Data Types and Binning

Qualitative data describe *qualities* or *categories*.

Quantitative data represent counts or measurements.

Before summarizing raw data, it is common to first organize the values into groups, or **bins**, based on value ranges.

Summarizing Raw Data

Consider the following 20 scores from a 100-point exam:

76 80 78 76 94 75 98 77 84 88 81 72 91 72 74 86 79 88 72 75

Determine appropriate bins and make a frequency table including columns for relative and cumulative frequency.

TABLE 5.3 Frequency Table for Binned Exam Scores

Scores	Frequency	Relative Frequency	Cumulative Frequency
95 to 99	1	0.05 = 5%	1
90 to 94	2	0.10 = 10%	3
85 to 89	3	0.15 = 15%	6
80 to 84	3	0.15 = 15%	9
75 to 79	7	0.35 = 35%	16
70 to 74	4	0.20 = 20%	20
Total	20	1.00 = 100%	20

Frequency Tables

A basic **frequency table** has two columns:

- First column: the **categories** of data.
- Second column: the **frequency**, which is the number of times each category appears in the data set.

Relative frequency: fraction or percentage of the total

Cumulative frequency: total of frequencies for the current and all previous categories

Bar and Pie Graphs

Bar chart: length of bar corresponds to the *frequency* or *relative frequency*.

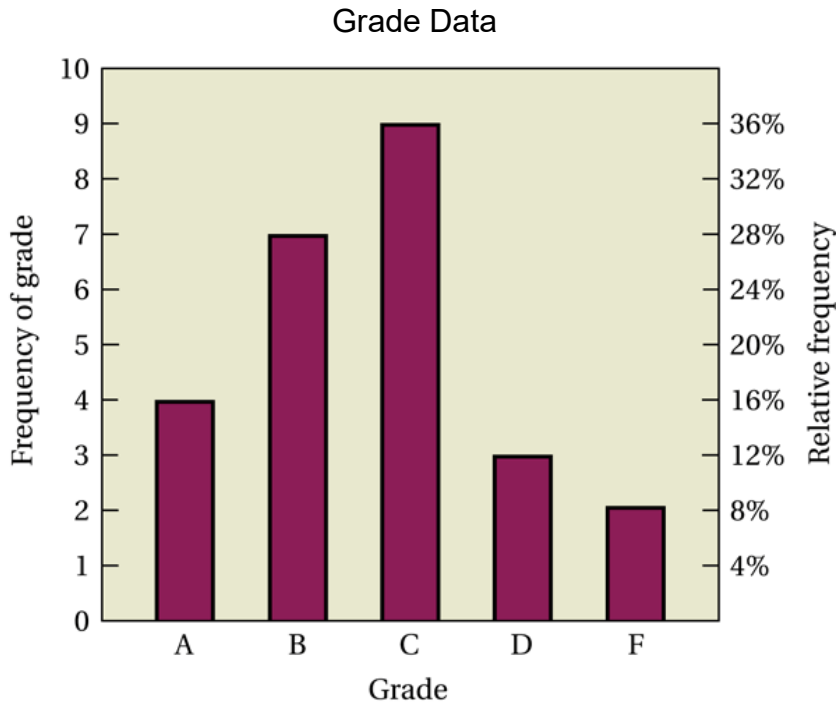
Pie chart: used primarily for *relative frequencies*. The size of each wedge is proportional to the relative frequency of the category it represents.

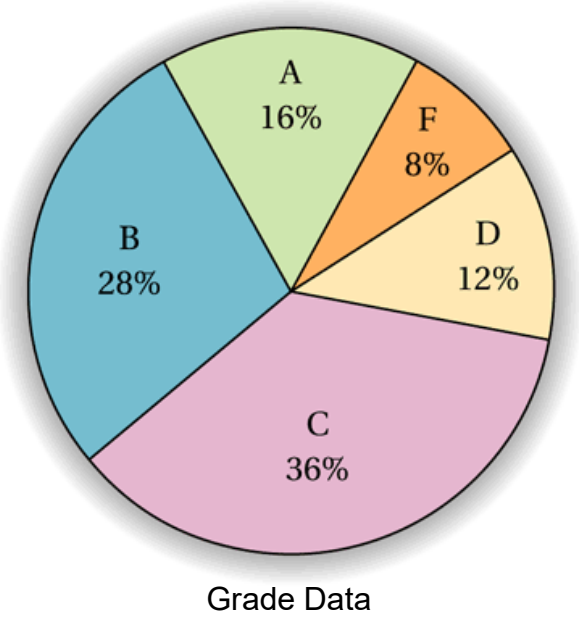
Bar and Pie Charts

The bar chart and pie chart below both show the data from table 5.1.

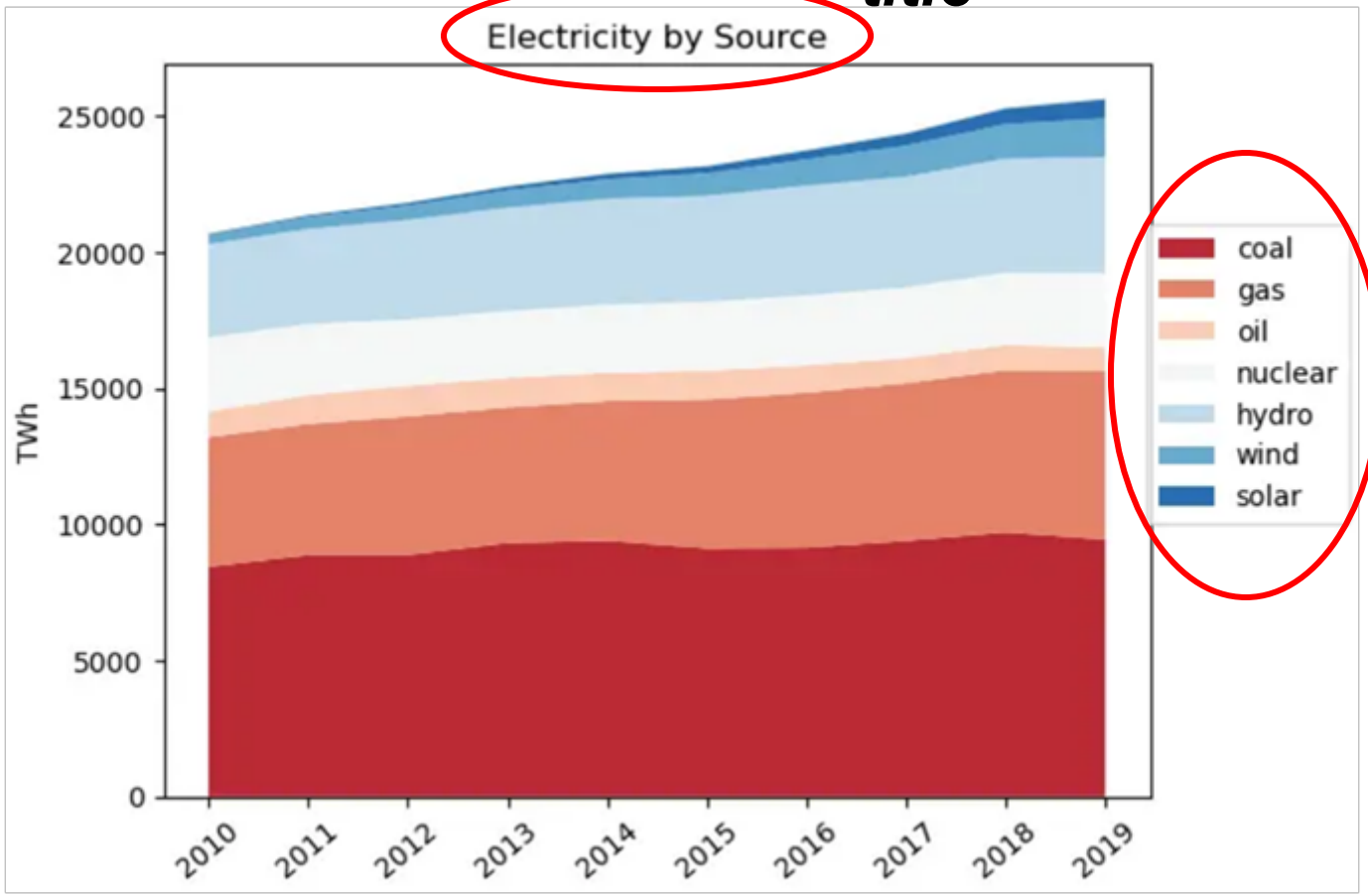
TABLE 5.1

Grade	Frequency
A	4
B	7
C	9
D	3
F	2
Total	25



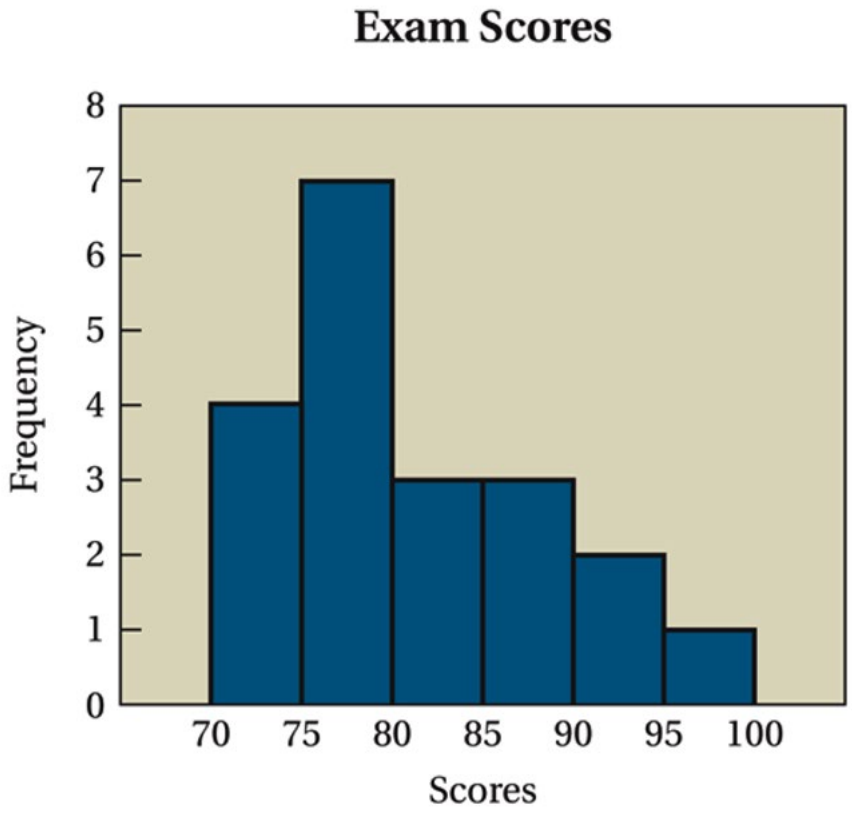


Important Labels for Graphs



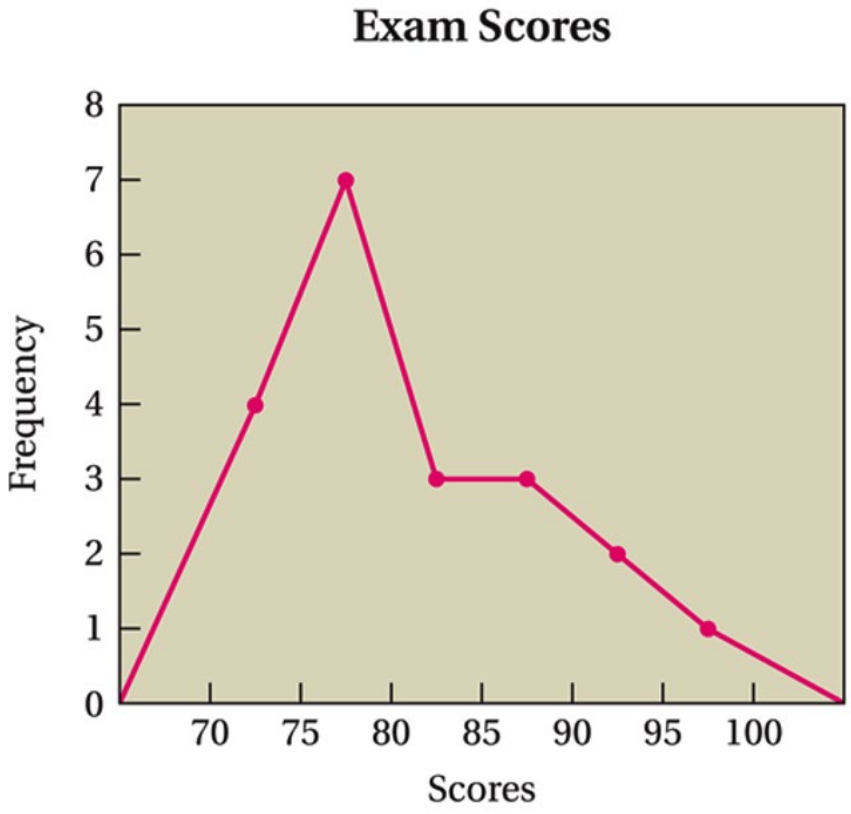
Histogram

- **Histogram** is a bar graph representing the distribution of quantitative data.



Line Chart

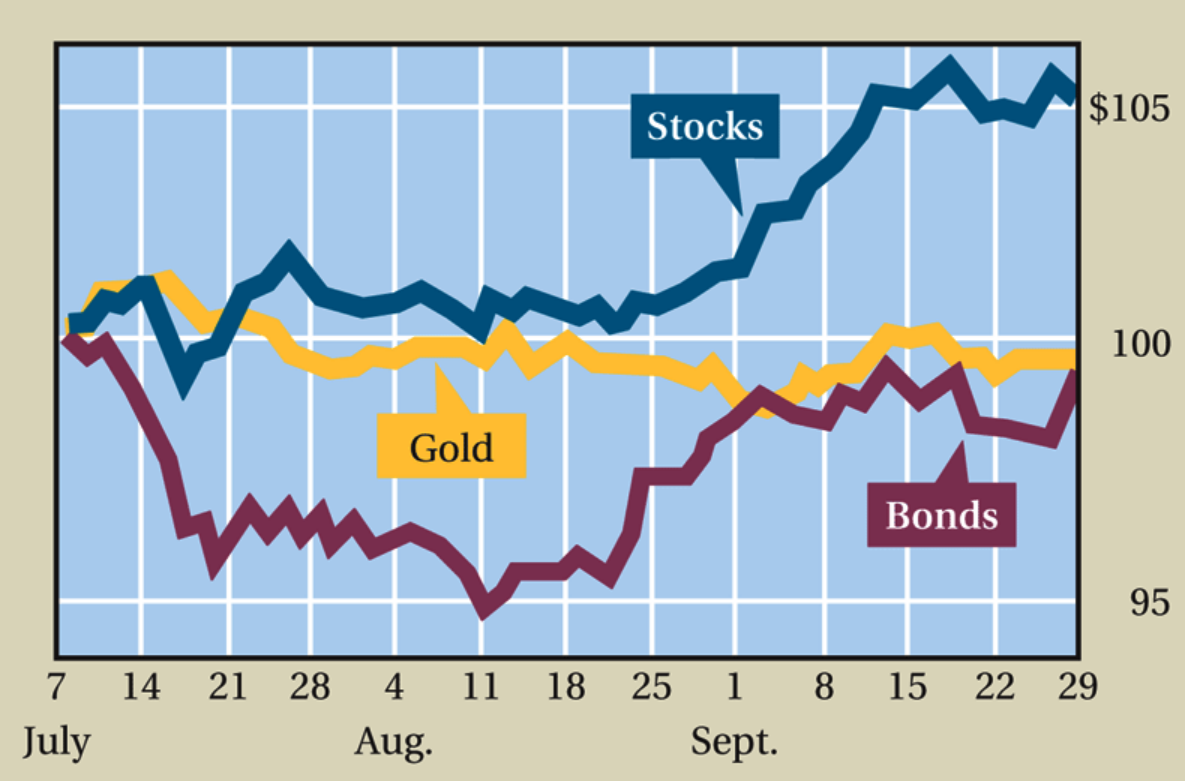
Line chart (line graph) is a type of chart that displays information as a series of data points connected by straight line segments.



Time-Series Diagram

A **time-series diagram** is a line chart in which the horizontal axis represents time.

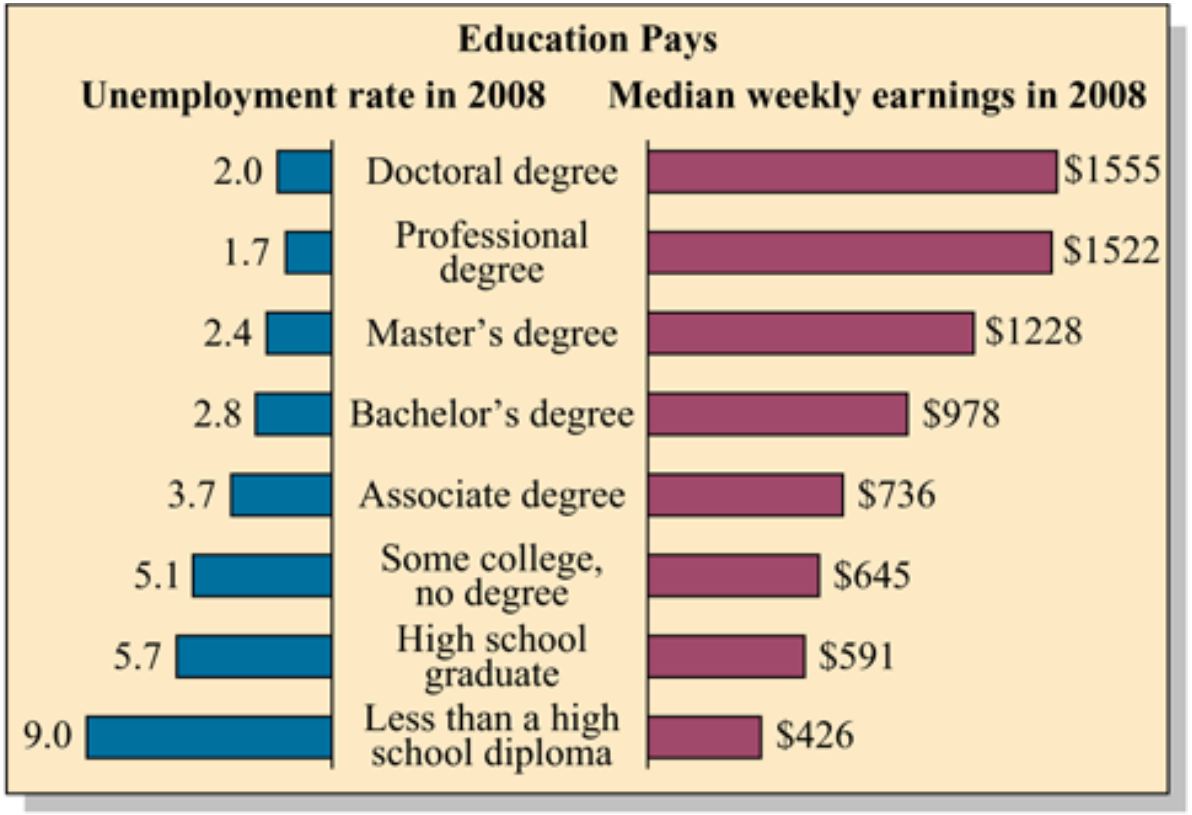
A time-series line chart of stock, bond, and gold prices for an initial $100 investment is shown below.



Multiple bar graph

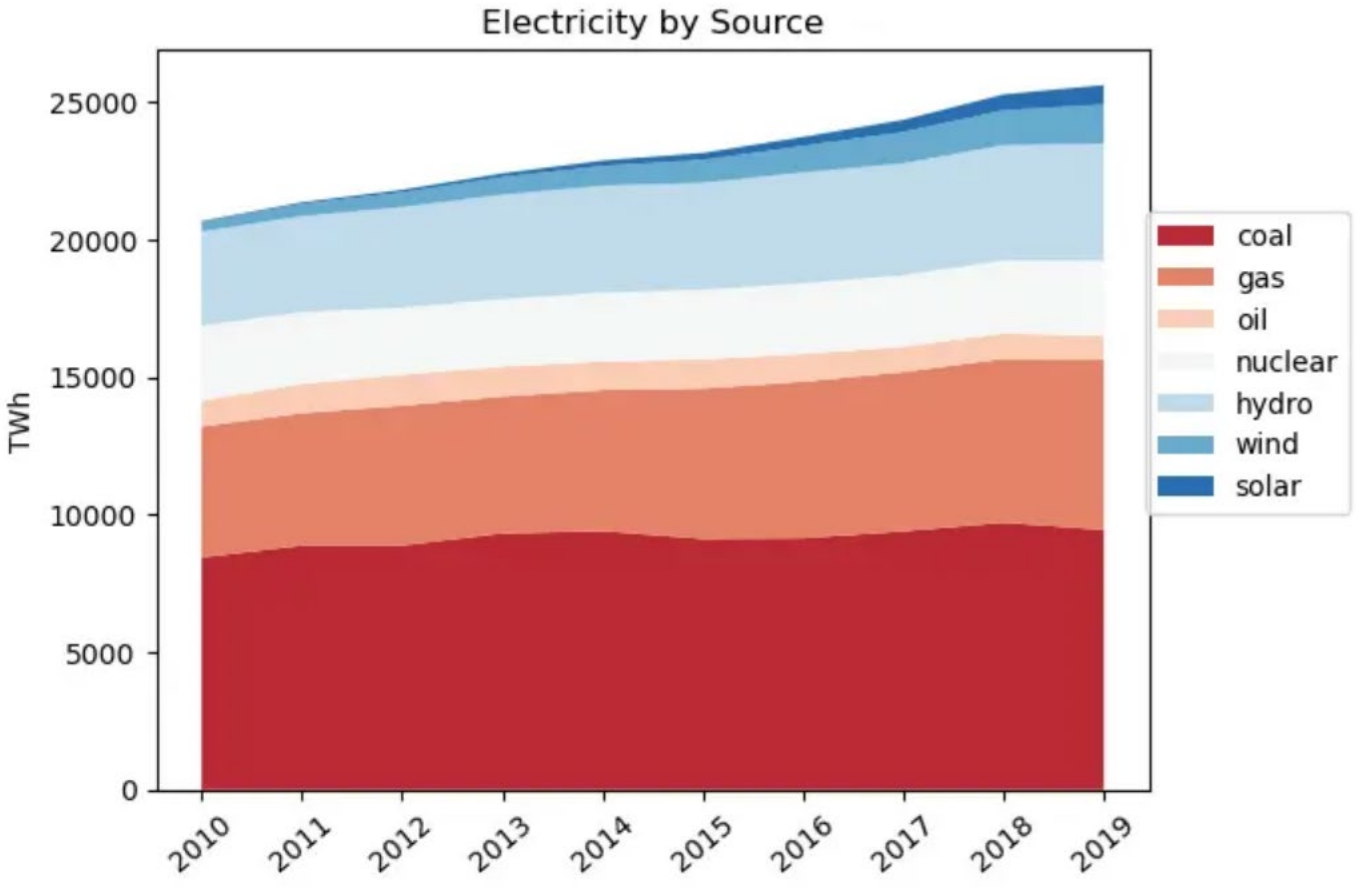
A **multiple bar graph** combines two or more bar graphs to compare two or more data sets.

The following multiple bar graph shows how education affects personal employment.



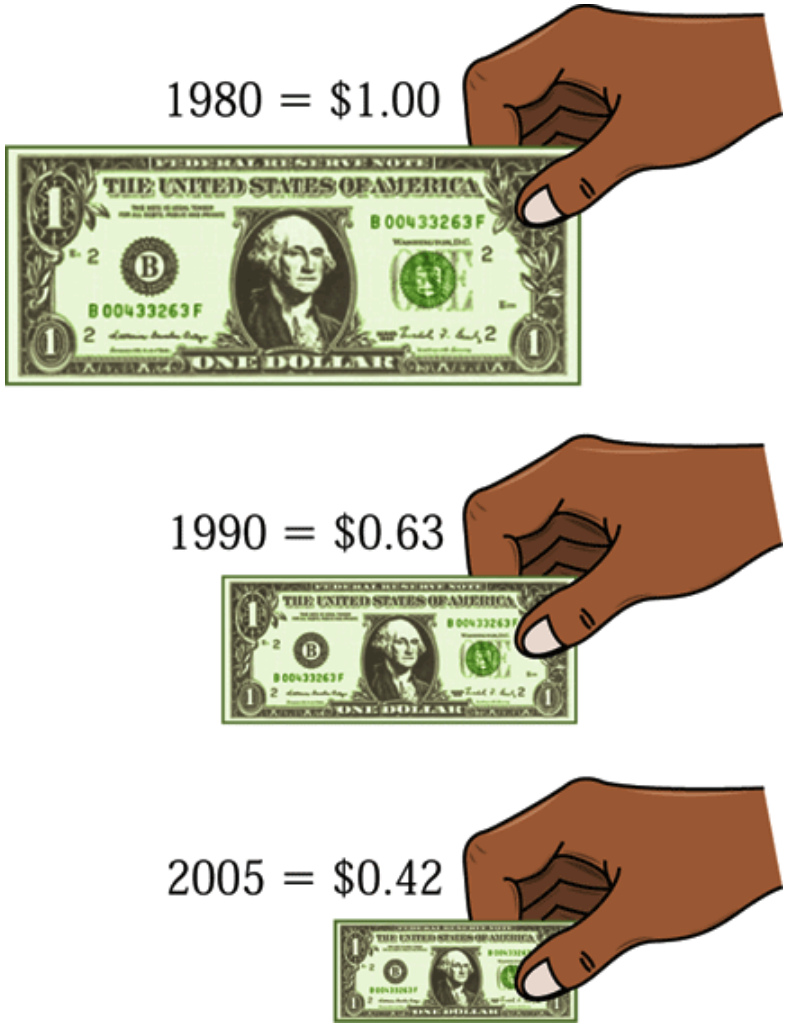
Stack Plot

A **stack plot** displays two or more data sets as vertically stacked areas. Stack plots are useful when both the individual data values and their cumulative value are important.

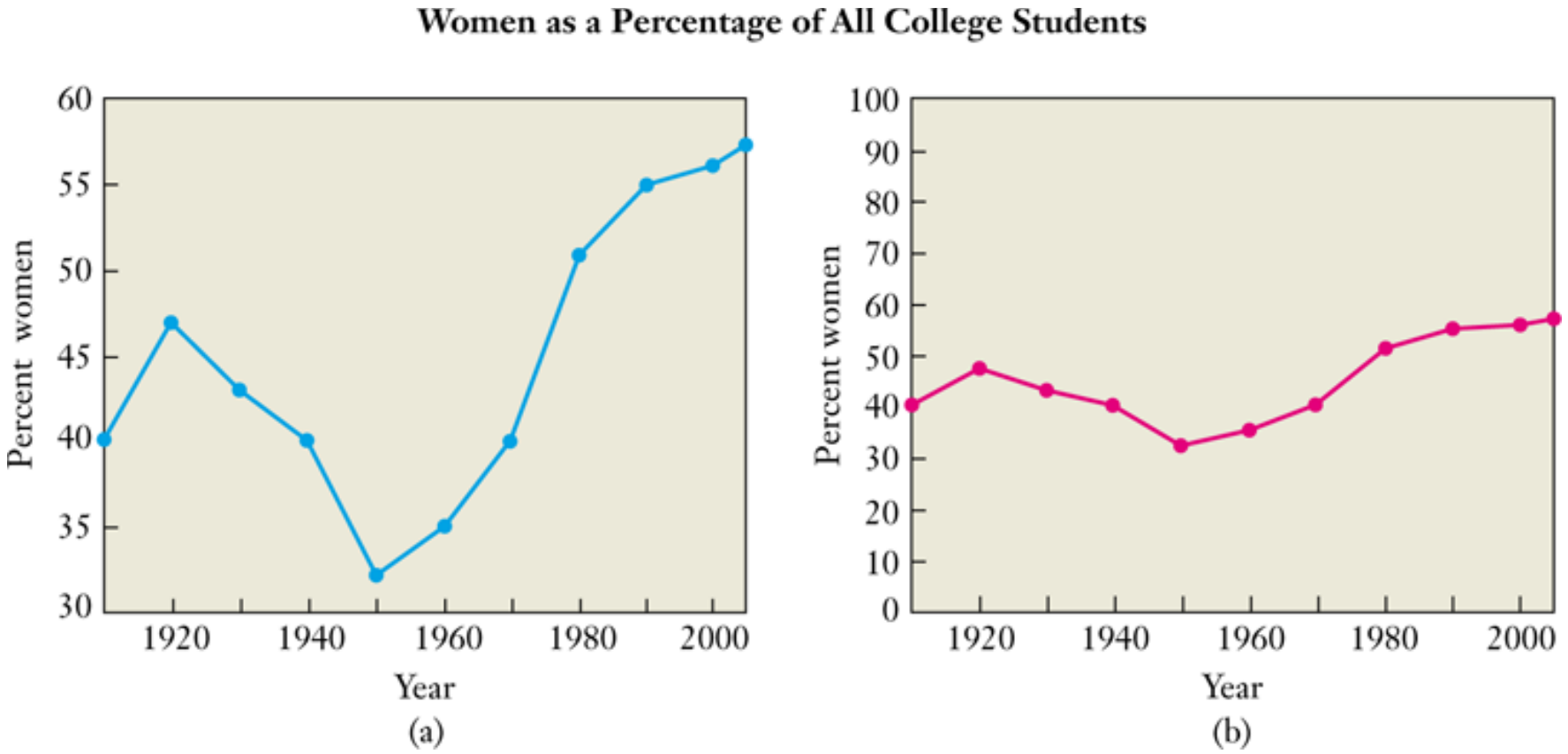


Perceptual Distortion

The lengths of the dollars represent the data, but your eyes tend to focus on the area.

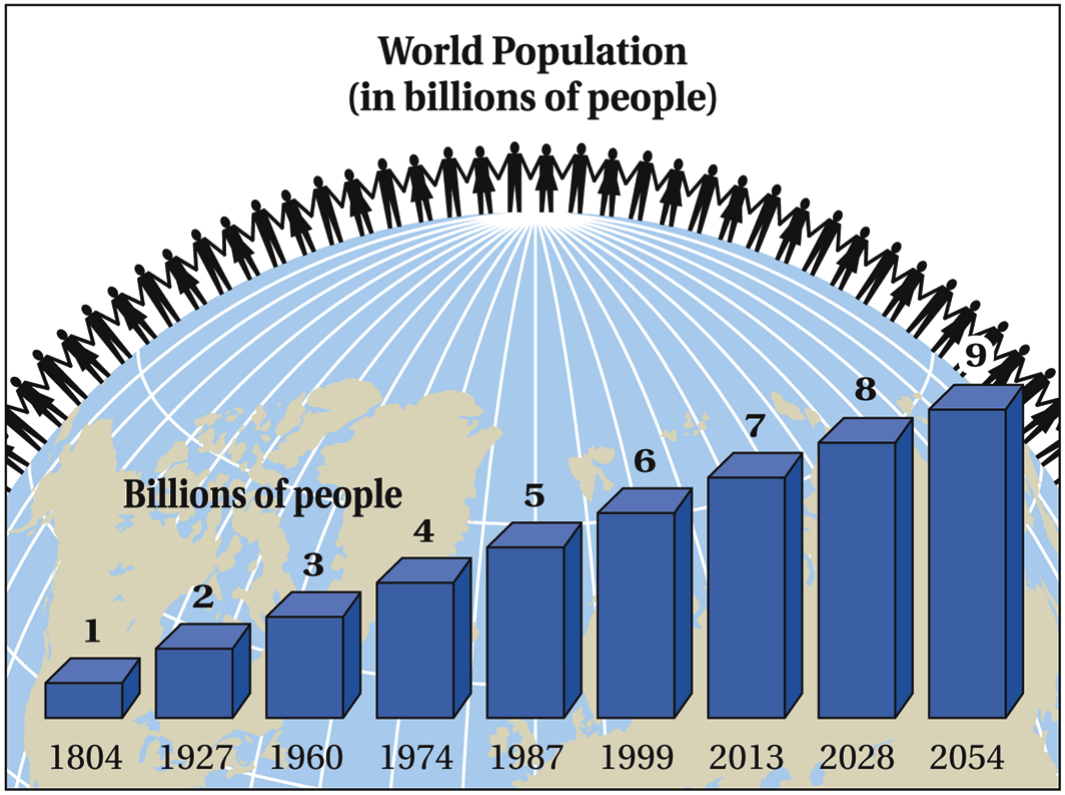


Different Vertical Scales



Both graphs show the same data, but they look very different because their vertical scales have different ranges.

Different Horizontal Scales



It appears that the world population has been rising linearly. However, the time intervals on the horizontal axis are not uniform in size.

Misleading graphs

Well designed graphs help us see patterns, but misleading graphs play tricks with our eyes and lead to wrong conclusions!