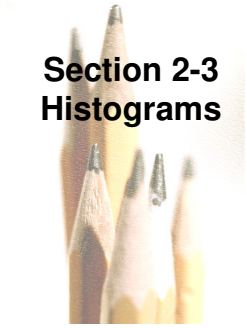
Chapter 2
Summarizing and Graphing Data

2-3 Histograms

2-4 Statistical Graphics

2-5 Critical Thinking: Bad Graphs

Section 2-3
Histograms



Key Concept

We use a visual tool called a **histogram** to analyze the shape of the distribution of the data.

Histogram

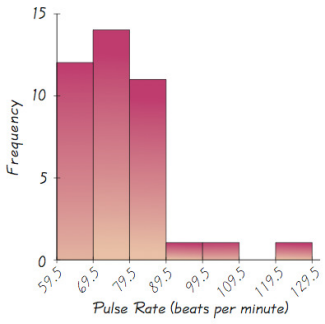
A graph consisting of bars of equal width drawn adjacent to each other (without gaps). The horizontal scale represents the classes of quantitative data values and the vertical scale represents the frequencies. The heights of the bars correspond to the frequency values.

Histogram

Basically a graphic version of a frequency distribution.

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60–69	12
70–79	14
80–89	11
90–99	1
100–109	1
110–119	0
120–129	1



Histogram

The bars on the horizontal scale are labeled with one of the following:

(1) Class boundaries

(2) Class midpoints

(3) Lower class limits (introduces a small error)

Horizontal Scale for Histogram: Use class boundaries or class midpoints.

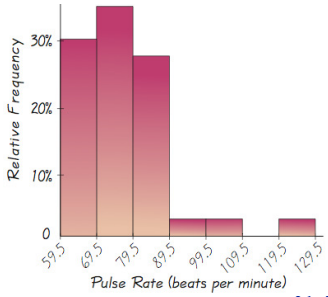
Vertical Scale for Histogram: Use the class frequencies.

Relative Frequency Histogram

Has the same shape and horizontal scale as a histogram, but the vertical scale is marked with relative frequencies instead of actual frequencies

Table 2-3 Relative Frequency Distribution of Pulse Rates of Females

Pulse Rate	Relative Frequency
60-69	30%
70-79	35%
80-89	27.5%
90-99	2.5%
100-109	2.5%
110-119	0
120-129	2.5%



Critical Thinking Interpreting Histograms

Objective is not simply to construct a histogram, but rather to *understand* something about the data.

When graphed, a normal distribution has a "bell" shape. Characteristic of the bell shape are

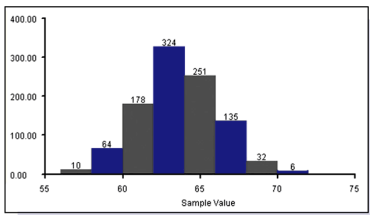
(1) The frequencies increase to a maximum, and then decrease, and

(2) symmetry, with the left half of the graph roughly a mirror image of the right half.

The histogram on the next slide illustrates this.

3

Critical Thinking Interpreting Histograms



Recap

In this Section we have discussed

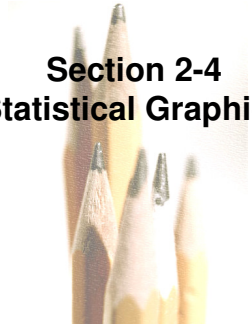
- Histograms
- Relative Frequency Histograms

Frequency Distribution Example

- Health Exam Results: Male – Ages of subjects
- 58 22 32 31 28 46 41 56 20 54 17 73 52 25 29 17 41 52 32 20 20 29 18 26 33 55 53 28 28 37 40 33 26 53 36 34 42 18 44 20
- Construct a Frequency Distribution using the above data with 7 classes.
- Construct the corresponding Frequency Histogram.

4

Section 2-4
Statistical Graphics



Key Concept

This section discusses other types of statistical graphs.

Our objective is to identify a suitable graph for representing the data set. The graph should be effective in revealing the important characteristics of the data.

Frequency Polygon

Uses line segments connected to points directly above class midpoint values

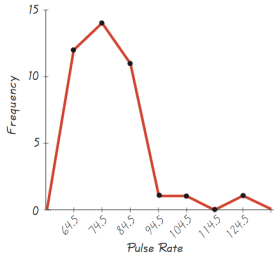


Figure 2-5 Frequency Polygon: Pulse Rates of Women

Relative Frequency Polygon

Uses relative frequencies (proportions or percentages) for the vertical scale.

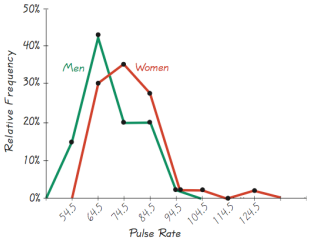


Figure 2-6 Relative Frequency Polygons: Pulse Rates of Women and Men

Ogive

A line graph that depicts cumulative frequencies

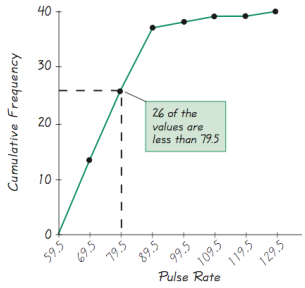
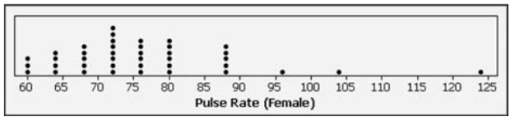


Figure 2-7 Ogive

Dot Plot

Consists of a graph in which each data value is plotted as a point (or dot) along a scale of values. Dots representing equal values are stacked.



6

Stemplot (or Stem-and-Leaf Plot)

Represents quantitative data by separating each value into two parts: the stem (such as the leftmost digit) and the leaf (such as the rightmost digit)

Stemplot

Stem (tens)	Leaves (units)	
6	000444488888	← Data values are 60, 60, 60, 64, . . . , 68.
7	22222222666666	
8	00000088888	
9	6	← Data value is 96.
10	4	← Data value is 104.
11		
12	4	

Pulse Rates of Females

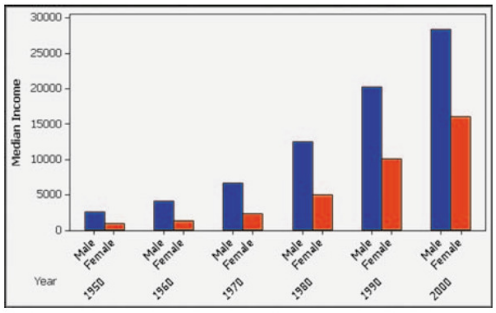
Bar Graph

Uses bars of equal width to show frequencies of categories of qualitative data. Vertical scale represents frequencies or relative frequencies. Horizontal scale identifies the different categories of qualitative data.

A *multiple bar graph* has two or more sets of bars, and is used to compare two or more data sets.

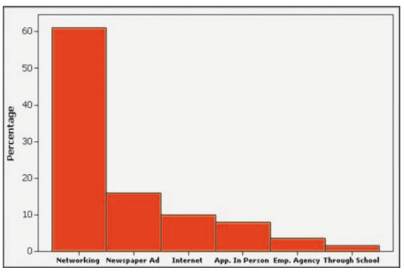
Multiple Bar Graph

Median Income of Males and Females



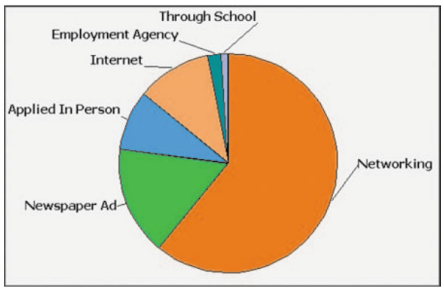
Pareto Chart

A bar graph for qualitative data, with the bars arranged in descending order according to frequencies



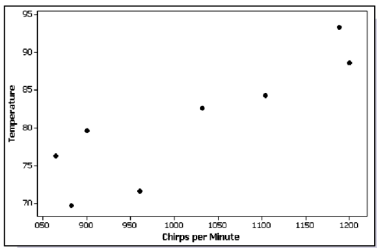
Pie Chart

A graph depicting qualitative data as slices of a circle, size of slice is proportional to frequency count



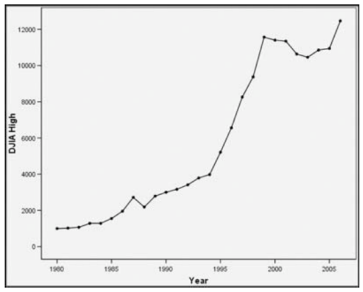
Scatter Plot (or Scatter Diagram)

A plot of paired (*x*,*y*) data with a horizontal *x*-axis and a vertical *y*-axis. Used to determine whether there is a relationship between the two variables



Time-Series Graph

Data that have been collected at different points in time: *time-series data*



Important Principles Suggested by Edward Tufte

For small data sets of 20 values or fewer, use a table instead of a graph.

A graph of data should make the viewer focus on the true nature of the data, not on other elements, such as eye-catching but distracting design features.

Do not distort data, construct a graph to reveal the true nature of the data.

Almost all of the ink in a graph should be used for the data, not the other design elements.

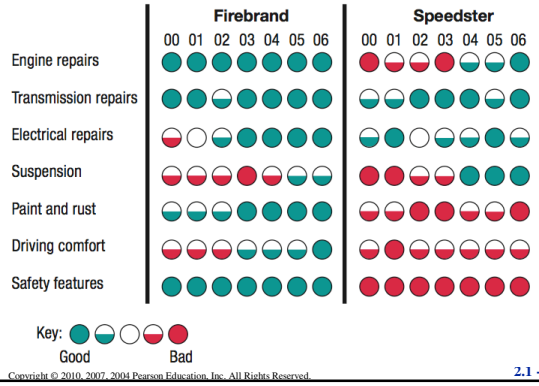
Important Principles Suggested by Edward Tufte

Don't use screening consisting of features such as slanted lines, dots, cross-hatching, because they create the uncomfortable illusion of movement.

Don't use area or volumes for data that are actually one-dimensional in nature. (Don't use drawings of dollar bills to represent budget amounts for different years.)

Never publish pie charts, because they waste ink on nondata components, and they lack appropriate scale.

Car Reliability Data



Recap

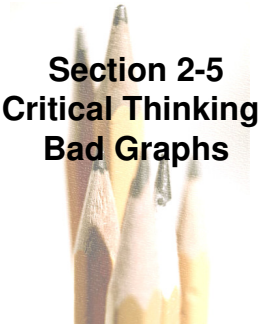
In this section we saw that graphs are excellent tools for describing, exploring and comparing data.

Describing data: Histogram - consider distribution, center, variation, and outliers.

Exploring data: features that reveal some useful and/or interesting characteristic of the data set.

Comparing data: Construct similar graphs to compare data sets.

Section 2-5
Critical Thinking: Bad Graphs



Key Concept

Some graphs are bad in the sense that they contain errors.

Some are bad because they are technically correct, but misleading.

It is important to develop the ability to recognize bad graphs and identify exactly how they are misleading.

Nonzero Axis

Are misleading because one or both of the axes begin at some value other than zero, so that differences are exaggerated.

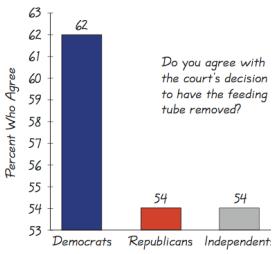


Figure 2-1 Survey Results by Party

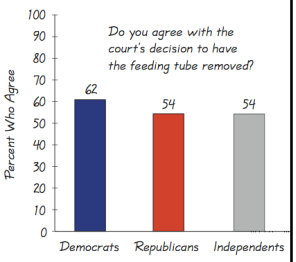


Figure 2-9 Survey Results by Party

Pictographs

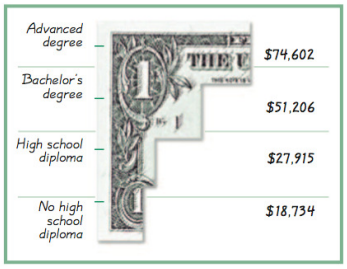
are drawings of objects. Three-dimensional objects - money bags, stacks of coins, army tanks (for army expenditures), people (for population sizes), barrels (for oil production), and houses (for home construction) are commonly used to depict data.

These drawings can create false impressions that distort the data.

If you double each side of a square, the area does not merely double; it increases by a factor of four;if you double each side of a cube, the volume does not merely double; it increases by a factor of eight.

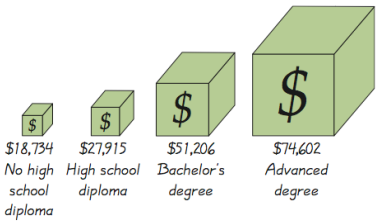
Pictographs using areas and volumes can therefore be very misleading.

Annual Incomes of Groups with Different Education Levels



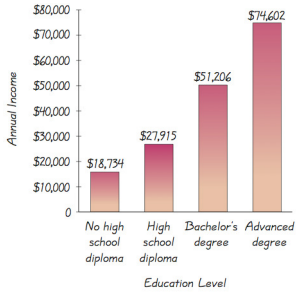
Bars have same width, too busy, too difficult to understand.

Annual Incomes of Groups with Different Education Levels



Misleading. Depicts one-dimensional data with three-dimensional boxes. Last box is 64 times as large as first box, but income is only 4 times as large.

Annual Incomes of Groups with Different Education Levels



Fair, objective, unencumbered by distracting features.