Chapter 2 Exploring Data with Tables and Graphs
Graphs that Enlighten and Graphs that Deceive (2.3)

Objectives

In this lesson, we will learn how to

- recognize several common graphs that help us understand data;
- select appropriate graphs for quantitative, categorical, and time-series data;
- construct and interpret dotplots, stemplots, bar graphs, Pareto charts, pie charts, and frequency polygons;
- compare distributions using relative frequency polygons;
- identify misleading graphical features such as truncated vertical scales and distorted pictographs;
- apply basic principles for designing graphs that reveal the true nature of the data.

1. Why More Graphs?

Technology gives us many ways to display data.

But simply creating a graph does not guarantee that the graph is useful.

Key Idea

Good graphs should enlighten us.

They should help us understand the true nature of the data instead of distracting us or creating a misleading impression.

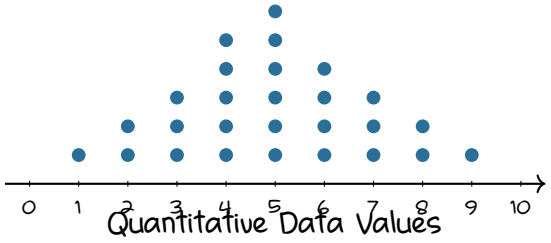
In this section, we study both

graphs that enlighten and graphs that deceive.

2. Dotplots

Dotplot

A dotplot is a graph of quantitative data in which each data value is plotted as a point or dot above a horizontal scale of values.
Dots representing equal values are stacked.



Features of a Dotplot

A dotplot

- displays the shape of the distribution;
- makes clusters, gaps, and unusual observations visible;
- usually allows us to recreate the original list of data values.

Example 1:
A quantitative data set contains only 18 observations.
Would a dotplot be a reasonable graph?

Solution

Yes.
A dotplot works especially well with a relatively small quantitative data set because it shows the distribution while preserving the individual data values.

3. Stemplots

A stemplot, or stem-and-leaf plot, separates each quantitative value into two parts.

Stemplot

Each value is separated into

- a stem, usually the leftmost digit or digits;
- a leaf, usually the rightmost digit.

For example,

$$47 \qquad \longrightarrow \qquad 4 \mid 7.$$

Here,

$$\text{stem} = 4, \qquad \text{leaf} = 7.$$

Example 2:
Construct a stemplot for

$$21,\ 24,\ 26,\ 31,\ 33,\ 34,\ 42.$$

Solution

Stem	Leaves
2	1 4 6
3	1 3 4
4	2

A key should also be included:

$$2 \mid 1 = 21.$$

Features of a Stemplot

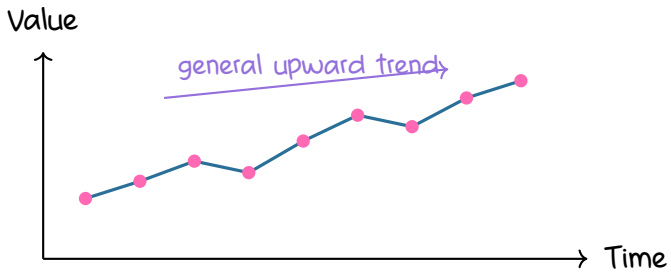
A stemplot

- shows the shape of the distribution;
- retains the original data values;
- automatically sorts the data values in order.

4. Time-Series Graphs

Time-Series Graph

A time-series graph displays quantitative data collected at different points in time, such as monthly, yearly, or semester data.



Main Purpose

A time-series graph is especially useful for revealing

trends over time.

A trend might show that values are generally

- increasing;
- decreasing;
- remaining relatively stable;
- or changing in a repeating pattern.

Example 3:
Suppose a university records online-course enrollment every semester for eight years.
Which graph would be especially useful?

Solution

A

time-series graph

would be appropriate because the observations are recorded over time.

5. Bar Graphs

Bar Graph

A bar graph uses bars of equal width to display frequencies of categories of categorical data.
The bars may or may not be separated by small gaps.

Important Difference

Histogram $\rightarrow$ quantitative data

Bar graph $\rightarrow$ categorical data

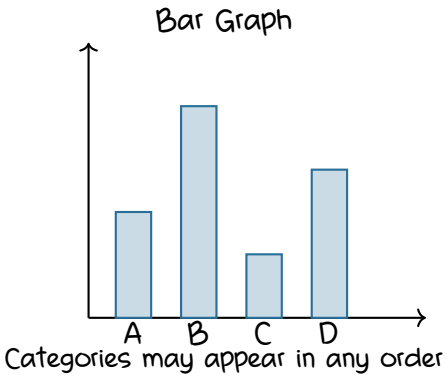
The order of categories in an ordinary bar graph usually has no numerical meaning.

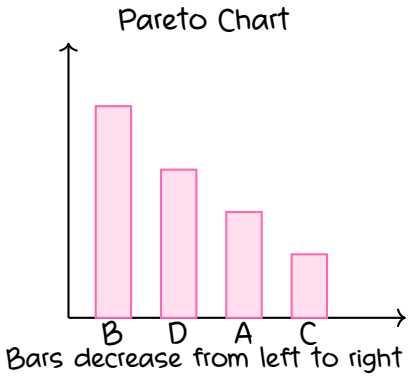
6. Pareto Charts

A Pareto chart is a special kind of bar graph.

Pareto Chart

A Pareto chart is a bar graph for categorical data in which the bars are arranged in descending order of frequency.
Therefore, the bars decrease in height from left to right.





Why Pareto Charts Are Useful

A Pareto chart

- makes categories easy to compare;
- immediately identifies the most frequent categories;
- draws attention to the categories that may deserve the greatest attention.

Bar Graph or Pareto Chart?

What feature distinguishes a Pareto chart from an ordinary bar graph?

Solution

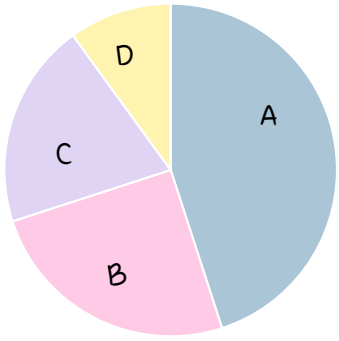
A Pareto chart arranges its bars

from highest frequency to lowest frequency.

7. Pie Charts

Pie Chart

A pie chart displays categorical data as slices of a circle.
The size of each slice is proportional to the frequency or relative frequency of that category.



Slice size represents category frequency

Interpretation

The entire circle represents

$$\boxed{100\%}$$

of the observations.
Larger slices represent categories containing larger proportions of the data.

Pie charts are widely used, but comparing similar slice sizes can be more difficult than comparing the heights of bars.

Example 4:
You want students to compare the frequencies of six college majors as accurately as possible.
Would a bar graph or pie chart usually be easier to compare?

Solution

A

bar graph

is usually easier to compare accurately because people can compare bar heights more easily than similar angles or slice areas.

8. Frequency Polygons

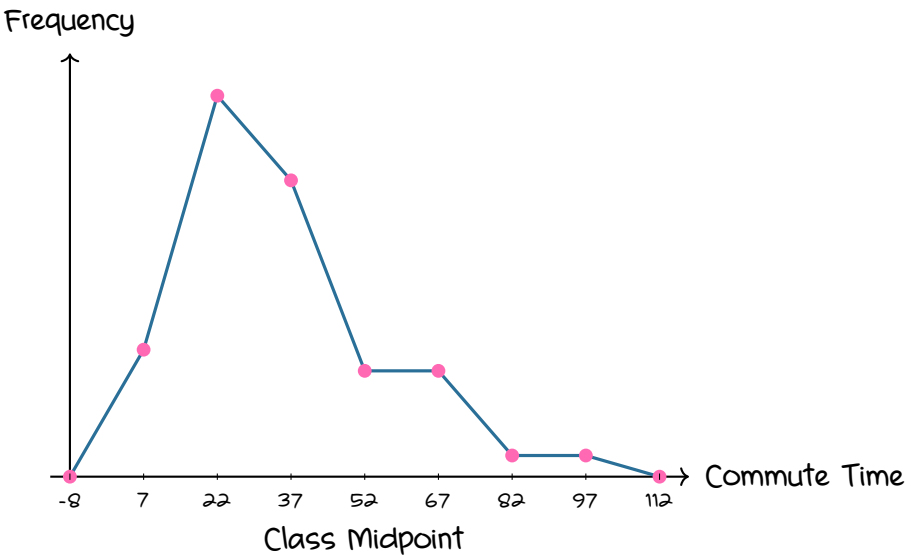
Frequency Polygon

A frequency polygon uses line segments connected to points located directly above the class midpoint values.
It is similar to a histogram, but it uses points and line segments instead of bars.

Using the Los Angeles commute-time data from the previous section:

Class Midpoint	Frequency
7	6
22	18
37	14
52	5
67	5
82	1
97	1

we obtain:



Notice that we add a point with frequency 0 before the first class midpoint and another after the last midpoint.

This visually closes the polygon.

9. Relative Frequency Polygons

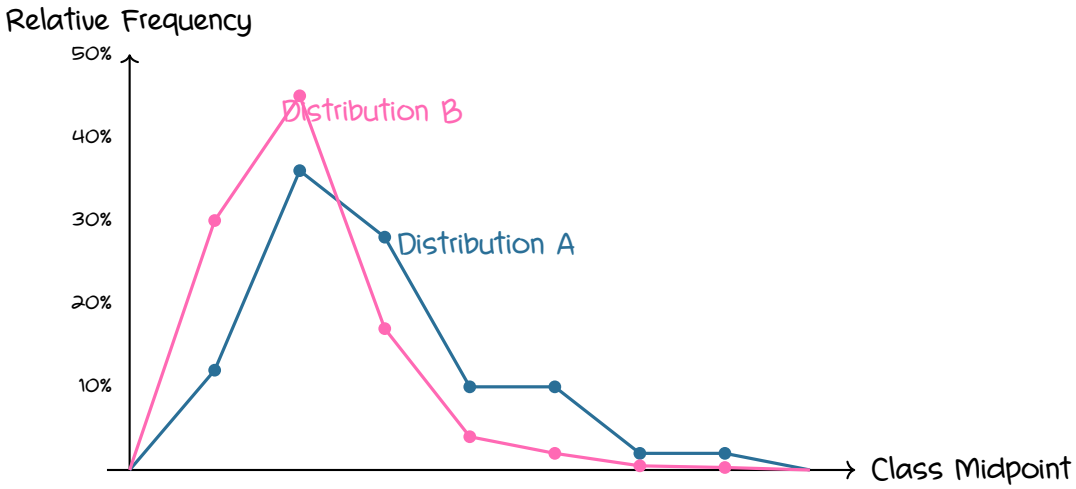
Relative Frequency Polygon

A relative frequency polygon is constructed in the same way as a frequency polygon, except that the vertical scale shows

relative frequencies or percentages

instead of ordinary frequencies.

One important advantage is that two distributions can be plotted on the same graph.



Why This Is Useful

Relative frequency polygons make it easier to compare

- the location of two distributions;
- their overall shapes;
- where each distribution is concentrated;
- and how their relative frequencies differ.

10. Graphs That Deceive

Graphs do not always tell the truth clearly.

A graph may contain correct numerical information while still creating a misleading visual impression.

Important Question

Whenever you examine a graph, ask:

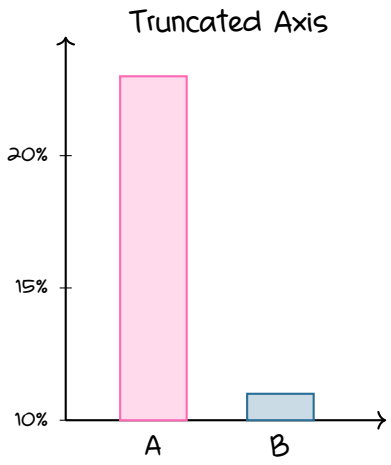
Does the graph accurately represent the size of the differences?

Nonzero Vertical Axis

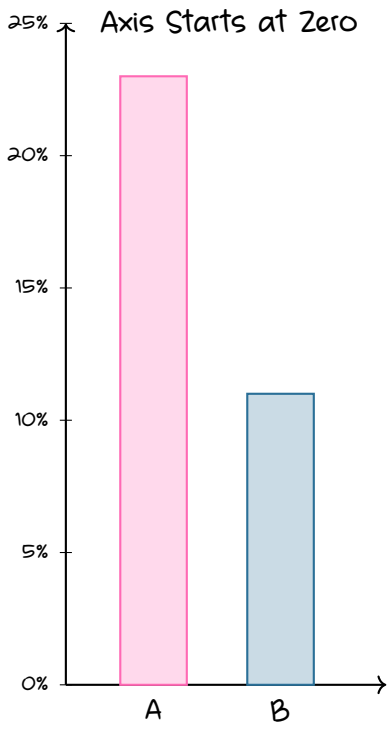
One of the most common deceptive techniques is starting the vertical scale at a number greater than zero.

Warning: Nonzero Vertical Axis

A vertical axis that begins above zero can make small or moderate differences appear dramatically larger.



Difference looks enormous



Same data, more honest scale

Both graphs above display the same two values.

The first graph starts at 10%, which visually exaggerates the difference.

The second graph starts at 0%, providing a more proportional view of the difference.

Graph-Reading Habit

Before interpreting the height of bars, always inspect the vertical scale.

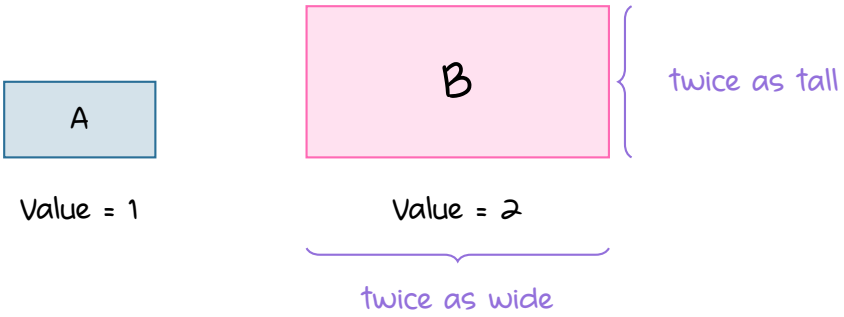
11. Pictographs

A pictograph uses pictures or drawings to represent data.

Pictographs may look attractive, but they can produce serious distortion.

Why Pictographs Can Mislead

One-dimensional numerical data are sometimes represented using two-dimensional or three-dimensional pictures.
Changing every dimension of the picture exaggerates the apparent change.



If both dimensions double,
the picture's area becomes 4 times as large.

Suppose value B is exactly twice value A.
If a picture for B is made

$$2 \times \text{ as wide}$$

and

$$2 \times \text{ as tall},$$

then its area becomes

$$2 \times 2 = 4$$

times as large.

So visually, B appears four times as large even though the data value is only twice as large.

Basic Geometry Behind the Distortion

For a square:

double each side $\Longrightarrow$ area becomes 4 times as large

For a cube:

double each dimension $\Longrightarrow$ volume becomes 8 times as large

Example 5:
A company compares two budgets using pictures of dollar bills.
Budget B is twice Budget A.
However, the picture for B is drawn twice as wide and twice as tall as the picture for A.
Why is the graph misleading?

Solution

The data value doubles, but the visual area becomes

$$2 \times 2 = 4$$

times as large.
Therefore, the picture makes Budget B appear four times as large instead of twice as large.

12. Principles for Better Graphs

Good Graph Design

- For a small data set of about 20 values or fewer, a table may be better than a graph.
- A graph should make us focus on the true nature of the data.
- Avoid decorative effects that distract from the data.

- Do not distort differences.
- Most of the visual information in the graph should communicate the data rather than decoration.

Example 6:
A colorful three-dimensional pie chart makes slices near the front appear larger than equally sized slices near the back.
Is this a good design?

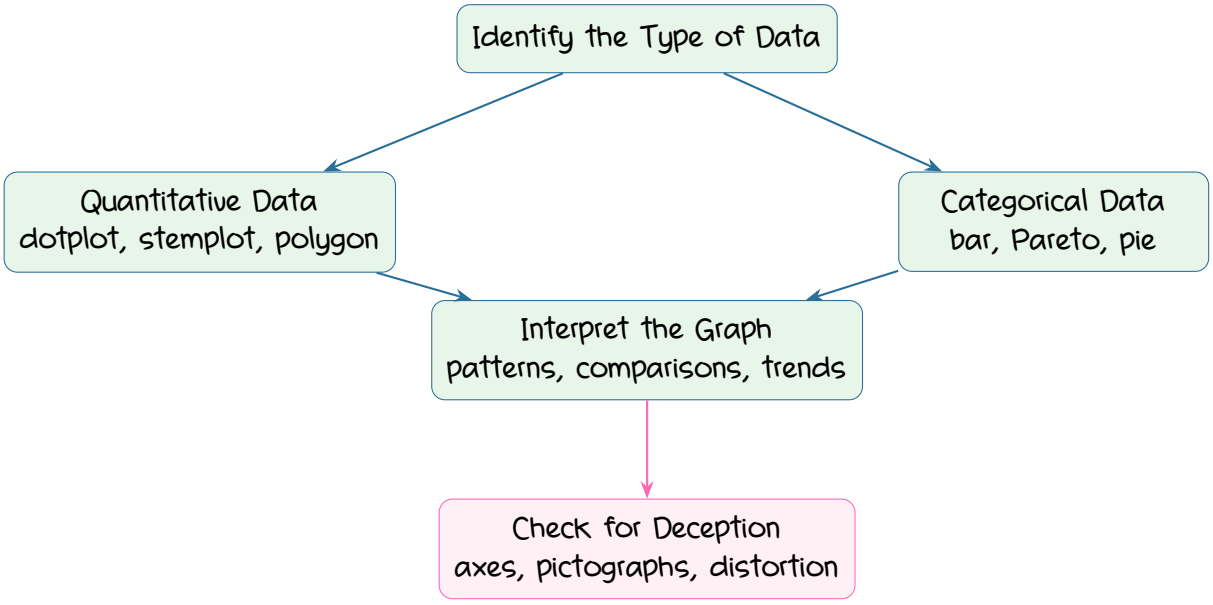
Solution

No.
The three-dimensional perspective can distort the apparent sizes of the categories.
A simpler two-dimensional graph would communicate the data more honestly.

13. Choosing an Appropriate Graph

Type of Data	Useful Graph	Main Purpose
Small quantitative data	Dotplot or stemplot	See individual values and distribution shape
Grouped quantitative data	Histogram or frequency polygon	See shape, center, and spread
Two quantitative distributions	Relative frequency polygons	Compare distributions
Categorical data	Bar graph	Compare categories
Categorical data ranked by frequency	Pareto chart	Identify most important categories
Categorical proportions	Pie chart	Show parts of a whole
Data collected over time	Time-series graph	Identify trends over time

14. The Big Picture



Summary

Key Ideas

- Dotplots display quantitative data using stacked dots and usually preserve the original observations.
- Stemplots show distribution shape while retaining and sorting the original data values.
- Time-series graphs reveal trends over time.
- Bar graphs display categorical data.
- A Pareto chart is a bar graph whose bars are arranged from greatest frequency to least frequency.
- A pie chart represents categorical data as portions of a whole.
- A frequency polygon connects points located above class midpoints.
- A relative frequency polygon uses proportions or percentages and is especially useful for comparing distributions.
- A vertical axis that begins above zero can exaggerate differences.
- Pictographs can distort data when two- or three-dimensional pictures are used to represent one-dimensional values.
- Good graphs should emphasize the data, avoid unnecessary decoration, and reveal the true nature of the data.