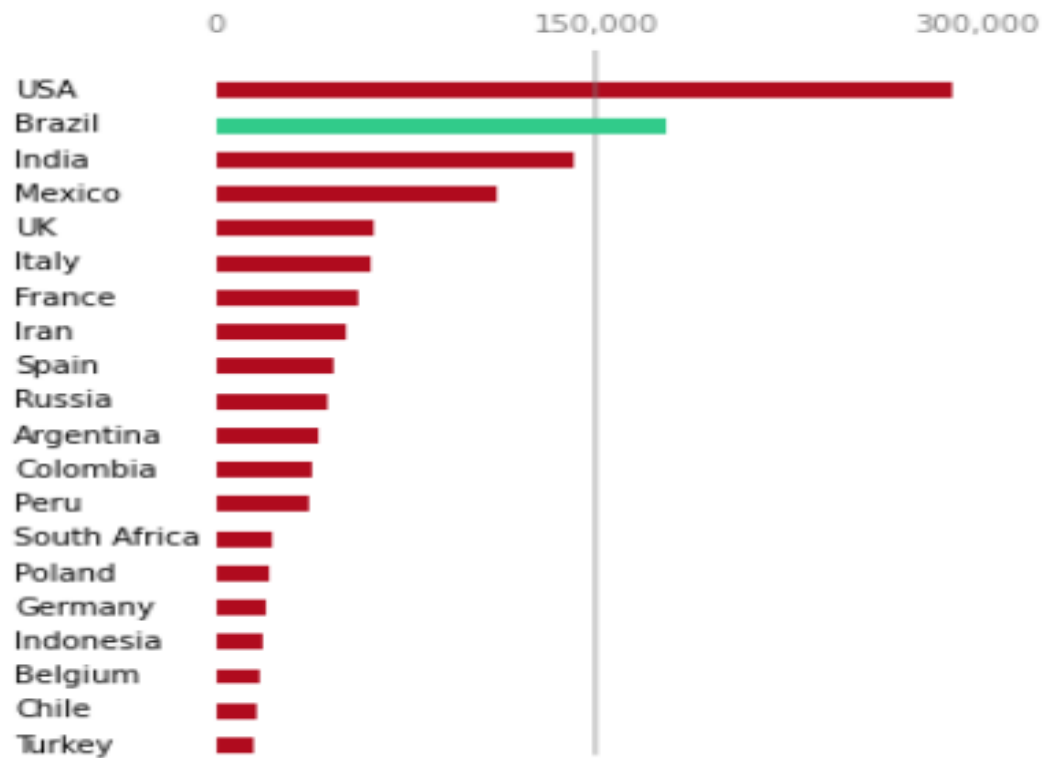


FIGURE 3.10 Graph with y-axis line removed

Focus your audience's attention

Preattentive attributes

The Death Toll Worldwide Is 1.5M+
Top 20 countries by death toll (December 2020)



- Preattentive attributes like size, color, position can be leveraged to help **direct your audience's attention** to where you want them to focus it
- They can be used to create a visual hierarchy of elements to **lead your audience through the information you want to communicate** in the way you want them to process it

— Preattentive attributes

756395068473
658663037576
860372658602
846589107830

FIGURE 4.2 Count the 3s example

— Preattentive attributes

Focus your audience's attention

Preattentive attributes

756**3**9506847**3**
65866**3**0**3**7576
860**3**72658602
8465891078**3**0

FIGURE 4.3 Count the 3s example with preattentive attributes

Focus your audience's attention

Preattentive attributes

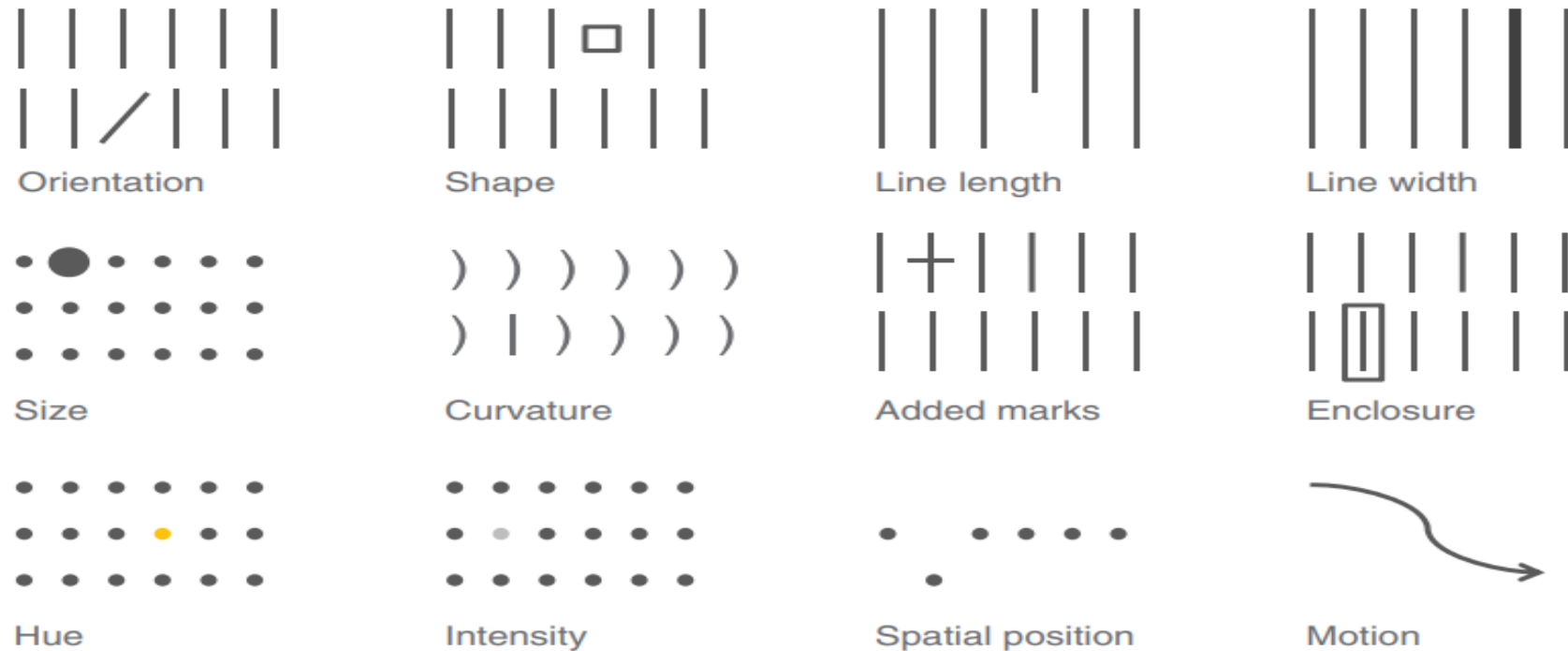


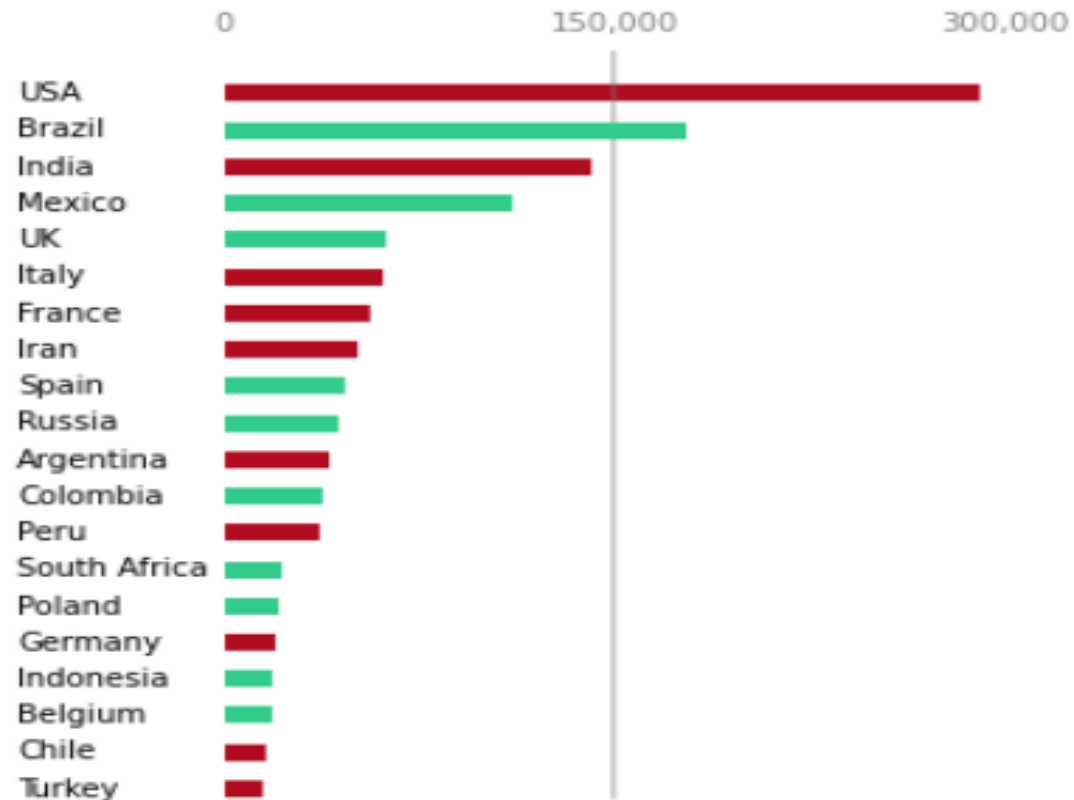
FIGURE 4.4 Preattentive attributes

Source: Adapted from Stephen Few's *Show Me the Numbers*, 2004.

Focus your audience's attention

Preattentive attributes

The Death Toll Worldwide Is 1.5M+
Top 20 countries by death toll (December 2020)



Pre-attentive attributes can become **inefficient** if we **overuse them**, so we must be very selective about what we make different. Below, for instance, nothing stands out.

Decluttering: step-by-step

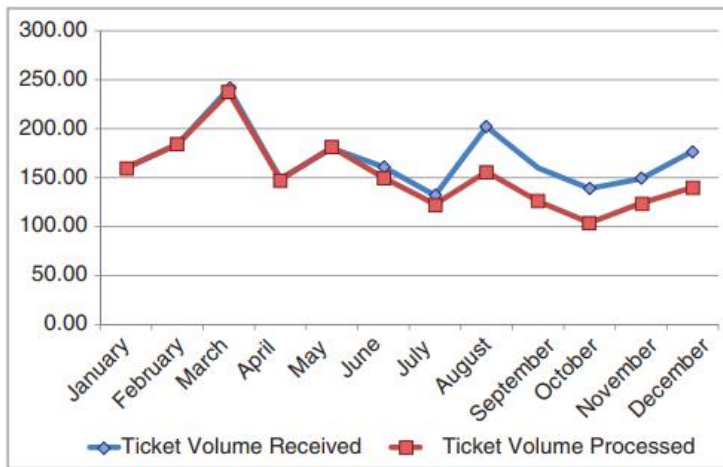
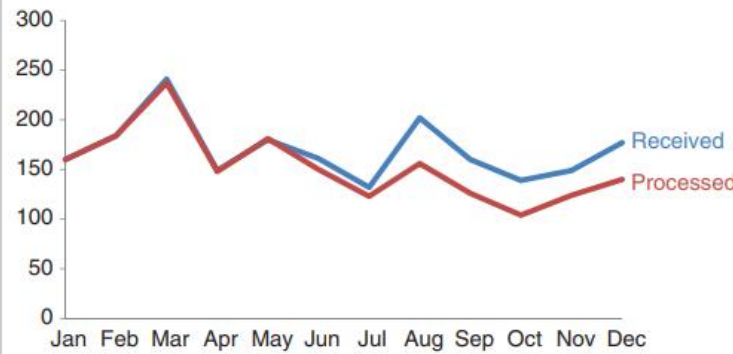
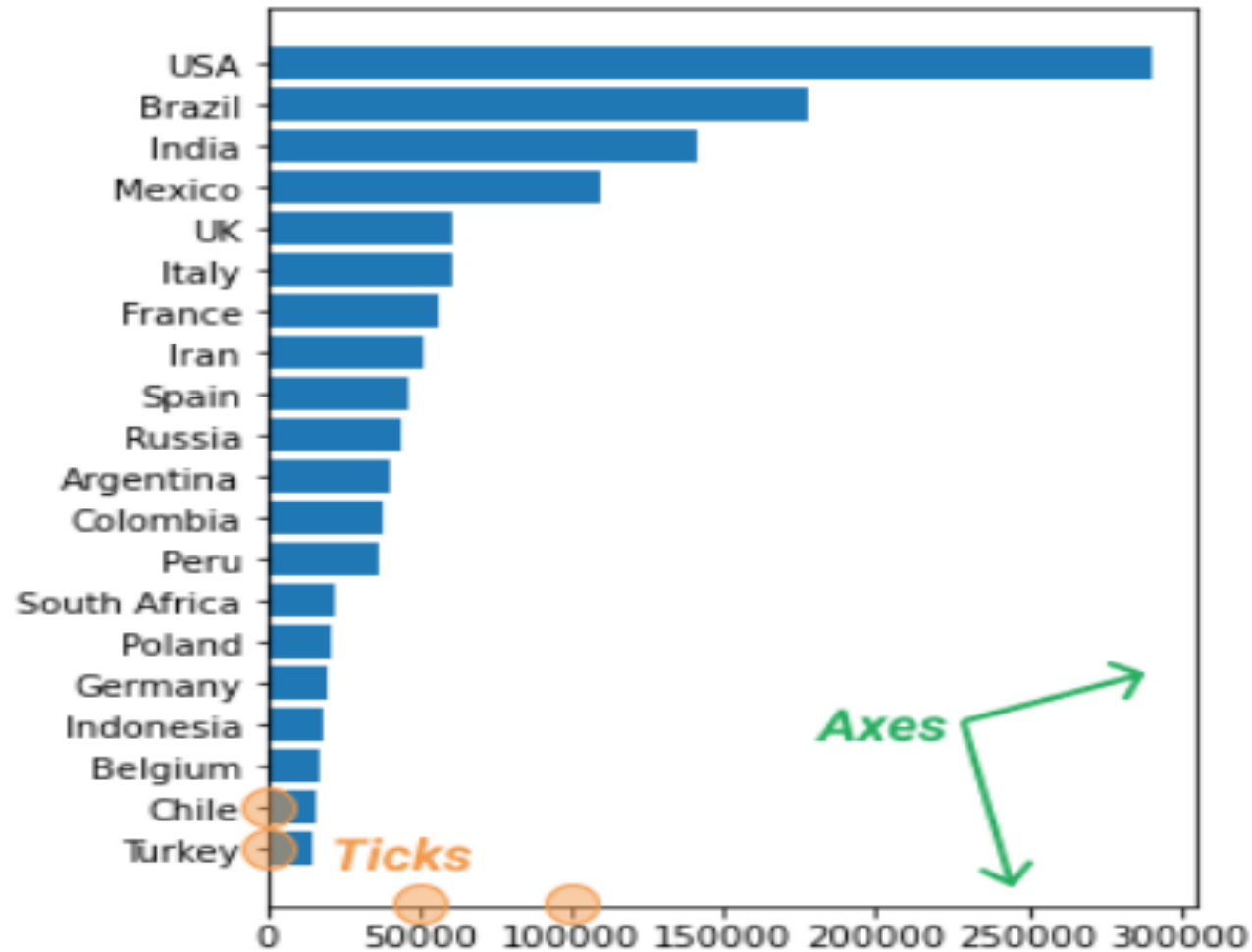


FIGURE 3.24 Before-and-after

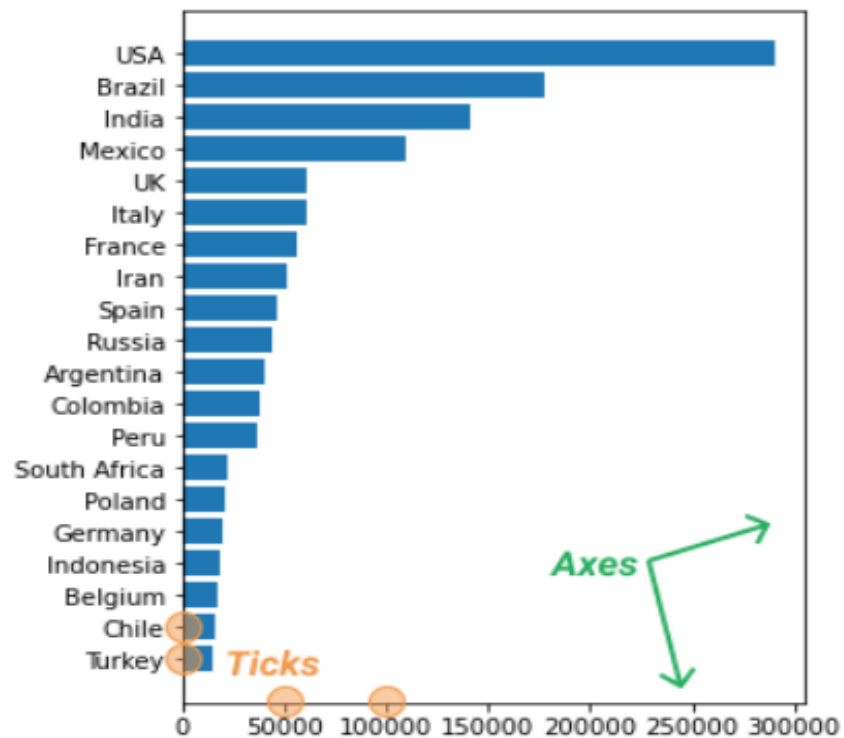


- Remove chart border
- Remove gridlines
- Remove data markers
- Clean up axis labels
- Label data directly
- Leverage consistent color

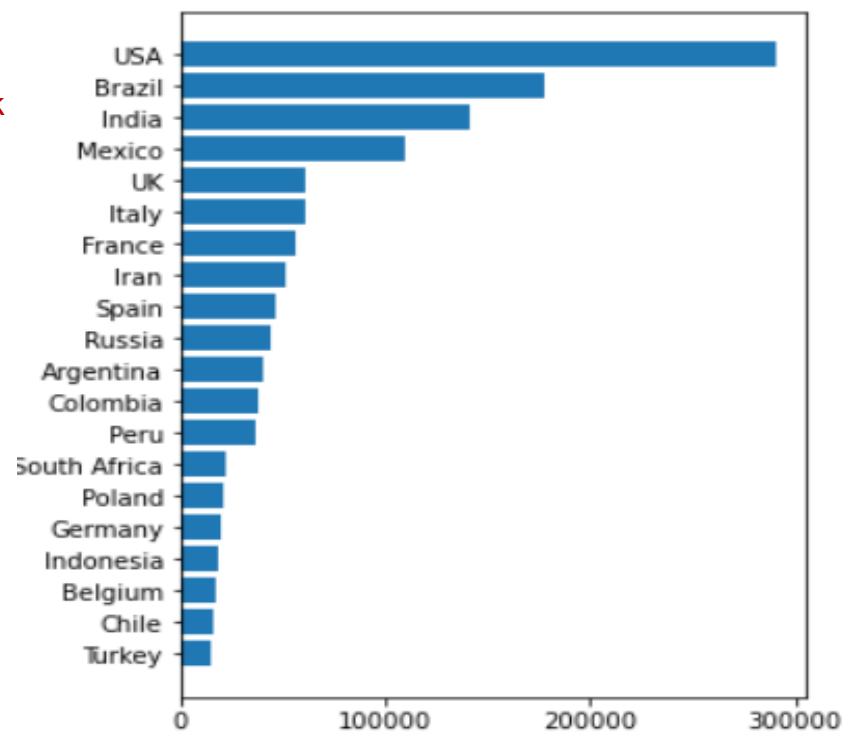
Decluttering: step-by-step



Decluttering: step-by-step

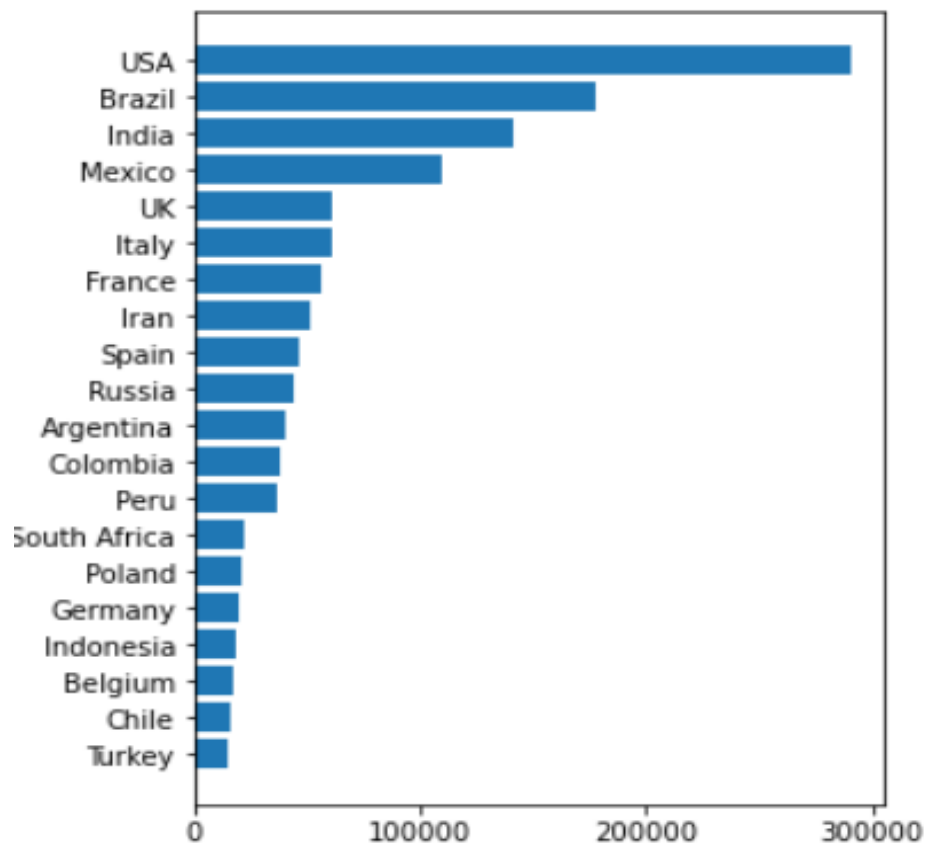


Remove some of the x-tick labels

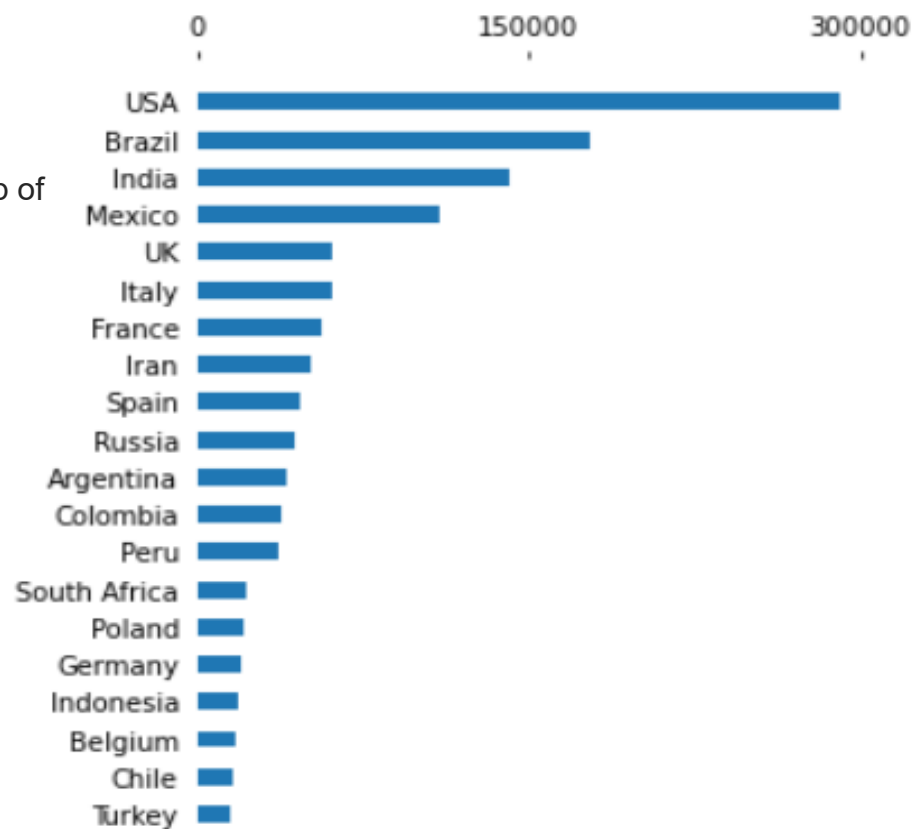


Decluttering: step-by-step

- People usually start from top left and follow a zigzag pattern until they reach bottom right.

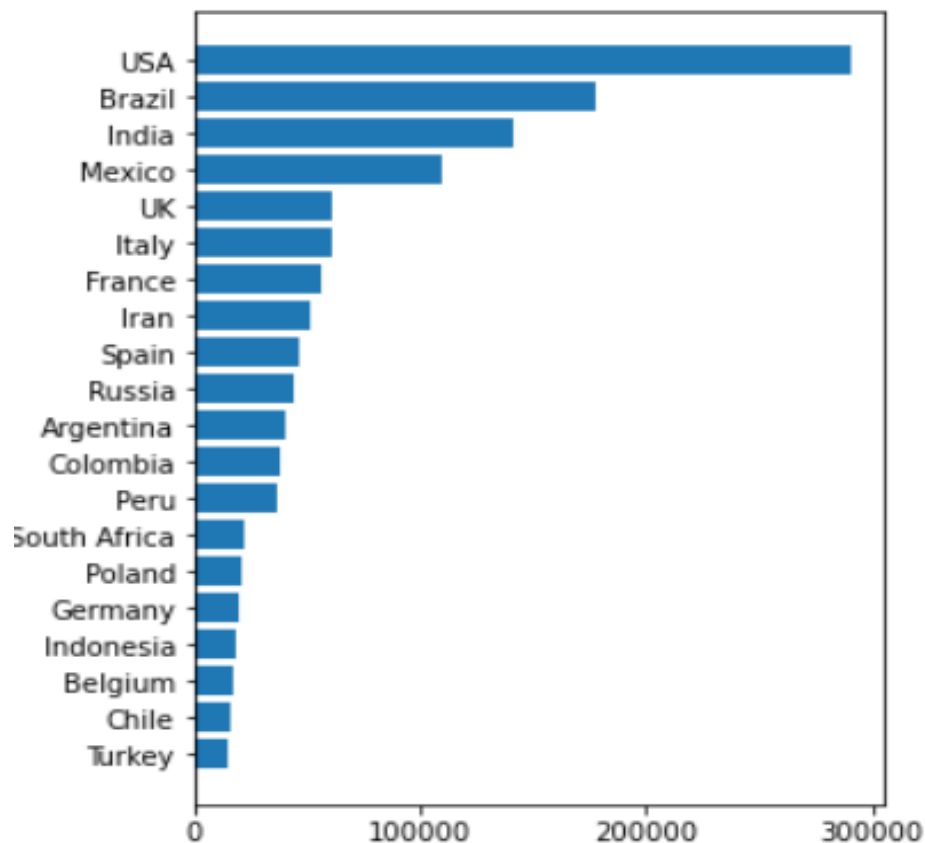


Move the tick labels at the top of the graph

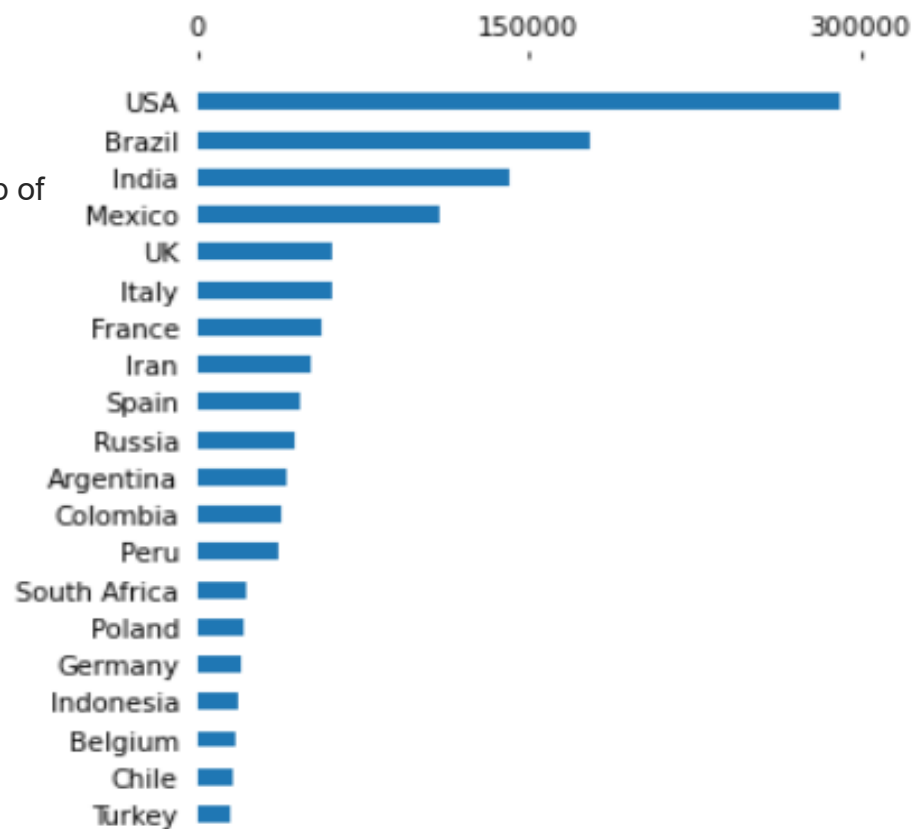


Decluttering: step-by-step

- People usually start from top left and follow a zigzag pattern until they reach bottom right.

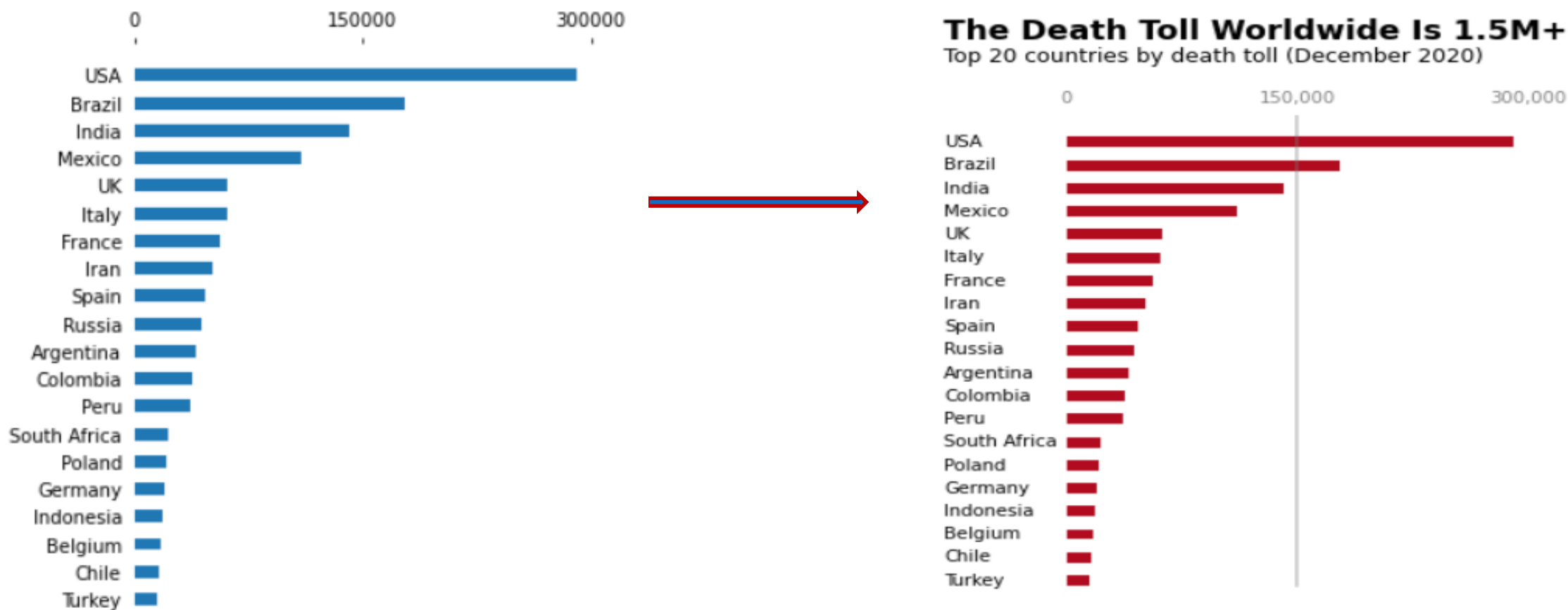


Move the tick labels at the top of the graph



Decluttering: step-by-step

- Instead of adding an x-axis label, we'll use the title and subtitle area to give the readers the necessary details



— Tell a story

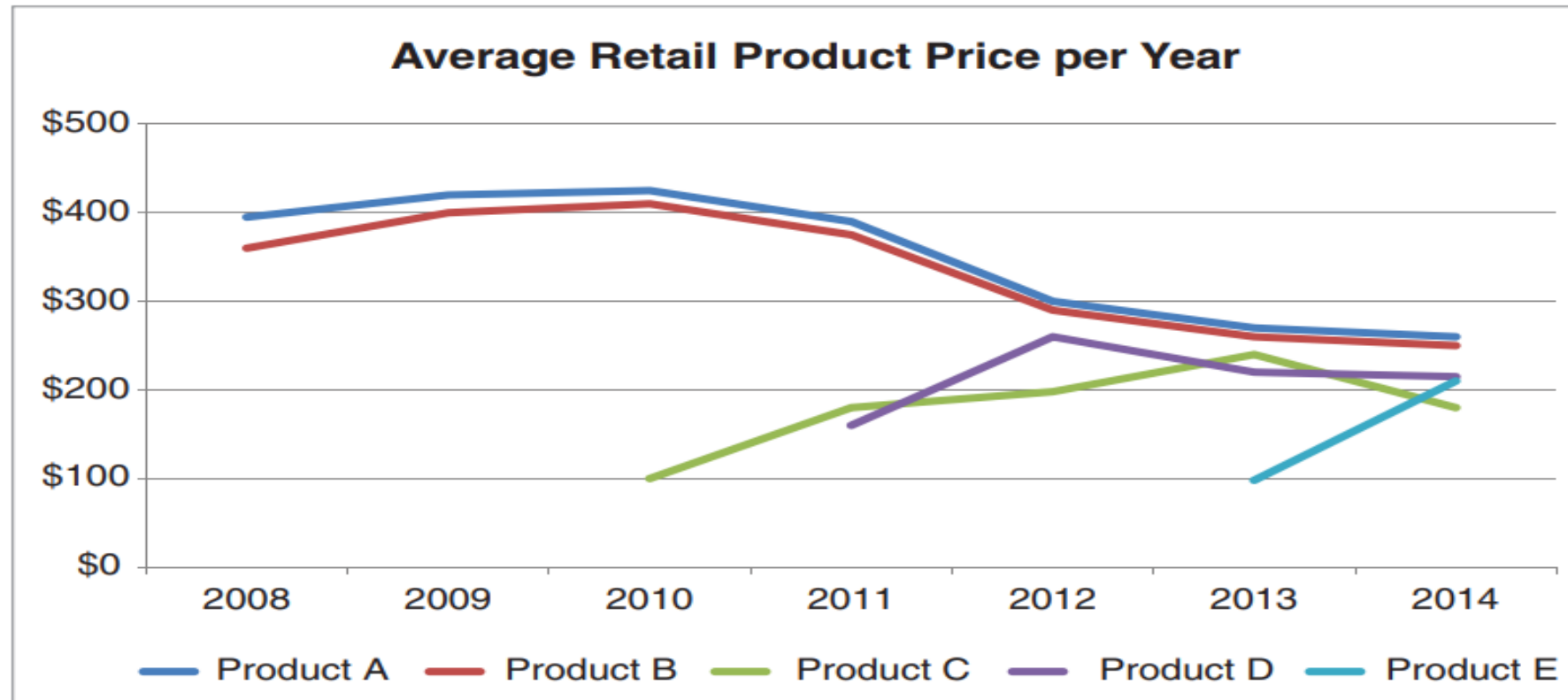


FIGURE 8.5 Single line graph for all products

— Tell a story

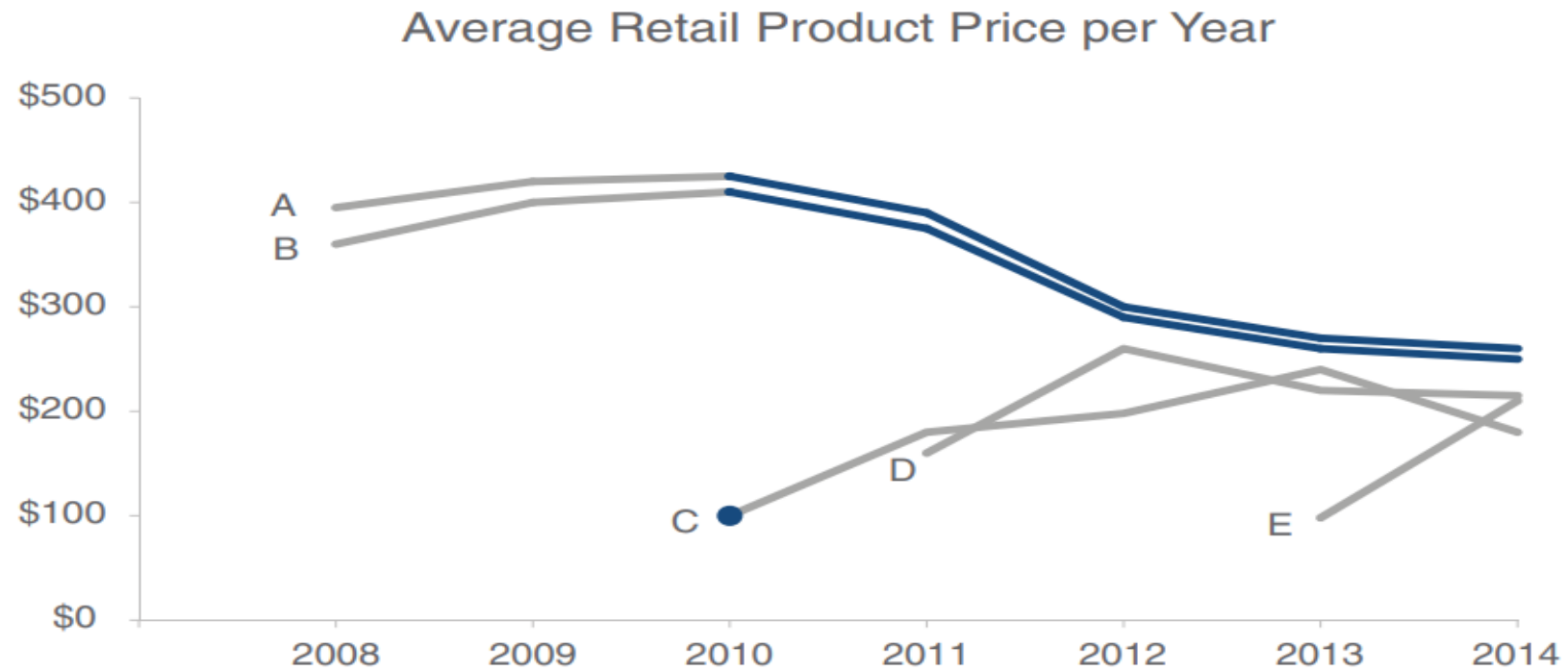
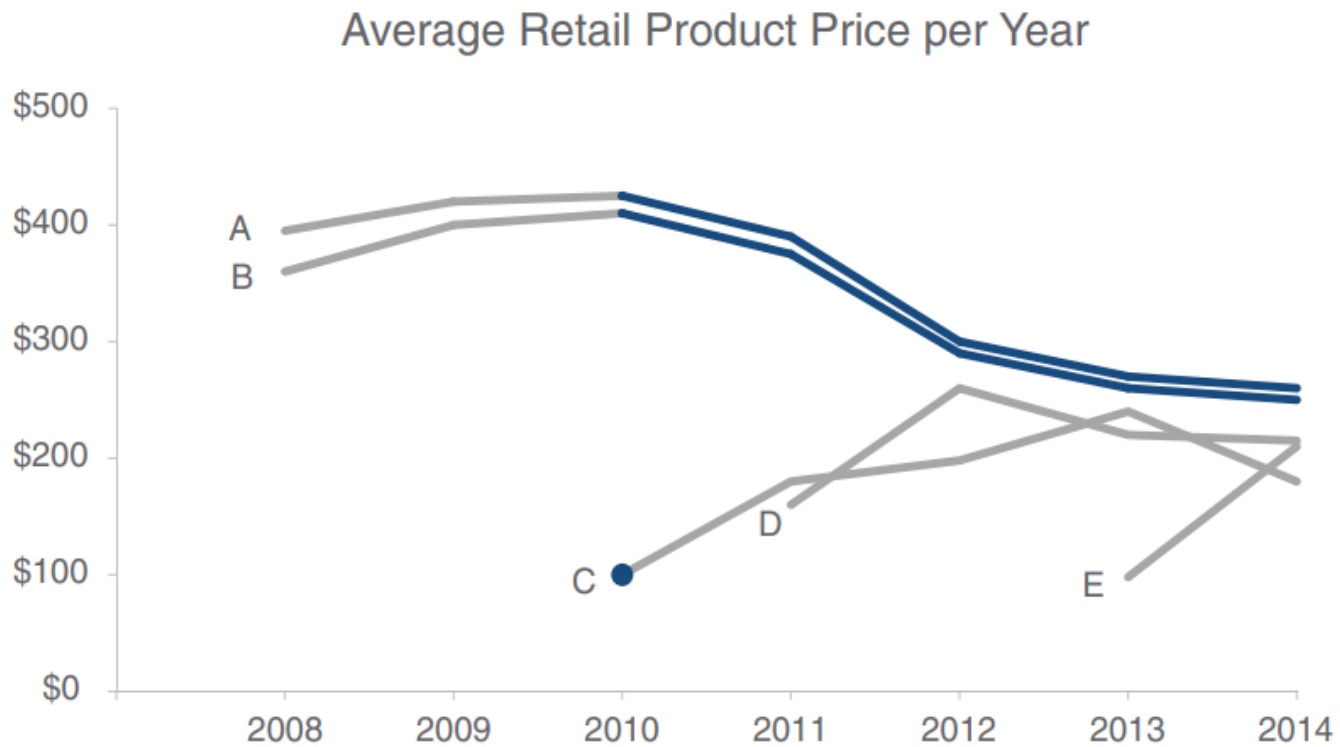


FIGURE 8.7 Focus the audience's attention

— Tell a story



After the launch of Product C in 2010, **the average retail price of existing products declined**

FIGURE 8.7 Focus the audience's attention

— Tell a story

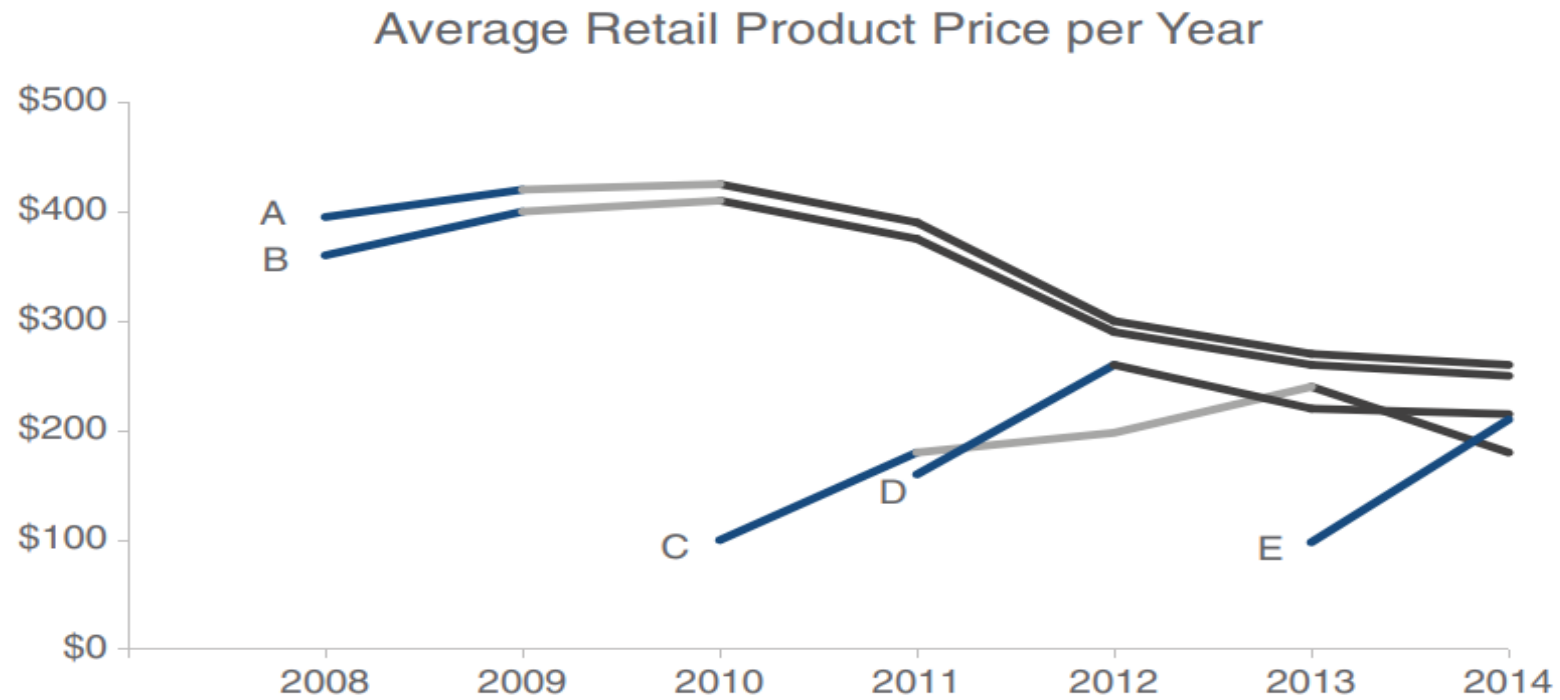


FIGURE 8.8 Refocus the audience's attention

— Tell a story

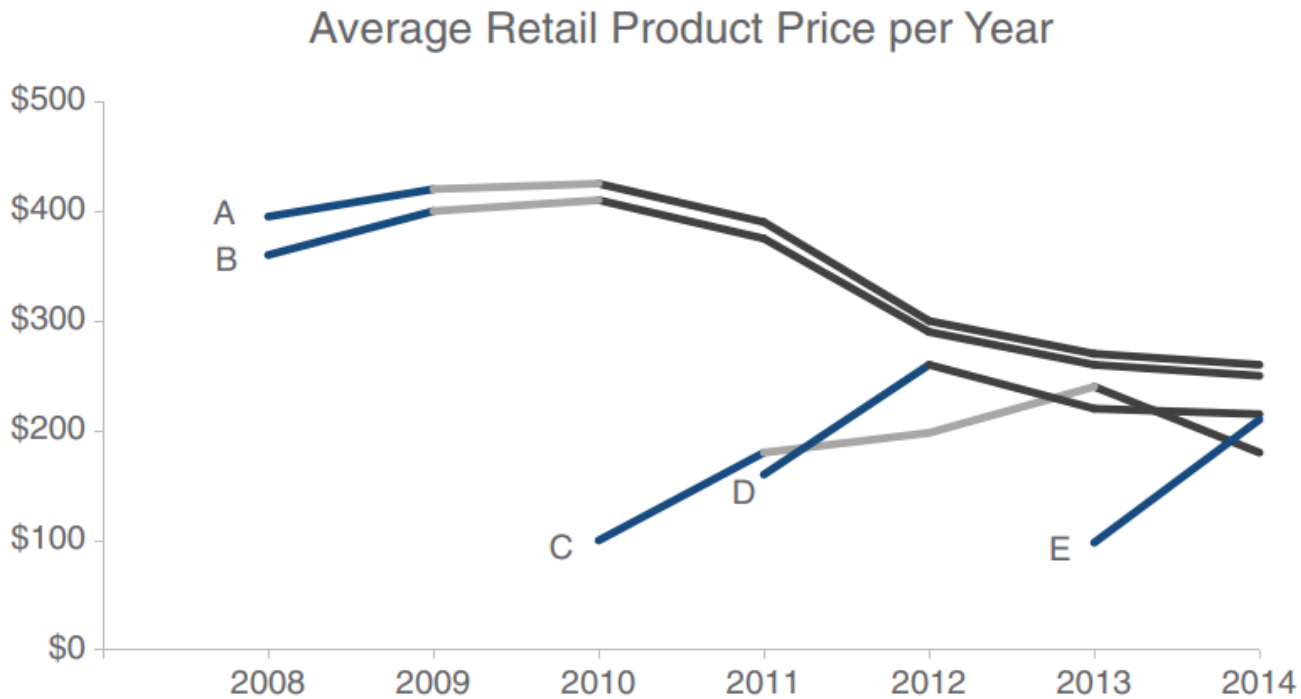


FIGURE 8.8 Refocus the audience's attention

With the launch a new product in this space, it is typical to see an initial average retail price **increase**, followed by a **decline**

— Tell a story

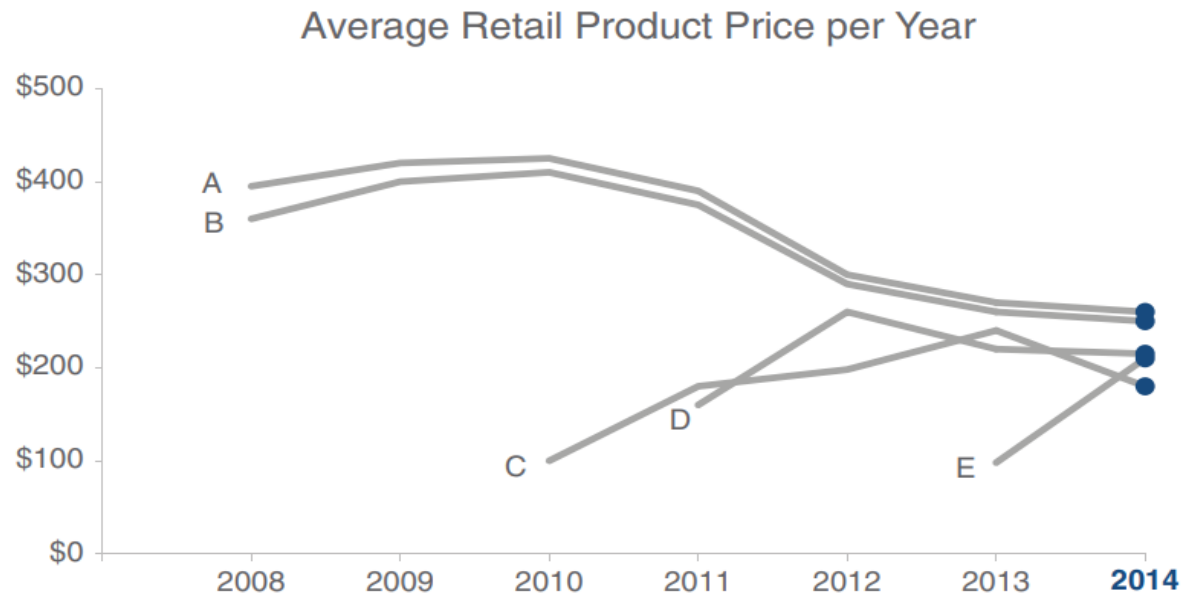


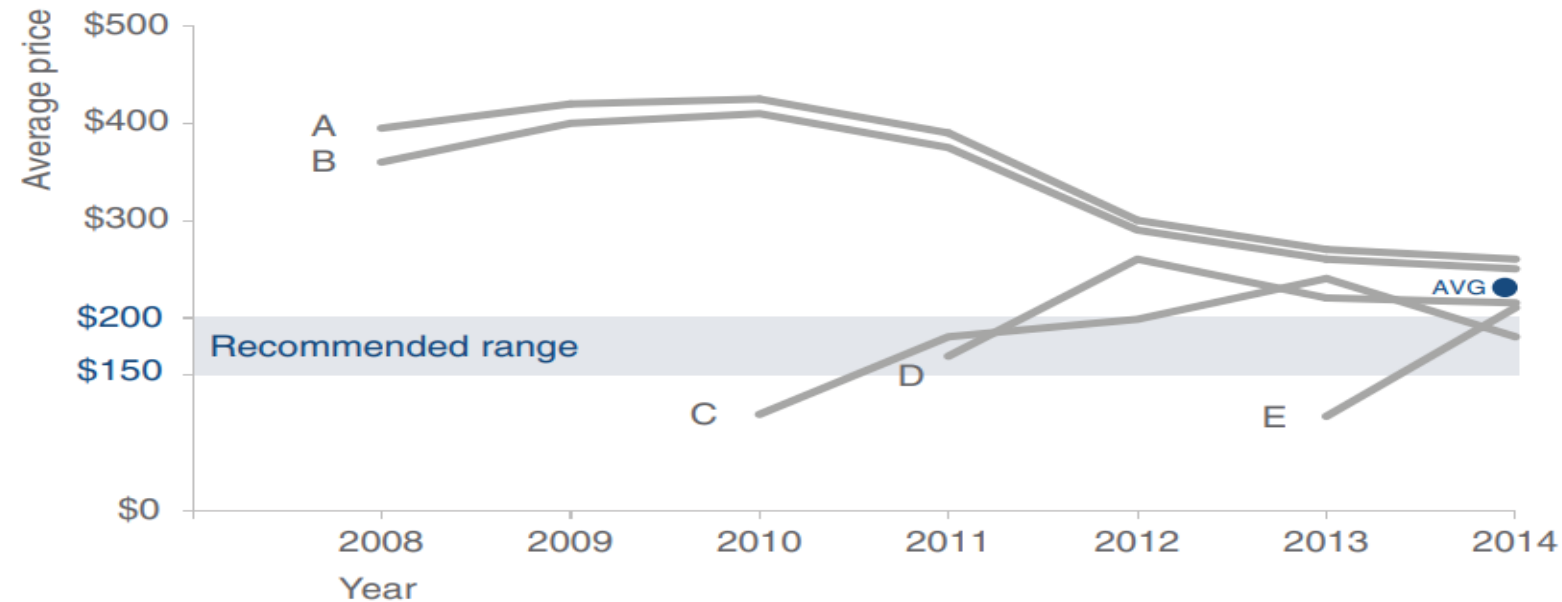
FIGURE 8.9 Refocus the audience's attention again

As of 2014, retail prices have converged, with an **average retail price of \$223**, ranging from a low of \$180 © to a high of \$260 (A)

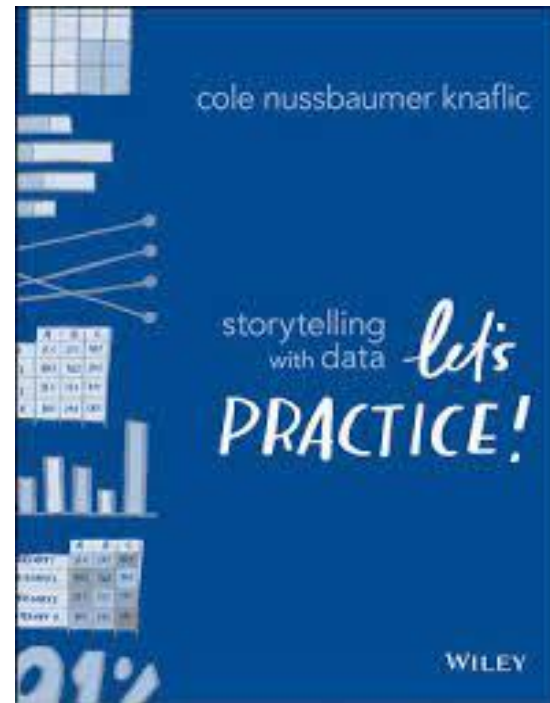
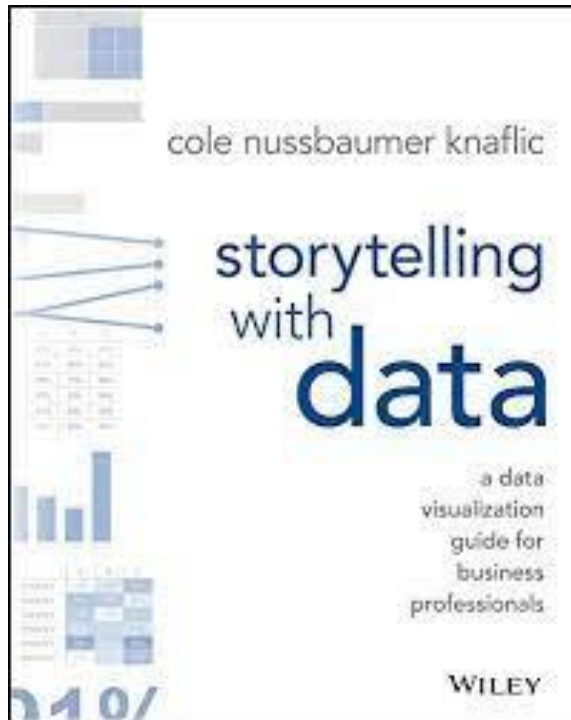
— Tell a story

To be competitive, we recommend introducing our product *below the \$223 average price point* in the **\$150–\$200 range**

Retail price over time



— References



Source: <https://www.dataquest.io/course/storytelling-data-visualization/>
https://matplotlib.org/stable/api/axes_api.html#matplotlib.axes.Axes