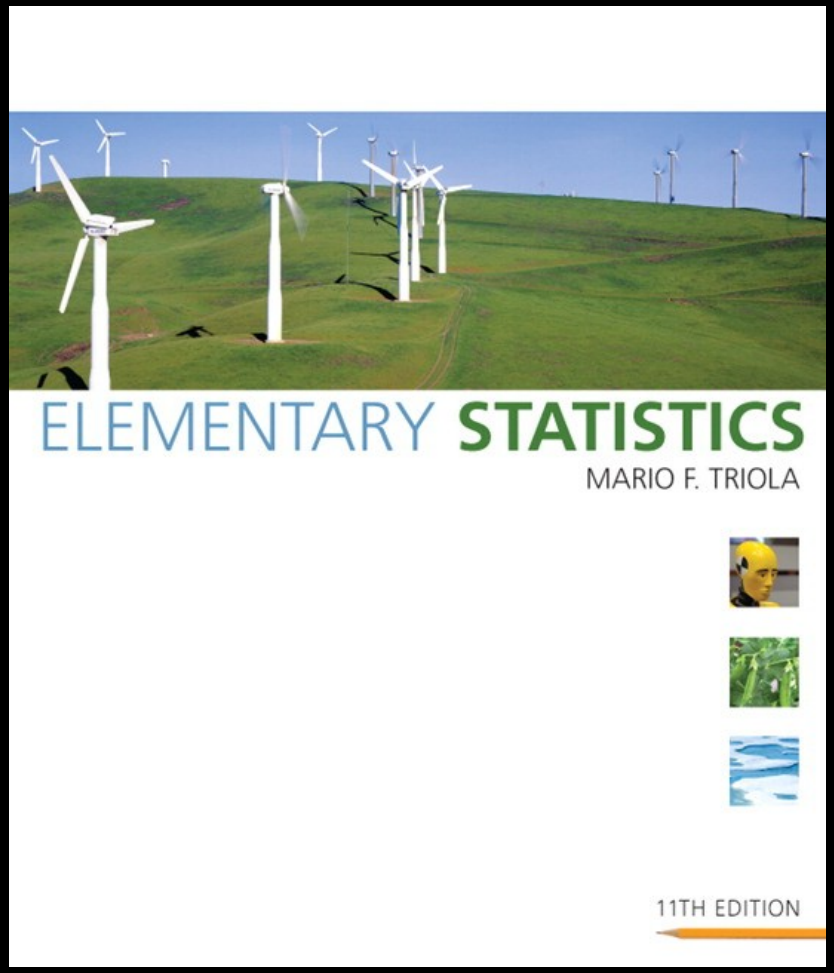
# Lecture Slides




## *Elementary Statistics*
## Eleventh Edition

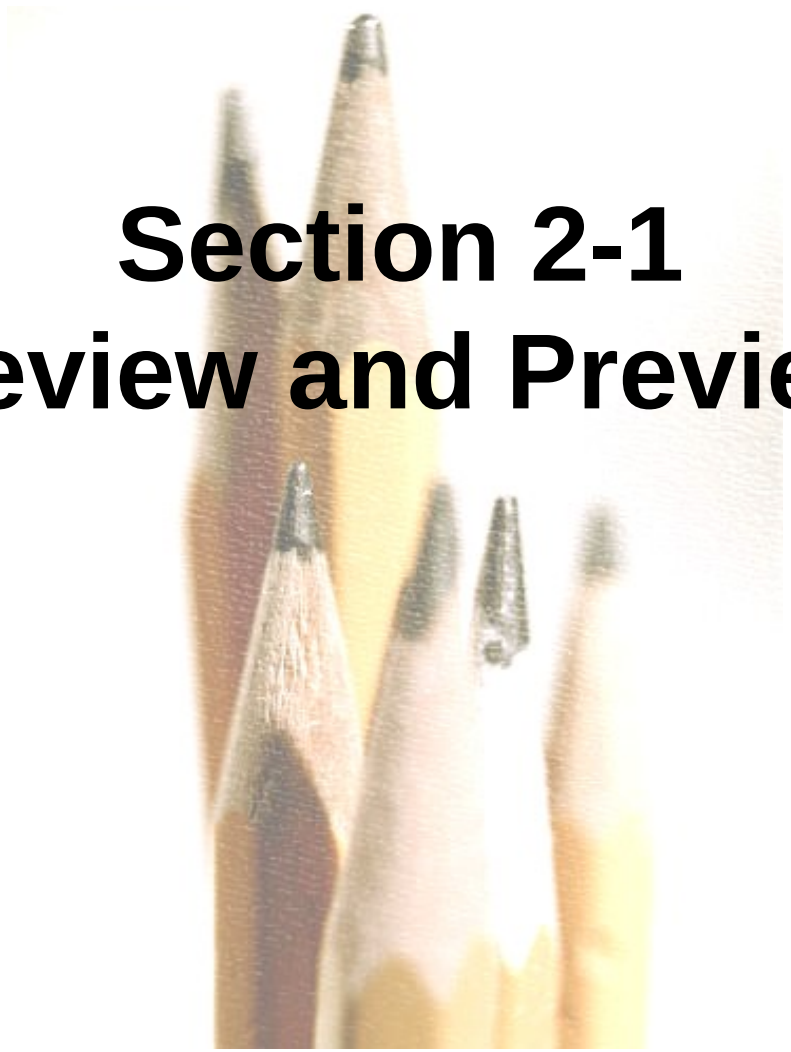
and the Triola Statistics Series

by Mario F. Triola

PEARSON

# Chapter 2
# Summarizing and Graphing Data

# Section 2-1
# Review and Preview



# Preview
# Important Characteristics of Data

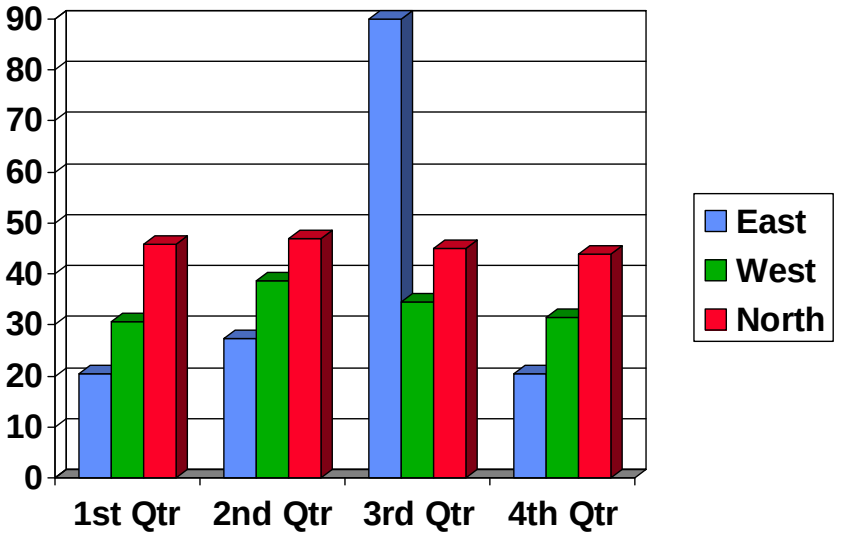
1. **Center**: A representative or average value that indicates where the middle of the data set is located.

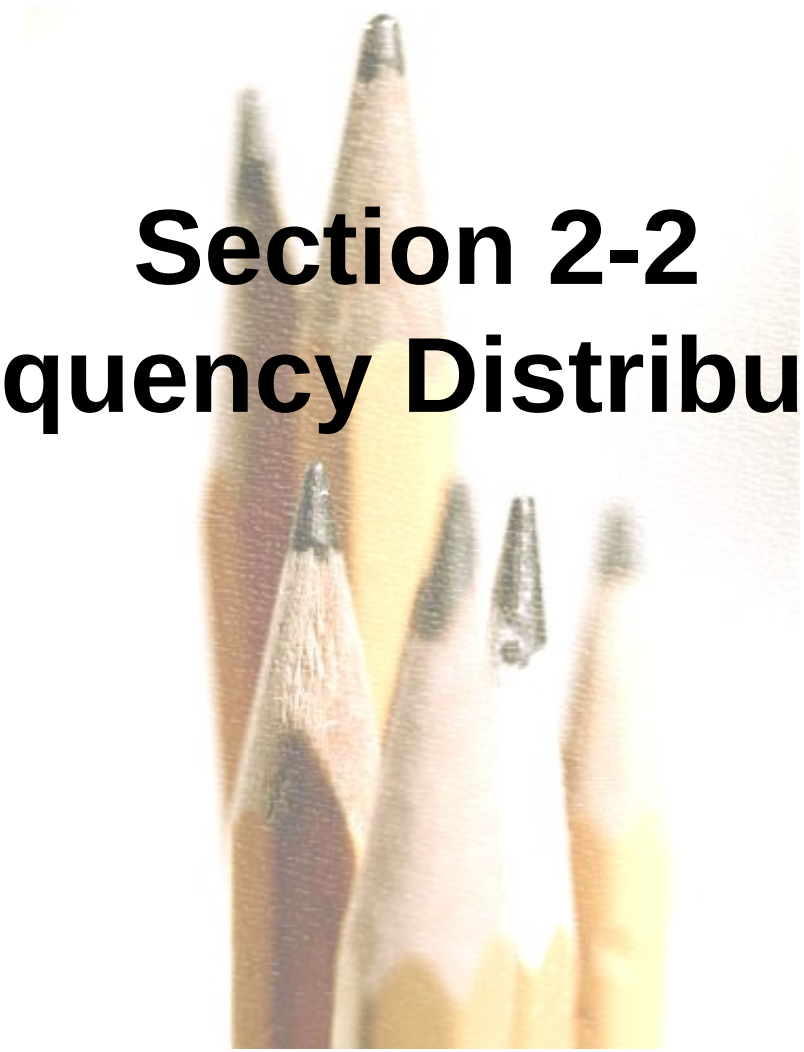
2. **Variation**: A measure of the amount that the data values vary.

3. **Distribution**: The nature or shape of the spread of data over the range of values (such as bell-shaped, uniform, or skewed).

4. **Outliers**: Sample values that lie very far away from the vast majority of other sample values.

5. **Time**: Changing characteristics of the data over time.





# Section 2-2
# Frequency Distributions

# Key Concept

**When working with large data sets, it is often helpful to organize and summarize data by constructing a table called a frequency distribution, defined later. Because computer software and calculators can generate frequency distributions, the details of constructing them are not as important as what they tell us about data sets. It helps us understand the nature of the *distribution* of a data set.**

# Definition

- **Frequency Distribution (or Frequency Table)**

  **shows how a data set is partitioned among all of several categories (or classes) by listing all of the categories along with the number of data values in each of the categories.**

# Pulse Rates of Females and Males

## Original Data

**Table 2-1** Pulse Rates (beats per minute) of Females and Males

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Females** | | | | | | | | | | | | | | | | | | |
| 76 | 72 | 88 | 60 | 72 | 68 | 80 | 64 | 68 | 68 | 80 | 76 | 68 | 72 | 96 | 72 | 68 | 72 | 64 | 80 |
| 64 | 80 | 76 | 76 | 76 | 80 | 104 | 88 | 60 | 76 | 72 | 72 | 88 | 80 | 60 | 72 | 88 | 88 | 124 | 64 |
| **Males** | | | | | | | | | | | | | | | | | | |
| 68 | 64 | 88 | 72 | 64 | 72 | 60 | 88 | 76 | 60 | 96 | 72 | 56 | 64 | 60 | 64 | 84 | 76 | 84 | 88 |
| 72 | 56 | 68 | 64 | 60 | 68 | 60 | 60 | 56 | 84 | 72 | 84 | 88 | 56 | 64 | 56 | 56 | 60 | 64 | 72 |

# Frequency Distribution
# Pulse Rates of Females

**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 |

**The *frequency* for a particular class is the number of original values that fall into that class.**

# Frequency Distributions

## Definitions

# Lower Class Limits

**are the smallest numbers that can actually belong to different classes**

**Lower Class Limits**

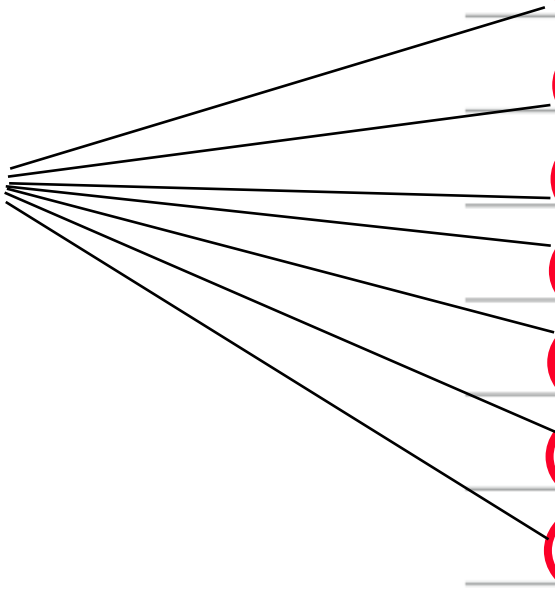


**Table 2-2 Pulse Rates of Females**

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

# Upper Class Limits

**are the largest numbers that can actually belong to different classes**

**Upper Class Limits**

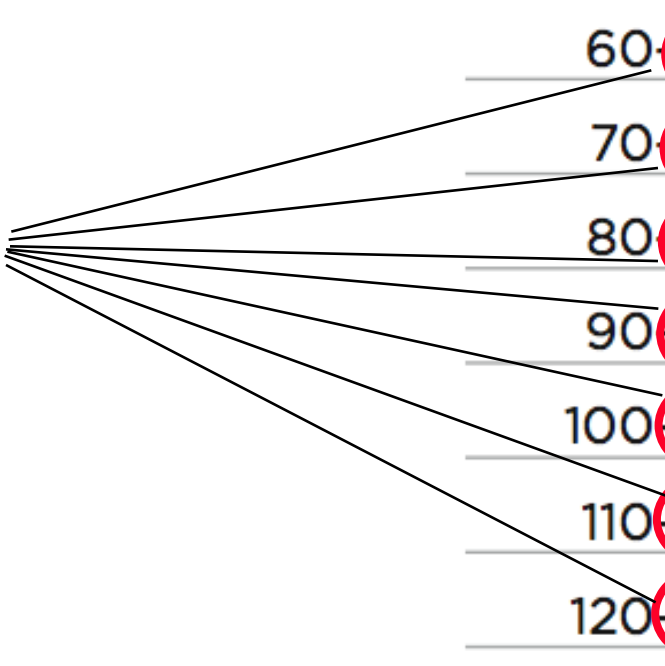


**Table 2-2 Pulse Rates of Females**

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

# Class Boundaries

**are the numbers used to separate classes, but without the gaps created by class limits**

**Class Boundaries**

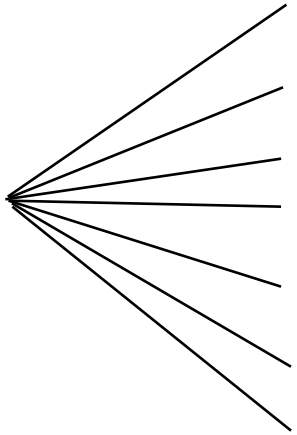


Table 2-2 Pulse Rates of Females

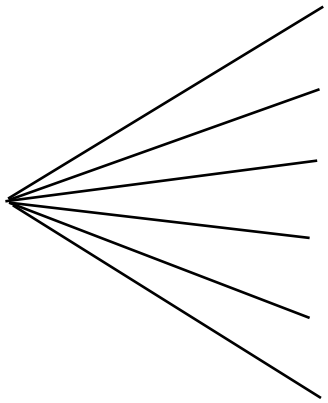
| Pulse Rate | Frequency | Class Boundaries |
|---|---|---|
| | | 59.5 |
| 60–69 | 12 | |
| | | 69.5 |
| 70–79 | 14 | |
| | | 79.5 |
| 80–89 | 11 | |
| | | 89.5 |
| 90–99 | 1 | |
| | | 99.5 |
| 100–109 | 1 | |
| | | 109.5 |
| 110–119 | 0 | |
| | | 119.5 |
| 120–129 | 1 | |
| | | 129.5 |

# Class Midpoints

**are the values in the middle of the classes and can be found by adding the lower class limit to the upper class limit and dividing the sum by two**

**Table 2-2 Pulse Rates of Females**

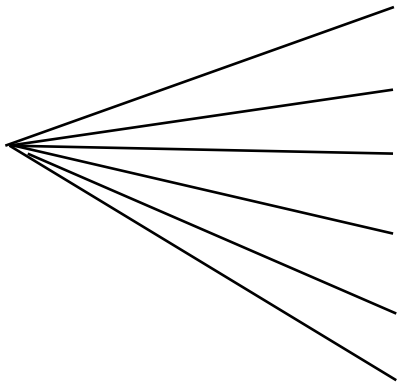
| Class Midpoints | Pulse Rate | Frequency |
|---|---|---|
| 64.5 | 60–69 | 12 |
| 74.5 | 70–79 | 14 |
| 84.5 | 80–89 | 11 |
| 94.5 | 90–99 | 1 |
| 104.5 | 100–109 | 1 |
| 114.5 | 110–119 | 0 |
| 124.5 | 120–129 | 1 |



# Class Width

**is the difference between two consecutive lower class limits or two consecutive lower class boundaries**

**Class Width**



| | Pulse Rate | Frequency |
|---|---|---|
| | 60–69 | 12 |
| 10 | 70–79 | 14 |
| 10 | 80–89 | 11 |
| 10 | 90–99 | 1 |
| 10 | 100–109 | 1 |
| 10 | 110–119 | 0 |
| 10 | 120–129 | 1 |

Table 2-2 Pulse Rates of Females

# Reasons for Constructing Frequency Distributions

1. Large data sets can be summarized.

2. We can analyze the nature of data.

3. We have a basis for constructing important graphs.

# Constructing A Frequency Distribution

1. Determine the number of classes (should be between 5 and 20).

2. Calculate the class width (round up).

$$\text{class width} \approx \frac{\text{(maximum value) – (minimum value)}}{\text{number of classes}}$$

3. Starting point: Choose the minimum data value or a convenient value below it as the first lower class limit.

4. Using the first lower class limit and class width, proceed to list the other lower class limits.

5. List the lower class limits in a vertical column and proceed to enter the upper class limits.

6. Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to get the frequency.

# Relative Frequency Distribution

**includes the same class limits as a frequency distribution, but the frequency of a class is replaced with a relative frequencies (a proportion) or a percentage frequency ( a percent)**

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}}$$

$$\text{percentage frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}} \times 100\%$$

# Relative 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 |

**Total Frequency = 40**

**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% |

*** 12/40 × 100 = 30%**

# Cumulative 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 |

**Table 2-4 Cumulative Frequency Distribution of Pulse Rates of Females**

| Pulse Rate | Cumulative Frequency |
|---|---|
| Less than 70 | 12 |
| Less than 80 | 26 |
| Less than 90 | 37 |
| Less than 100 | 38 |
| Less than 110 | 39 |
| Less than 120 | 39 |
| Less than 130 | 40 |

Cumulative Frequencies

# Frequency Tables

**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 |

**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% |

**Table 2-4 Cumulative Frequency Distribution of Pulse Rates of Females**

| Pulse Rate | Cumulative Frequency |
|---|---|
| Less than 70 | 12 |
| Less than 80 | 26 |
| Less than 90 | 37 |
| Less than 100 | 38 |
| Less than 110 | 39 |
| Less than 120 | 39 |
| Less than 130 | 40 |

# Critical Thinking Interpreting Frequency Distributions

**In later chapters, there will be frequent reference to data with a normal distribution. One key characteristic of a normal distribution is that it has a "bell" shape.**

- **The frequencies start low, then increase to one or two high frequencies, then decrease to a low frequency.**

- **The distribution is approximately symmetric, with frequencies preceding the maximum being roughly a mirror image of those that follow the maximum.**

# Gaps

## Gaps

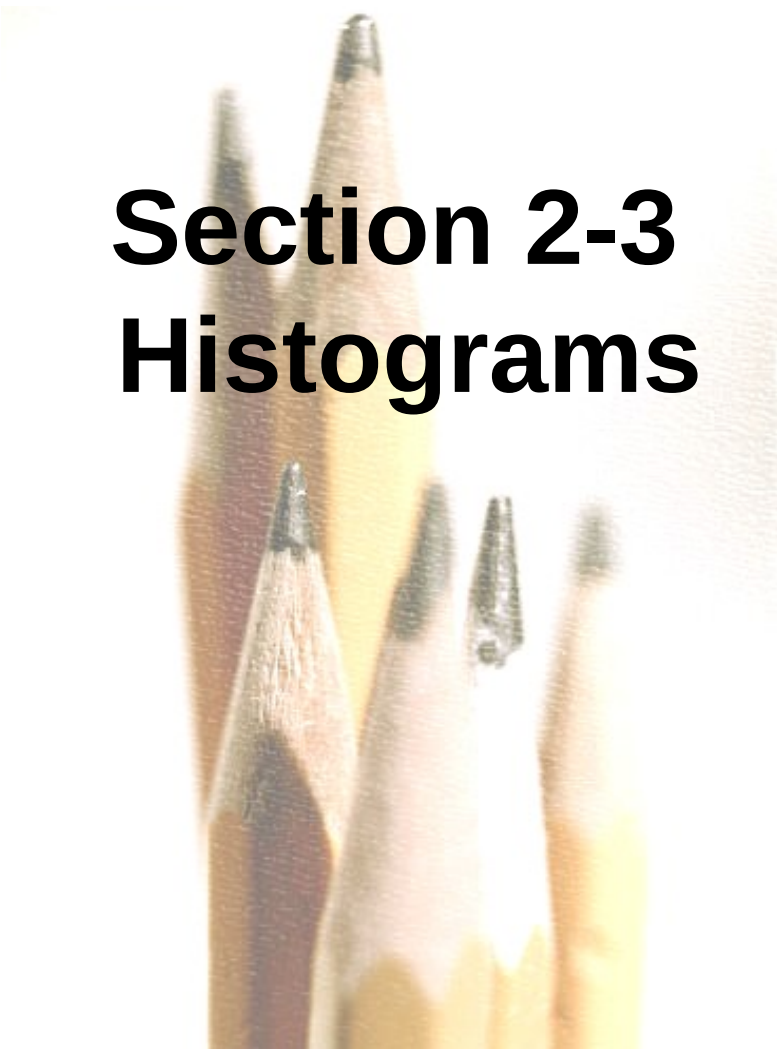
***The presence of gaps can show that we have data from two or more different populations.***
**However, the converse is not true, because data from different populations do not necessarily result in gaps.**

# Recap

**In this Section we have discussed**

- **Important characteristics of data**
- **Frequency distributions**
- **Procedures for constructing frequency distributions**
- **Relative frequency distributions**
- **Cumulative frequency distributions**

# 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

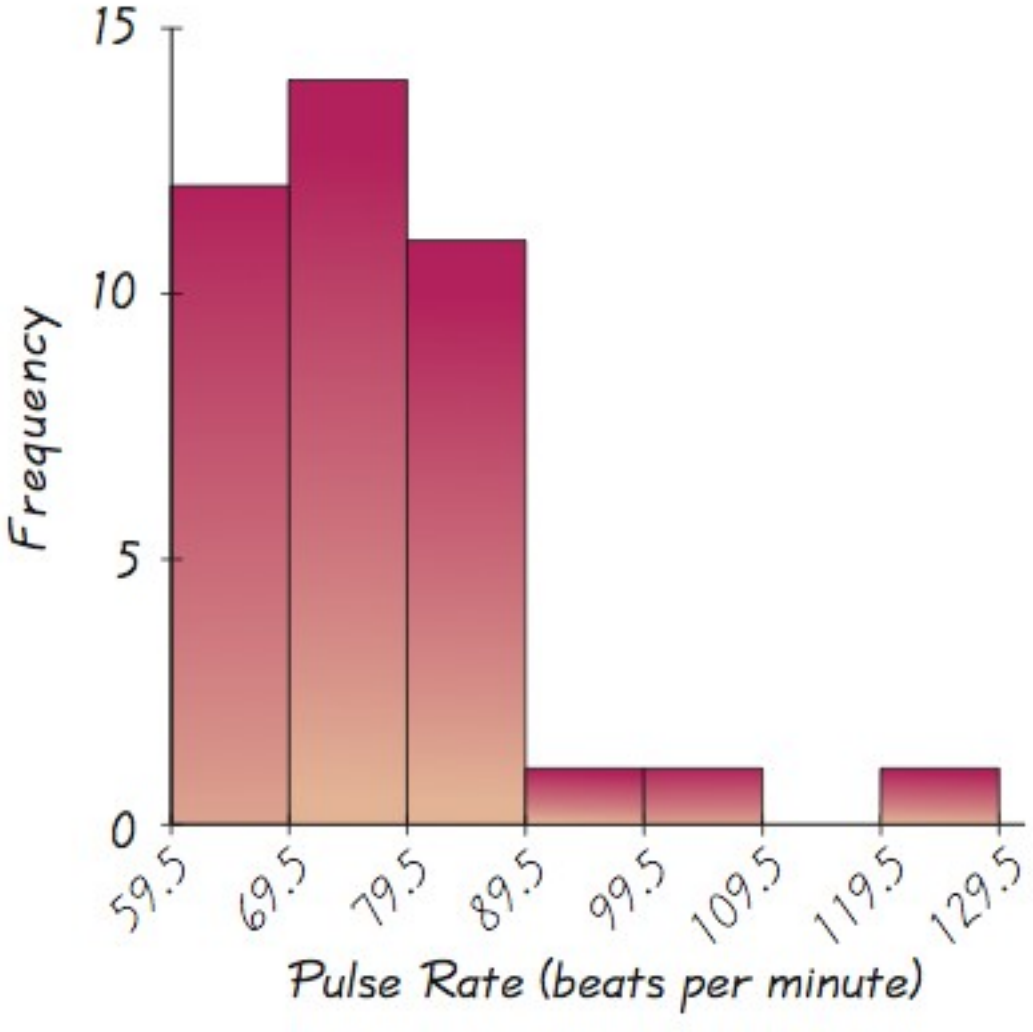
**A graph consisting of bars of equal width drawn adjacent to each other (without gaps). The horizontal scale represents the classes of quantitative data values and the vertical scale represents the frequencies. The heights of the bars correspond to the frequency values.**

# 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.**

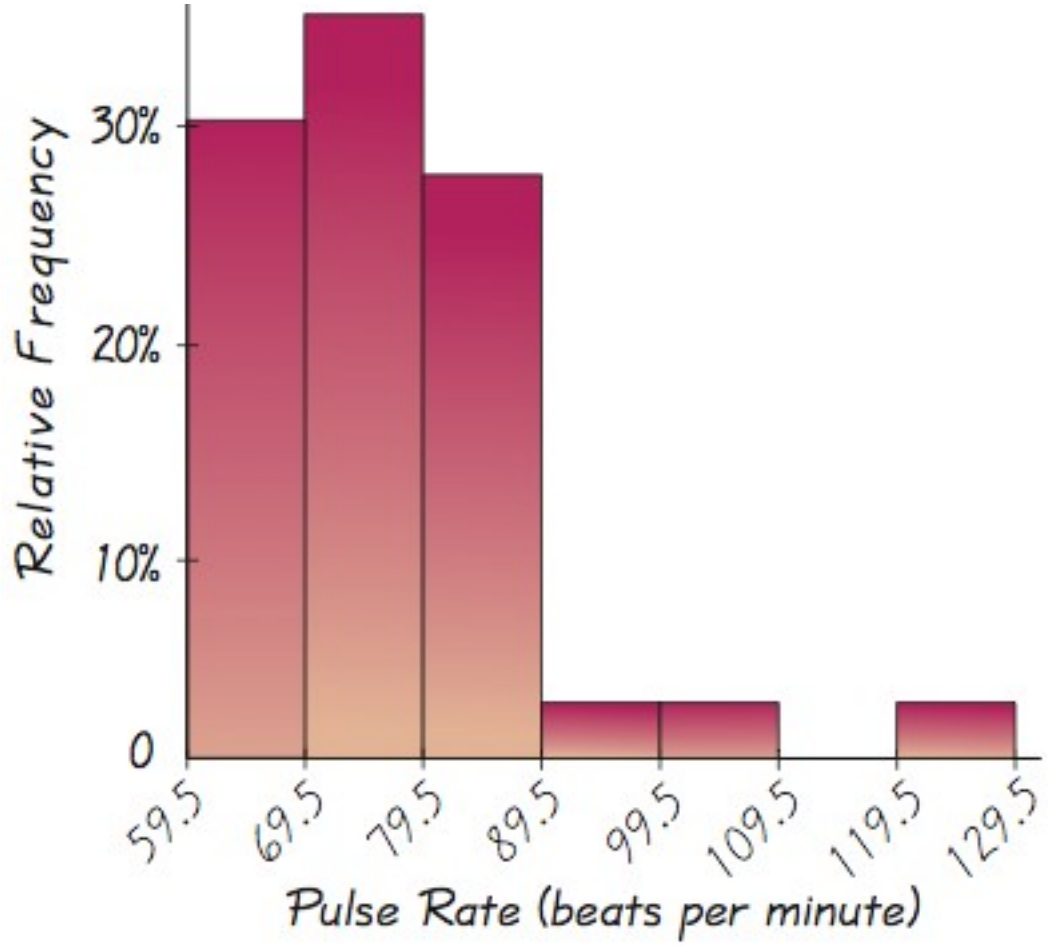
**Vertical Scale for Histogram: Use the class frequencies.**

# 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.**

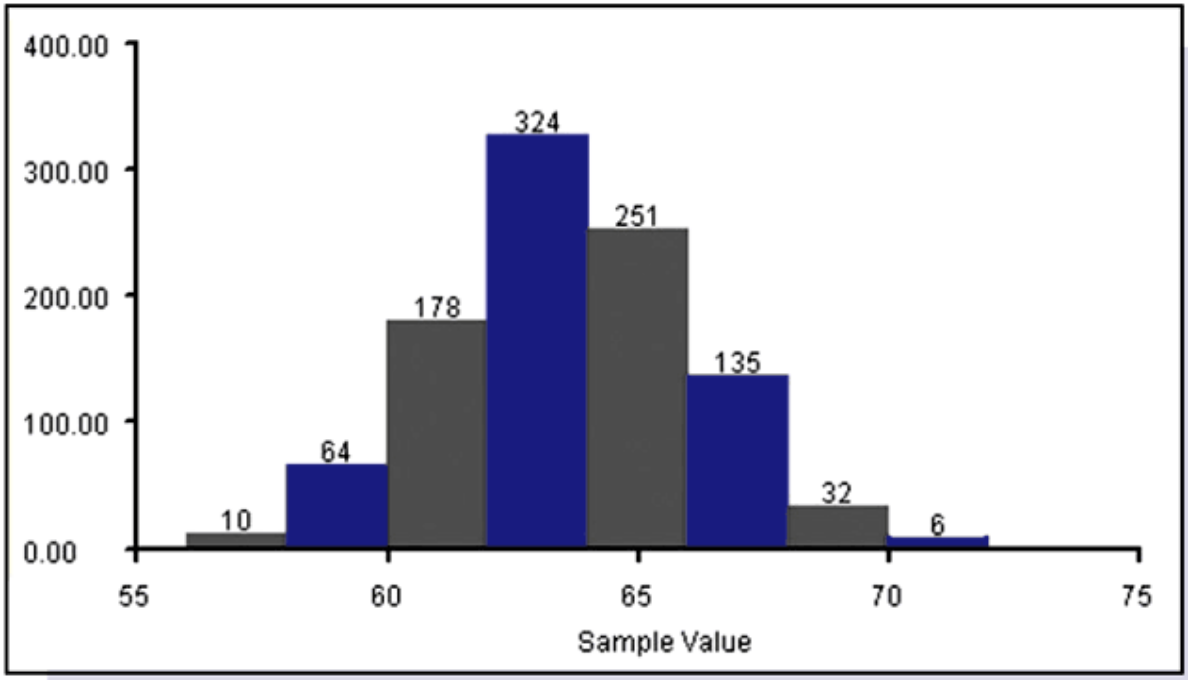
**When graphed, a normal distribution has a "bell" shape. Characteristic of the bell shape are**

**(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.**

# Critical Thinking Interpreting Histograms

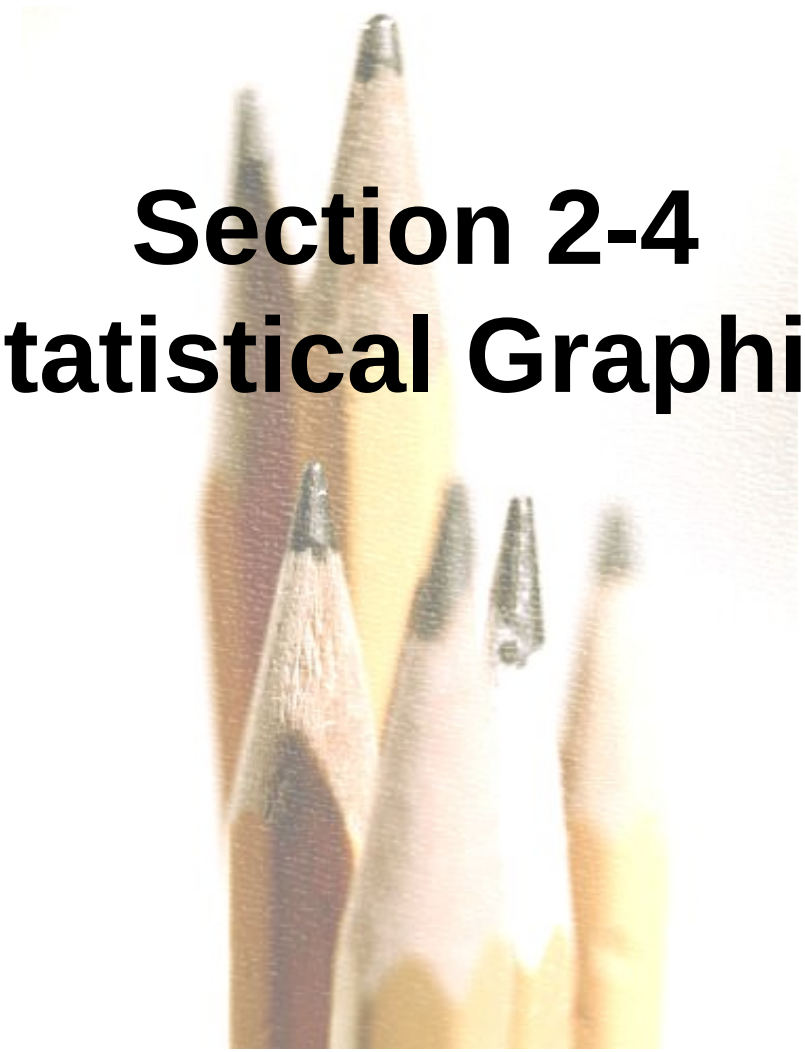


# Recap

**In this Section we have discussed**

- **Histograms**
- **Relative Frequency Histograms**

# 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**

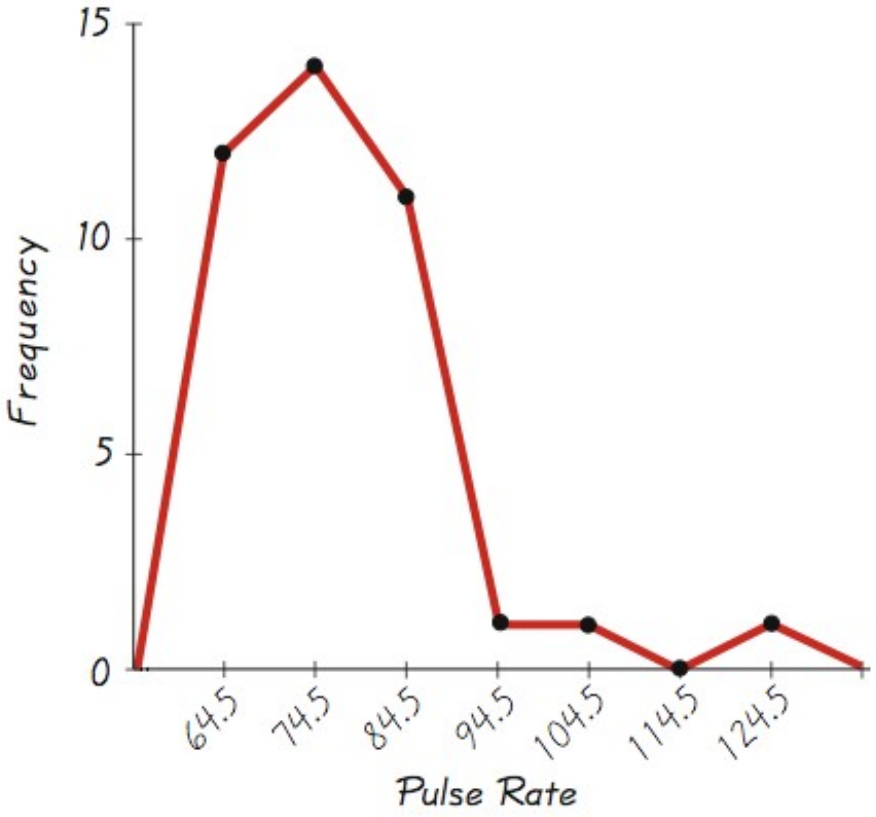



**Figure 2-5 Frequency Polygon: Pulse Rates of Women**

# Relative Frequency Polygon

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

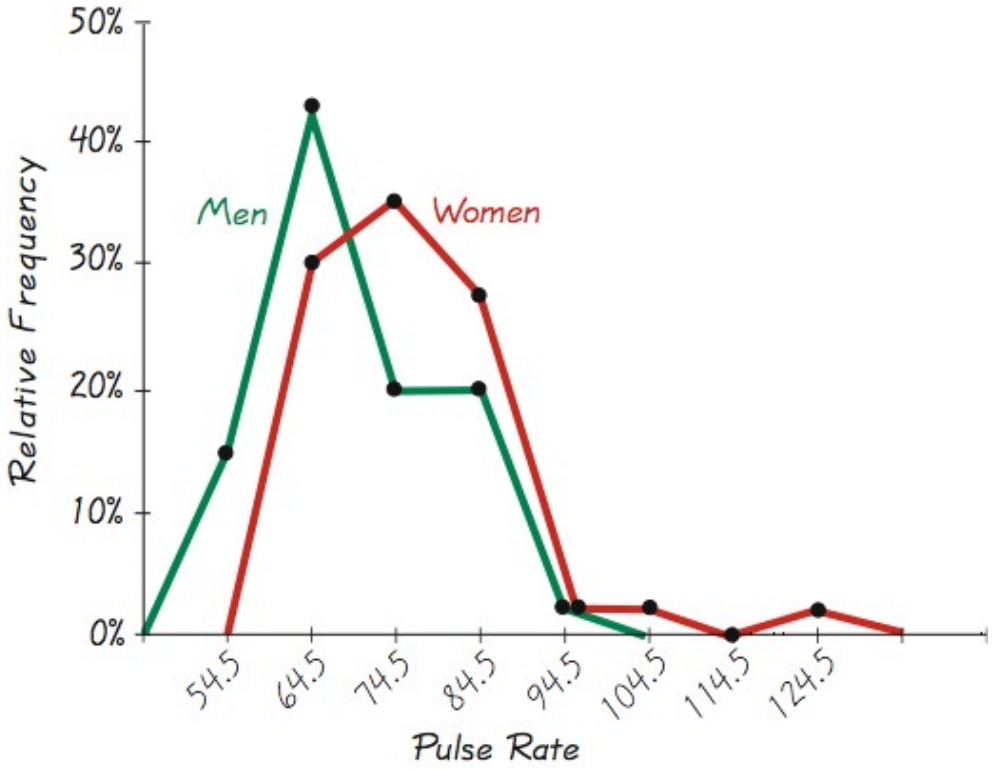



**Figure 2-6 Relative Frequency Polygons: Pulse Rates of Women and Men**

# Ogive

**A line graph that depicts cumulative frequencies**

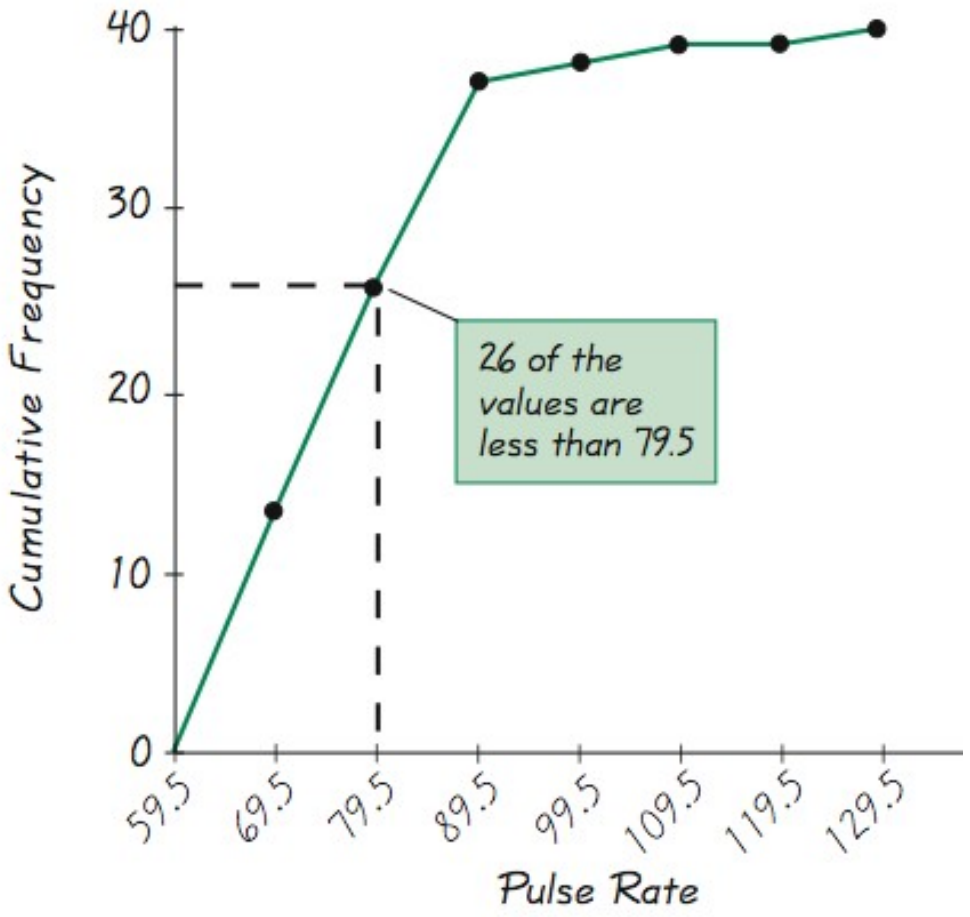
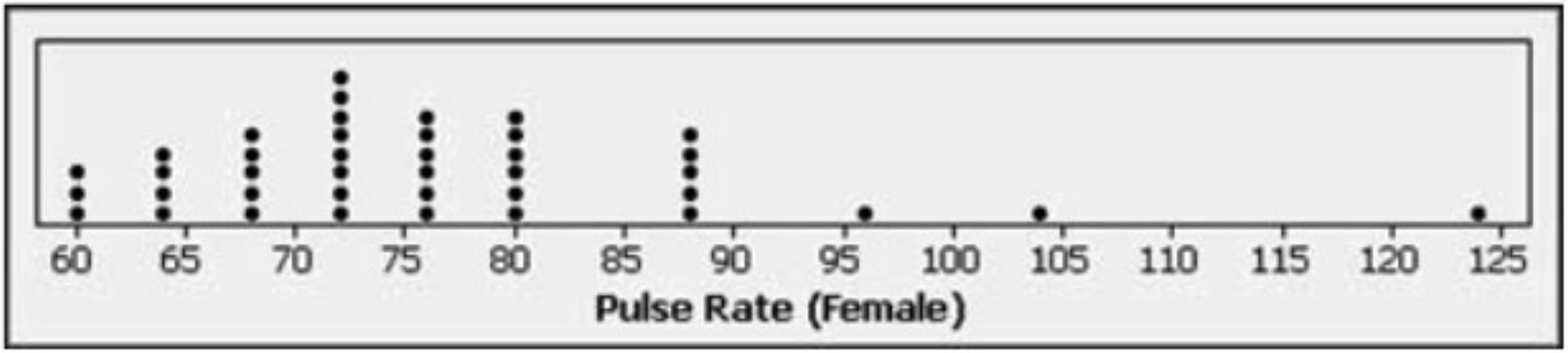



**Figure 2-7 Ogive**

# 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.**



# 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

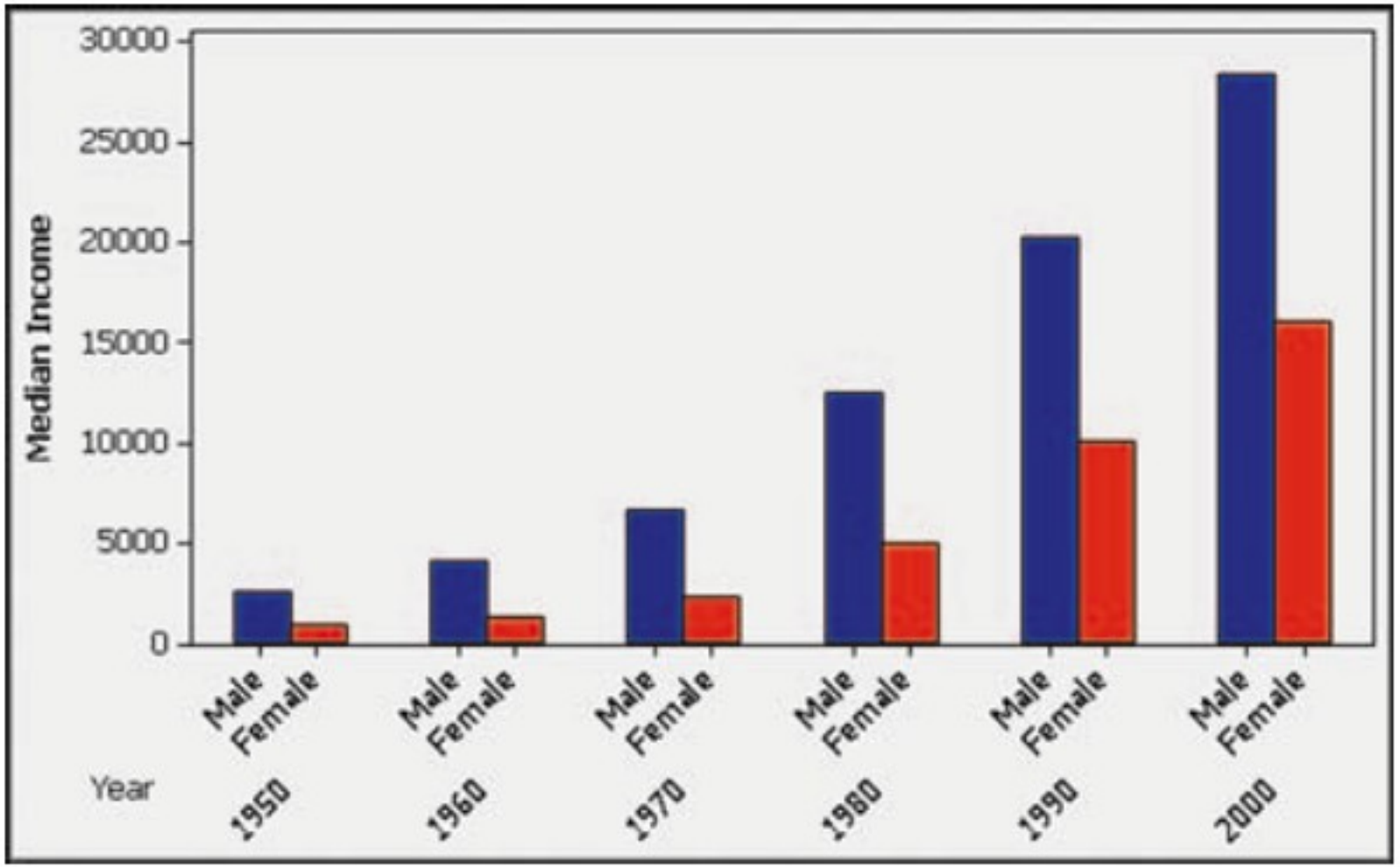
# Bar Graph

**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.**

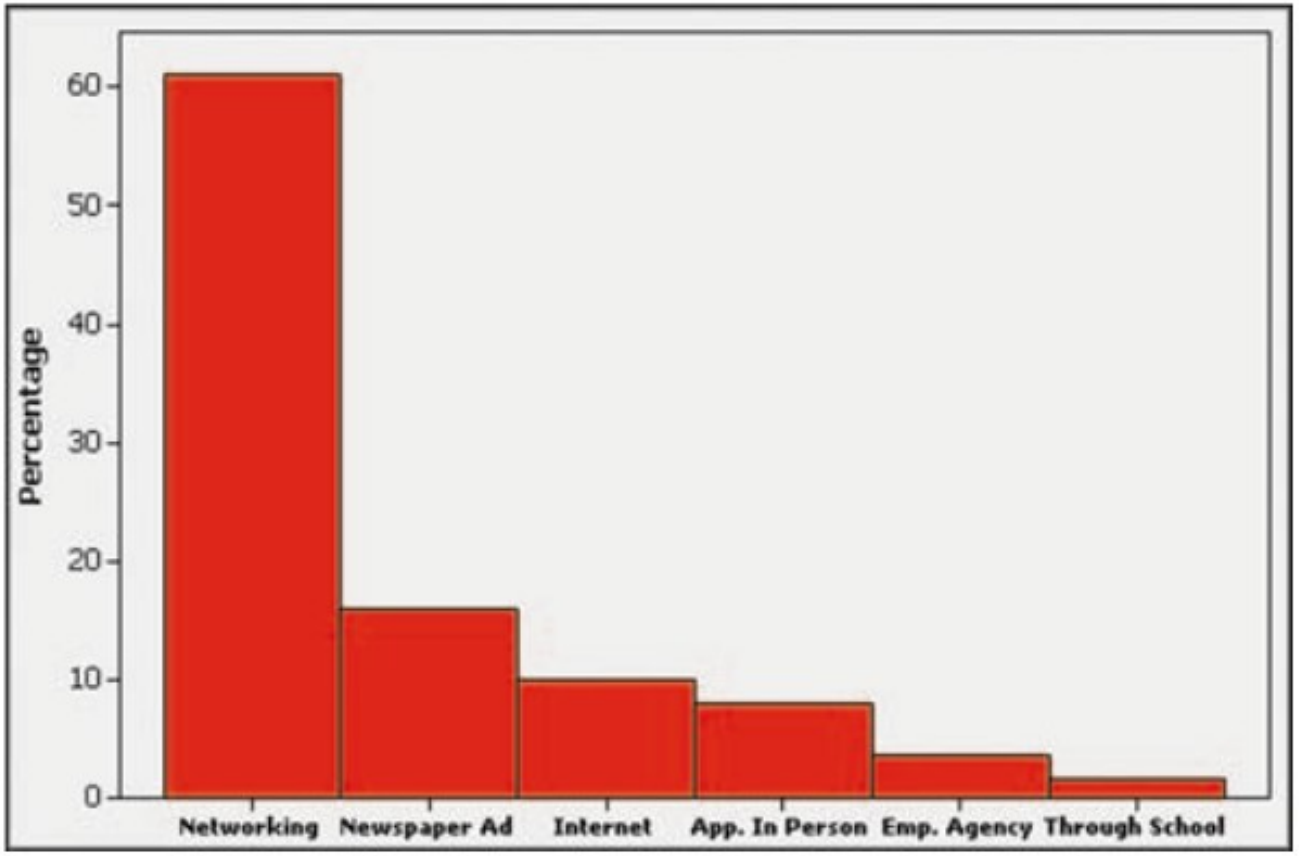
# Multiple Bar Graph

Median Income of Males and Females



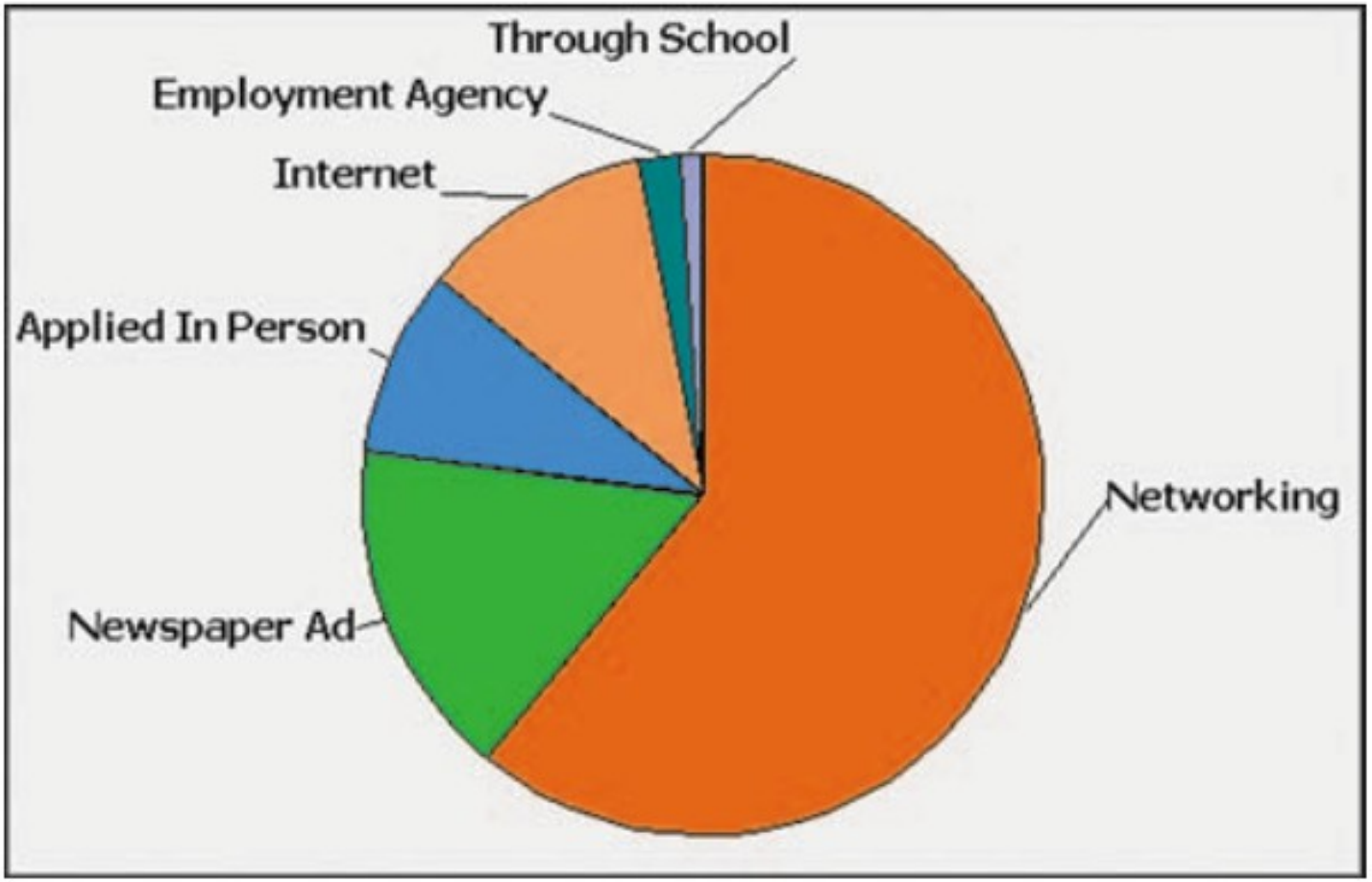
# Pareto Chart

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



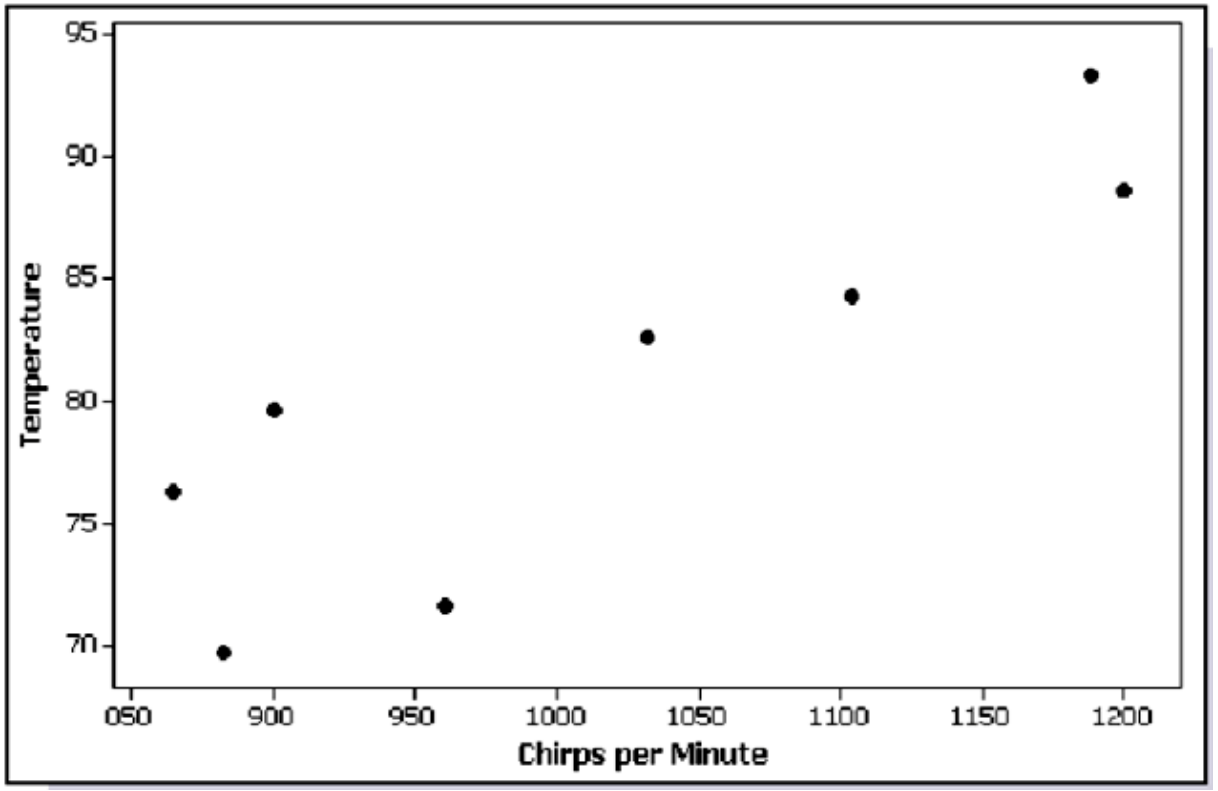
# Pie Chart

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



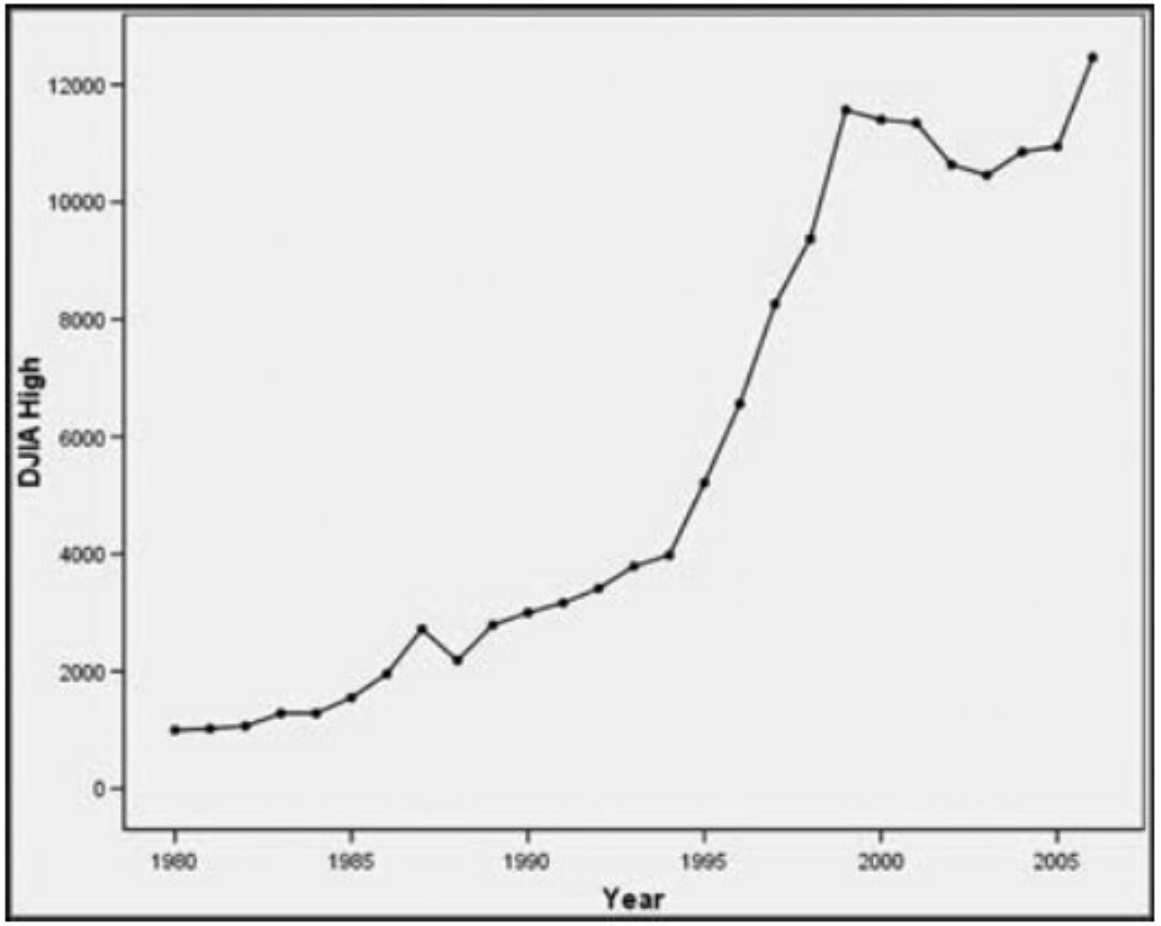
# Scatter Plot (or Scatter Diagram)

**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.**

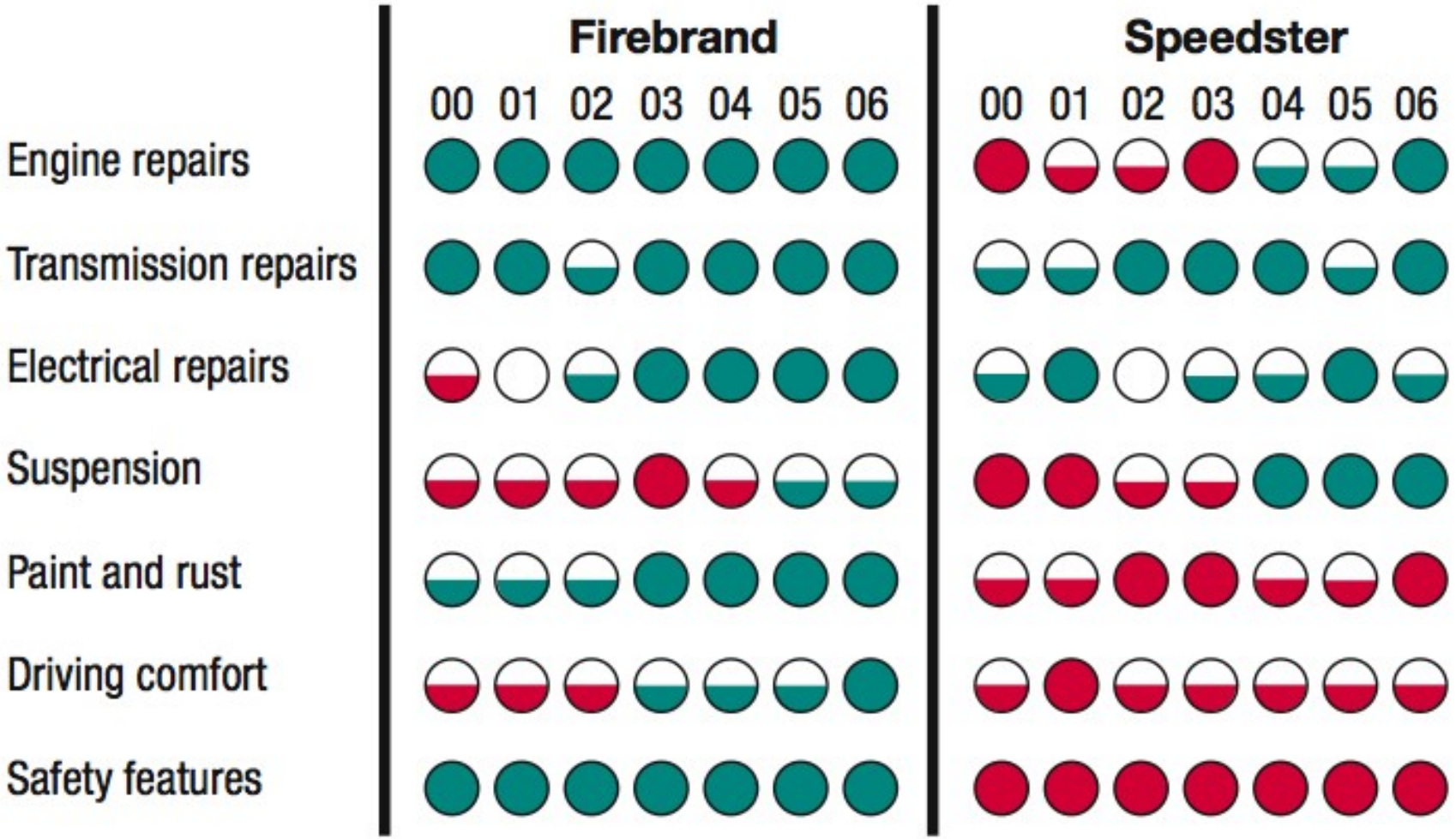
# Important Principles Suggested by Edward Tufte

**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



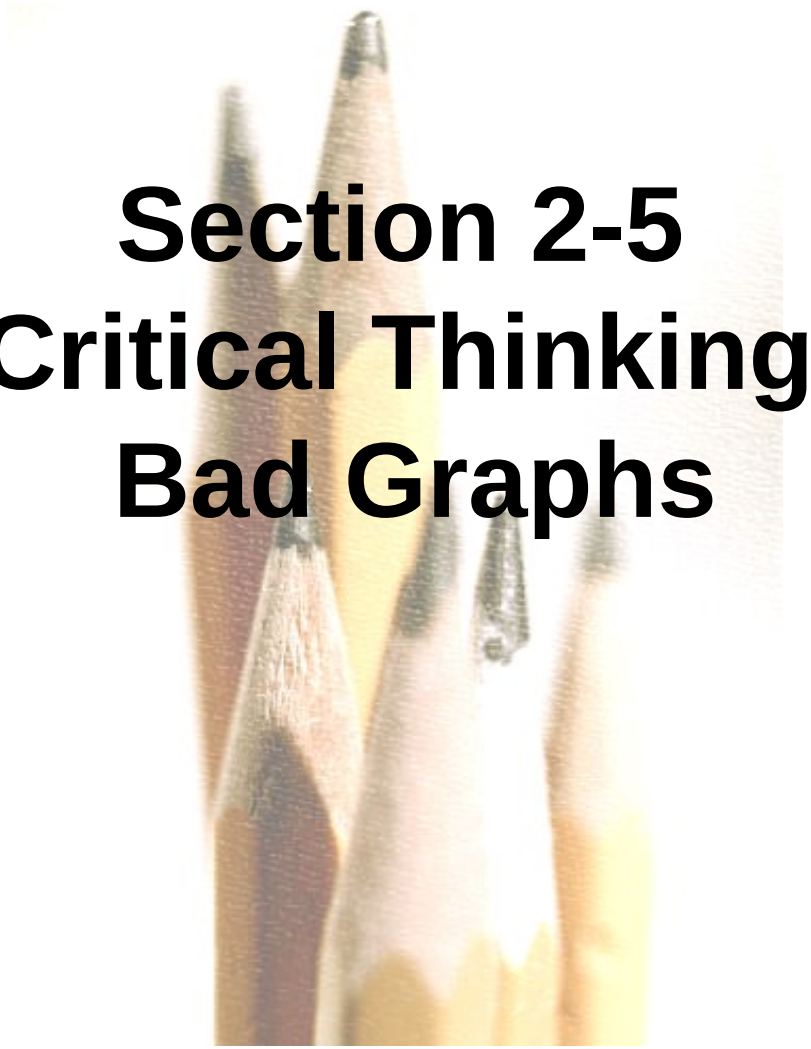
# Recap

**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.**

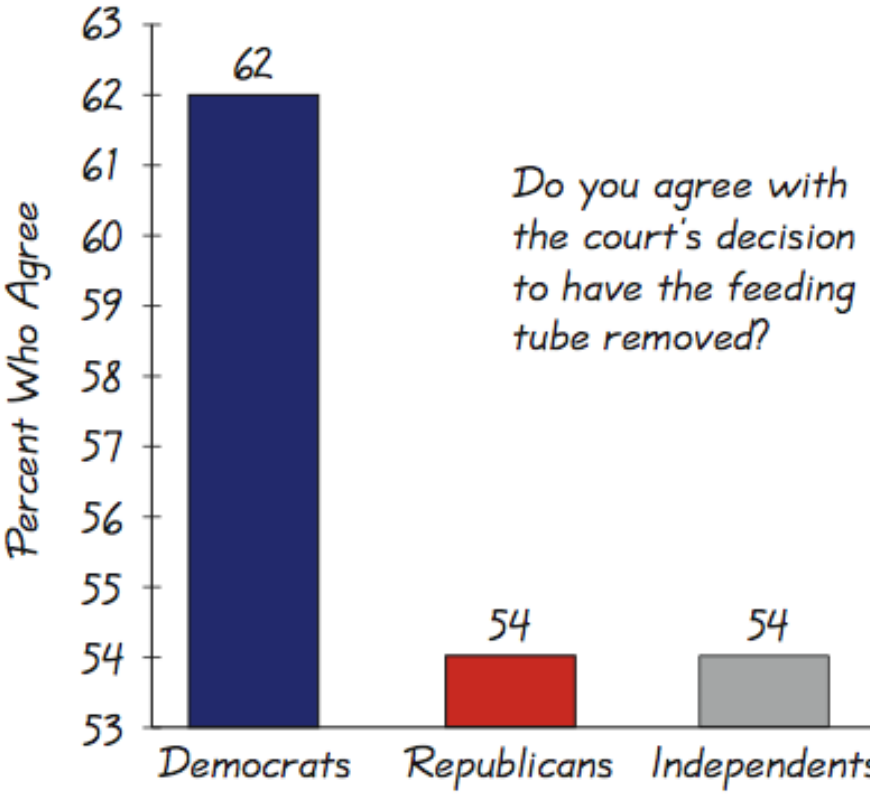



**Figure 2-1 Survey Results by Party**

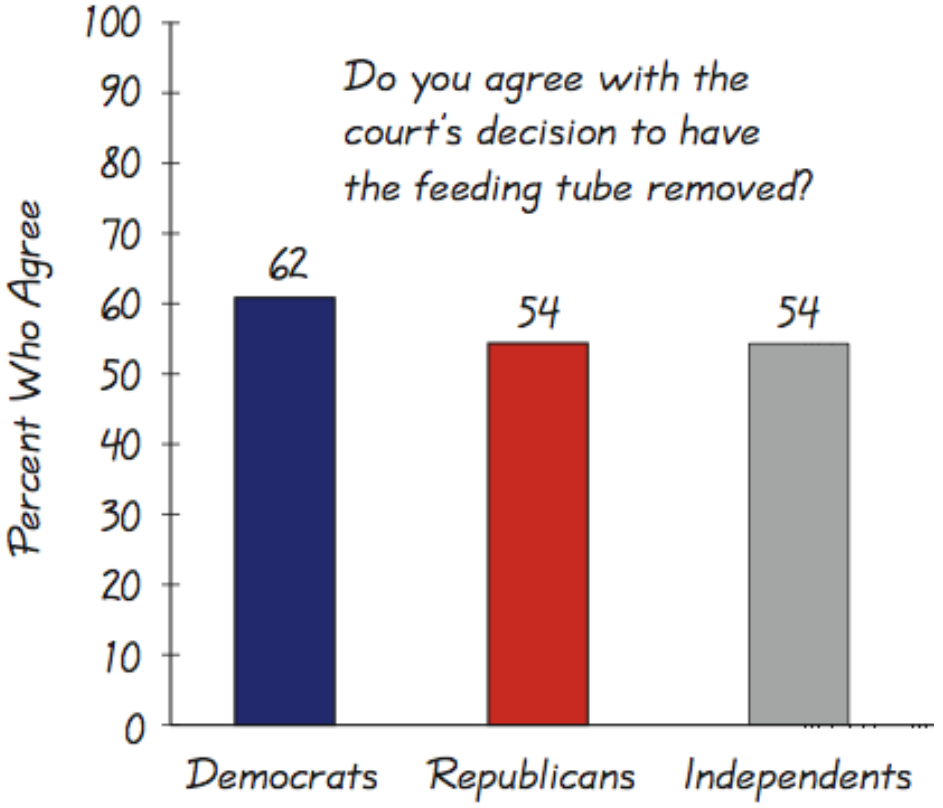



**Figure 2-9 Survey Results by Party**

# Pictographs

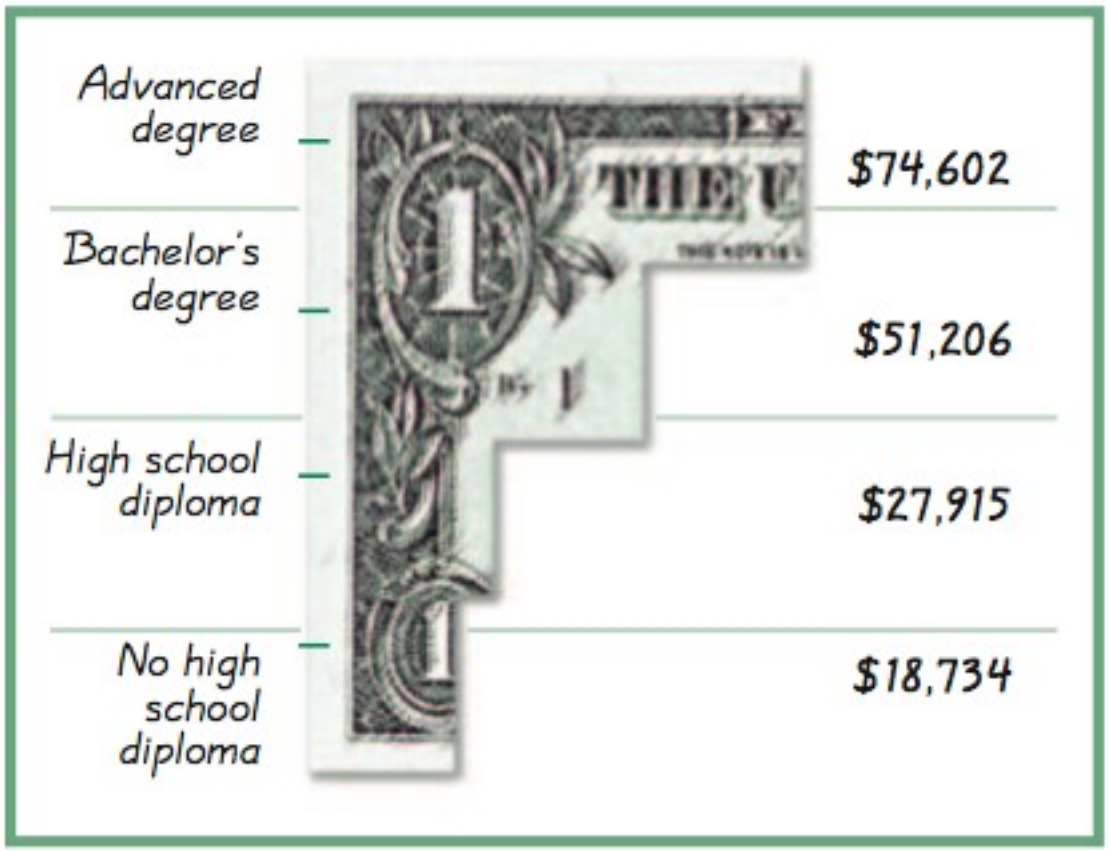
**are drawings of objects. Three-dimensional objects - money bags, stacks of coins, army tanks (for army expenditures), people (for population sizes), barrels (for oil production), and houses (for home construction) are commonly used to depict data.**

**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.**

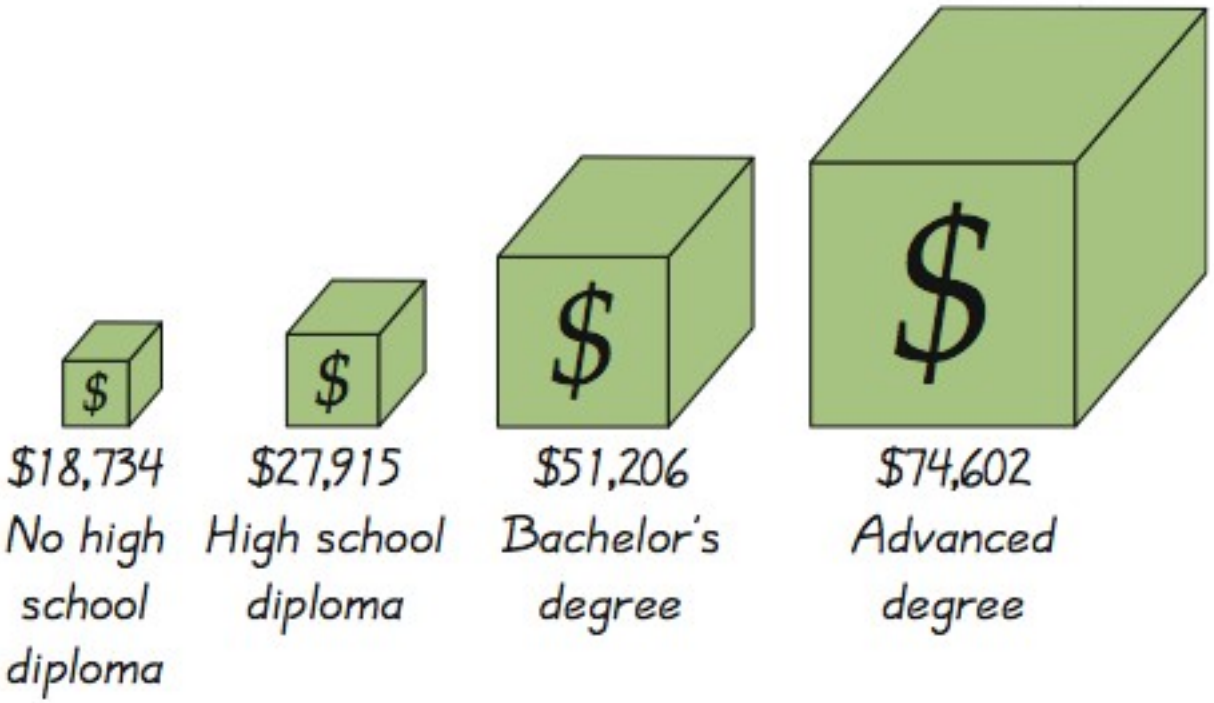
**Pictographs using areas and volumes can therefore be very misleading.**

# Annual Incomes of Groups with Different Education Levels



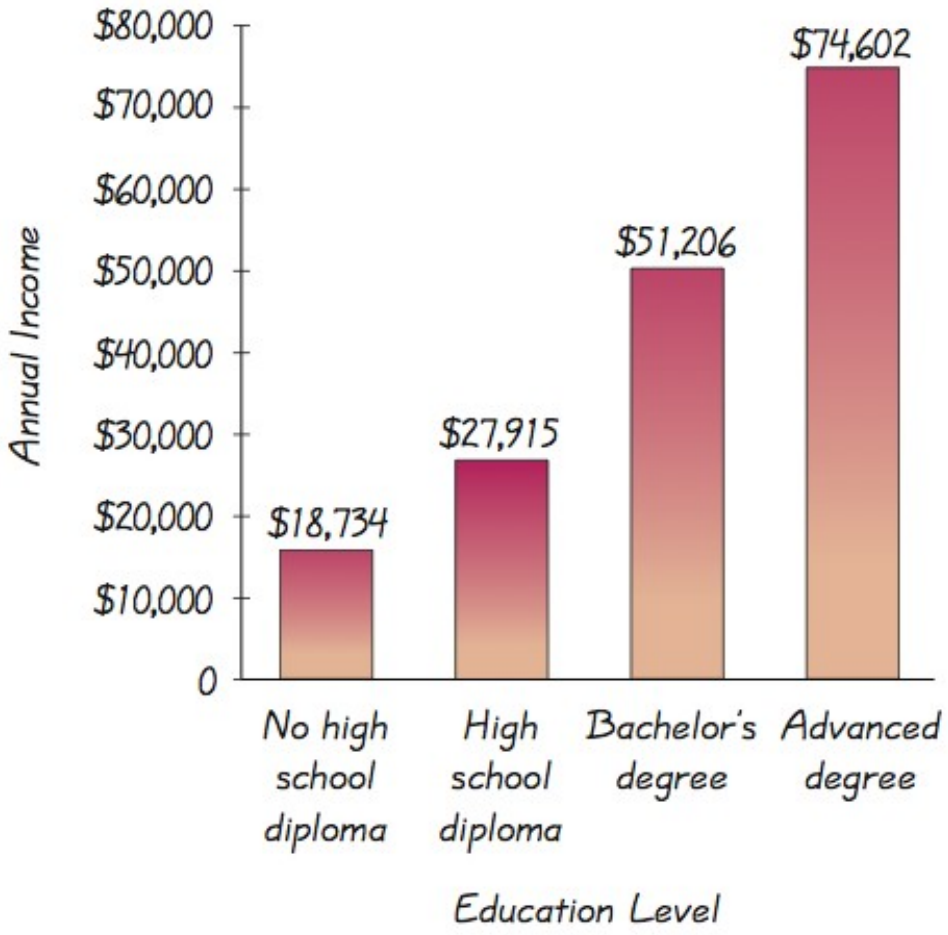

**Bars have same width, too busy, too difficult to understand.**

# 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.**