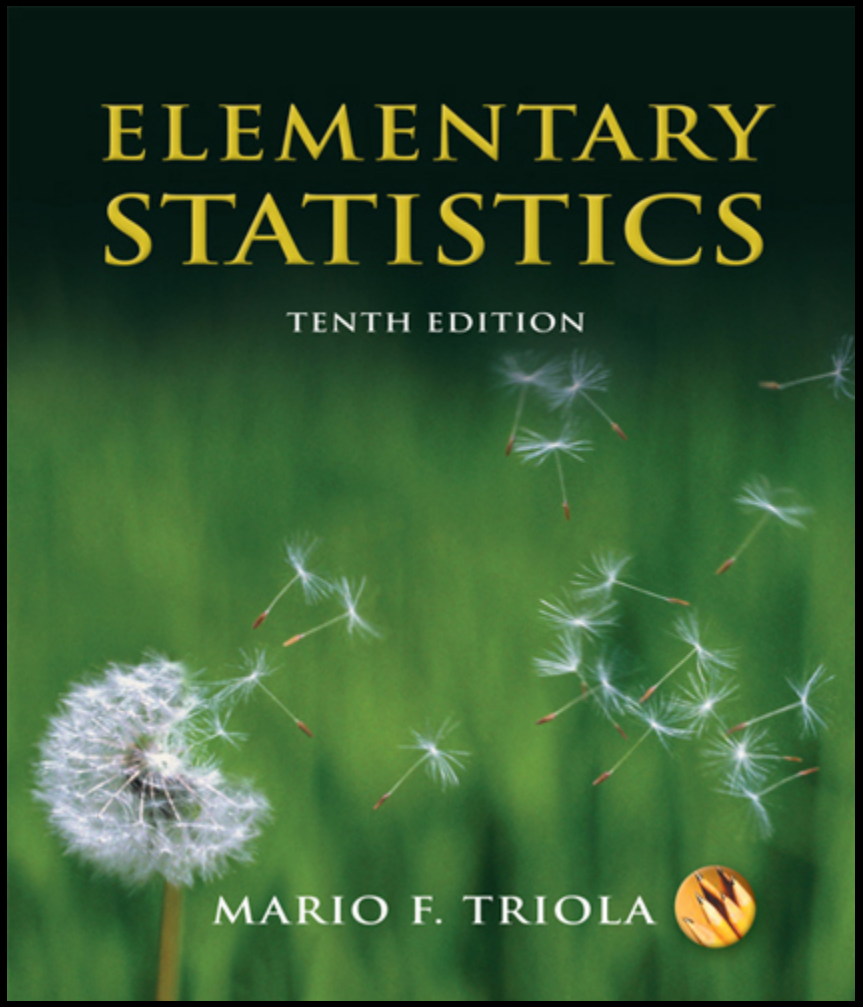
Lecture Slides



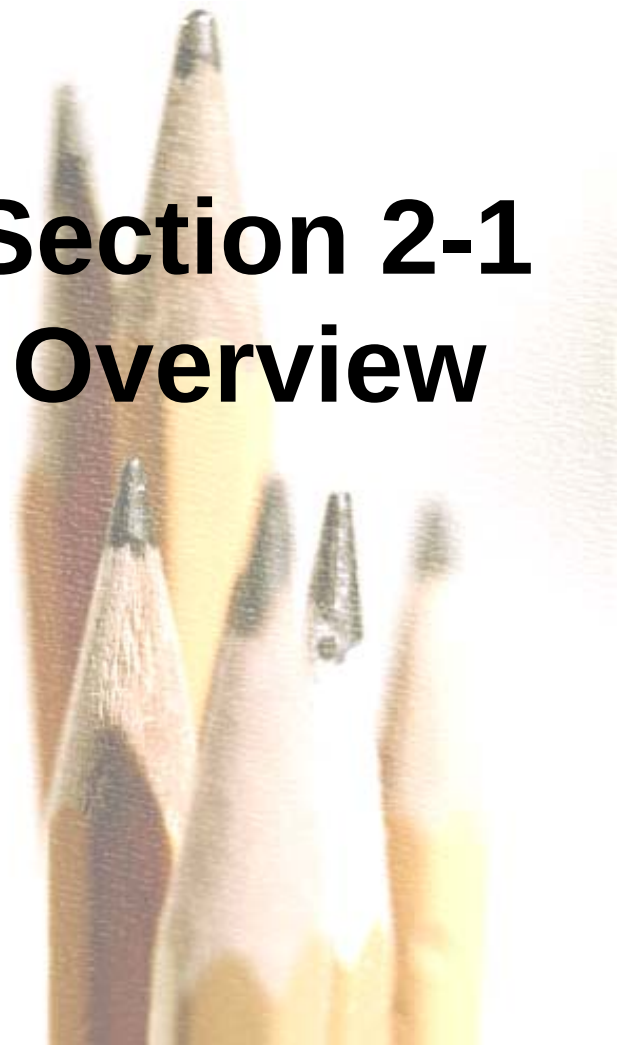
Elementary Statistics
Tenth Edition

and the Triola Statistics Series

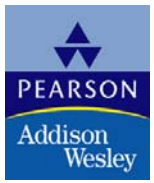
by Mario F. Triola

Chapter 2
Summarizing and Graphing Data

Section 2-1 Overview



Created by Tom Wegleitner, Centreville, Virginia



Overview
Important Characteristics of Data

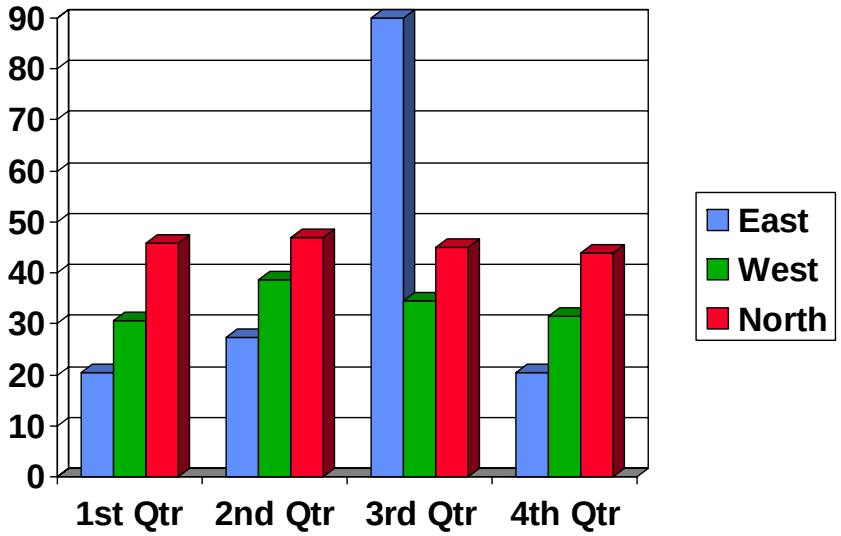
1. **Center**: A representative or average value that indicates where the middle of the data set is located.

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

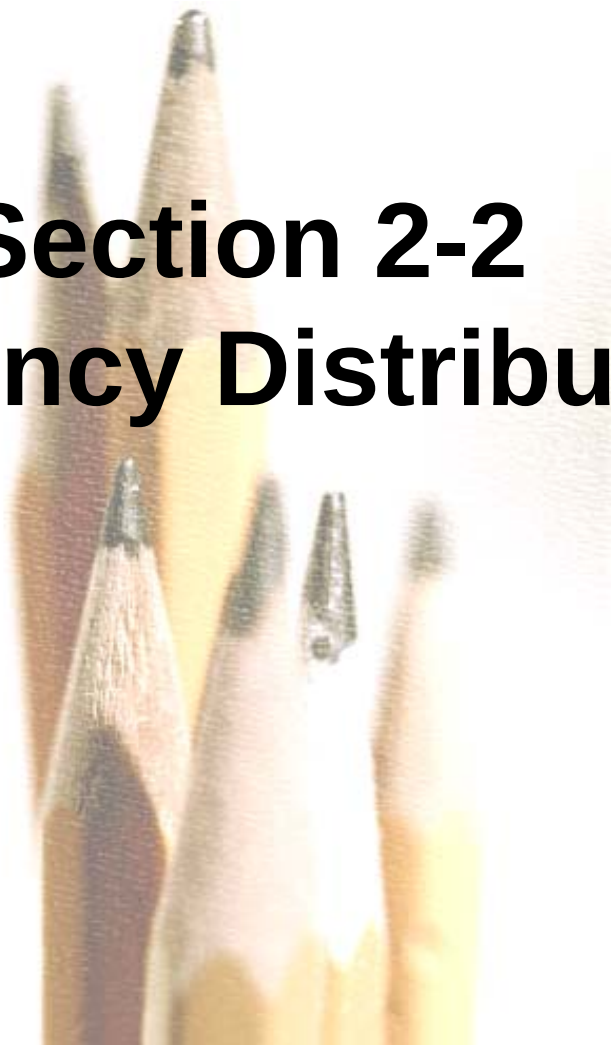
3. **Distribution**: The nature or shape of the distribution of data (such as bell-shaped, uniform, or skewed).

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

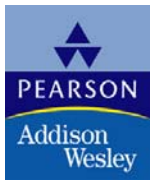
5. **Time**: Changing characteristics of the data over time.



Section 2-2
Frequency Distributions



Created by Tom Wegleitner, Centreville, Virginia



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.

Definition

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

  lists data values (either individually or by groups of intervals), along with their corresponding frequencies or counts

Frequency Distribution Ages of Best Actresses

Table 2-1 Academy Awards: Ages of Best Actresses and Best Actors

The ages (in years) are listed in order, beginning with the first awards ceremony.

Best Actresses

22	37	28	63	32	26	31	27	27	28
30	26	29	24	38	25	29	41	30	35
35	33	29	38	54	24	25	46	41	28
40	39	29	27	31	38	29	25	35	60
43	35	34	34	27	37	42	41	36	32
41	33	31	74	33	50	38	61	21	41
26	80	42	29	33	35	45	49	39	34
26	25	33	35	35	28				

Best Actors

44	41	62	52	41	34	34	52	41	37
38	34	32	40	43	56	41	39	49	57
41	38	42	52	51	35	30	39	41	44
49	35	47	31	47	37	57	42	45	42
44	62	43	42	48	49	56	38	60	30
40	42	36	76	39	53	45	36	62	43
51	32	42	54	52	37	38	32	45	60
46	40	36	47	29	43				

Original Data

Table 2-2
Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Frequency Distribution

Frequency Distributions

Definitions

Lower Class Limits

are the smallest numbers that can actually belong to different classes

Lower Class Limits

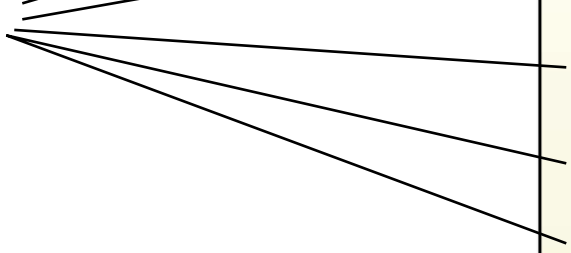


Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Upper Class Limits

are the largest numbers that can actually belong to different classes

Upper Class Limits

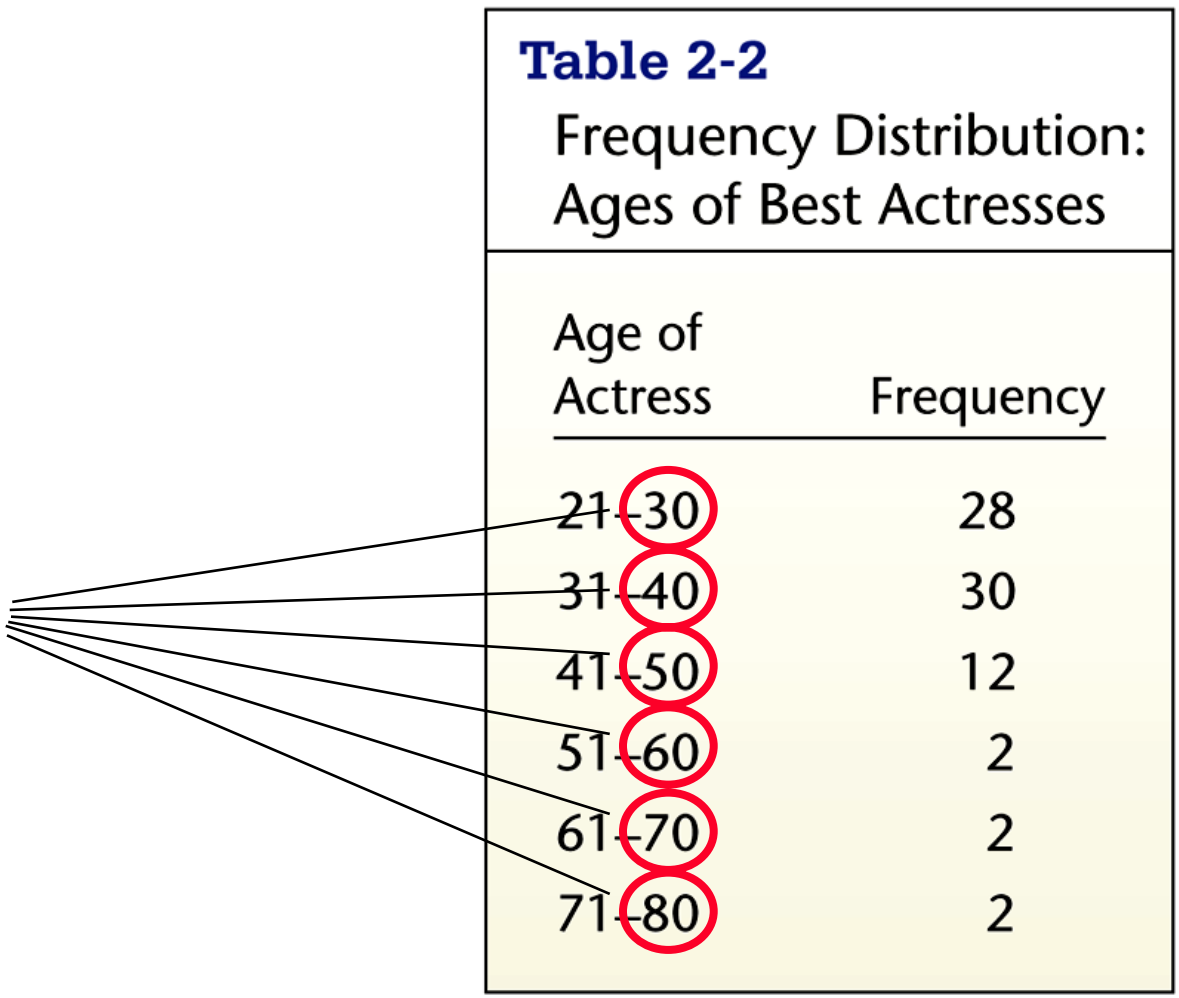


Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Class Boundaries

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

Class Boundaries

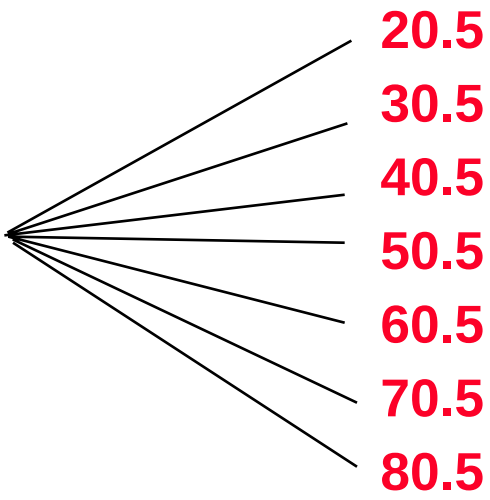


Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Class Midpoints

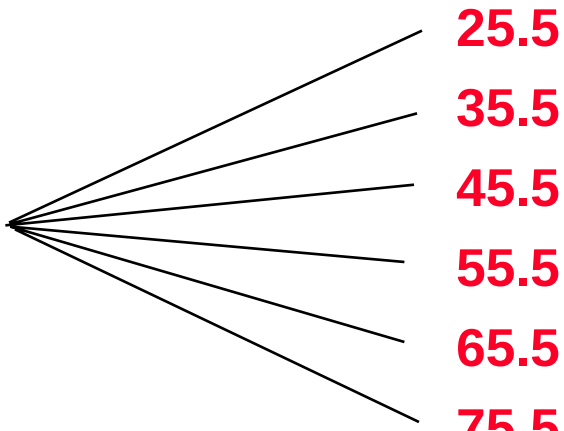
can be found by adding the lower class limit to the upper class limit and dividing the sum by two

Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Class Midpoints



Class Width

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

Class Width

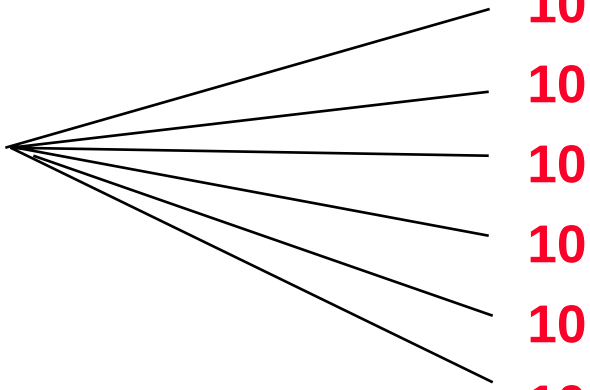


Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Reasons for Constructing Frequency Distributions

1. Large data sets can be summarized.

2. We can gain some insight into the nature of data.

3. We have a basis for constructing important graphs.

Constructing A Frequency Distribution

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

2. Calculate (round up).

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

3. Starting point: Begin by choosing a lower limit of the first class.

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

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

6. Go through the data set putting a tally in the appropriate class for each data value.

Relative Frequency Distribution

includes the same class limits as a frequency distribution, but relative frequencies are used instead of actual frequencies

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

Relative Frequency Distribution

Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Table 2-3

Relative Frequency Distribution of Best Actress Ages

Age of Actress	Relative Frequency
21–30	37%
31–40	39%
41–50	16%
51–60	3%
61–70	3%
71–80	3%

28/76 = 37%

30/76 = 39%

etc.

Total Frequency = 76

Cumulative Frequency Distribution

Table 2-2
Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Table 2-4
Cumulative Frequency Distribution of Best Actress Ages

Age of Actress	Cumulative Frequency
Less than 31	28
Less than 41	58
Less than 51	70
Less than 61	72
Less than 71	74
Less than 81	76

Cumulative Frequencies

Frequency Tables

Table 2-2
Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Table 2-3
Relative Frequency Distribution of Best Actress Ages

Age of Actress	Relative Frequency
21–30	37%
31–40	39%
41–50	16%
51–60	3%
61–70	3%
71–80	3%

Table 2-4
Cumulative Frequency Distribution of Best Actress Ages

Age of Actress	Cumulative Frequency
Less than 31	28
Less than 41	58
Less than 51	70
Less than 61	72
Less than 71	74
Less than 81	76

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 some maximum frequency, then decrease to a low frequency.**

- **The distribution should be approximately symmetric.**

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

Created by Tom Wegleitner, Centreville, Virginia

PEARSON
Addison Wesley

Key Concept

A histogram is an important type of graph that portrays the nature of the distribution.

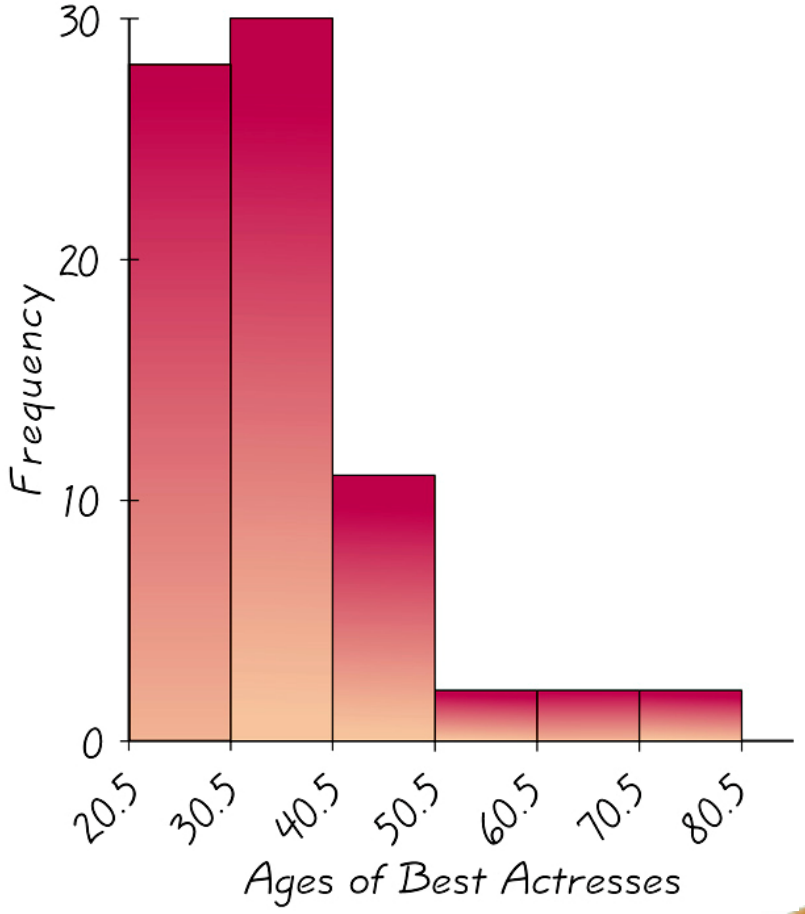
Histogram

A bar graph in which the horizontal scale represents the classes of data values and the vertical scale represents the frequencies

Table 2-2

Frequency Distribution: Ages of Best Actresses

Age of Actress	Frequency
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2



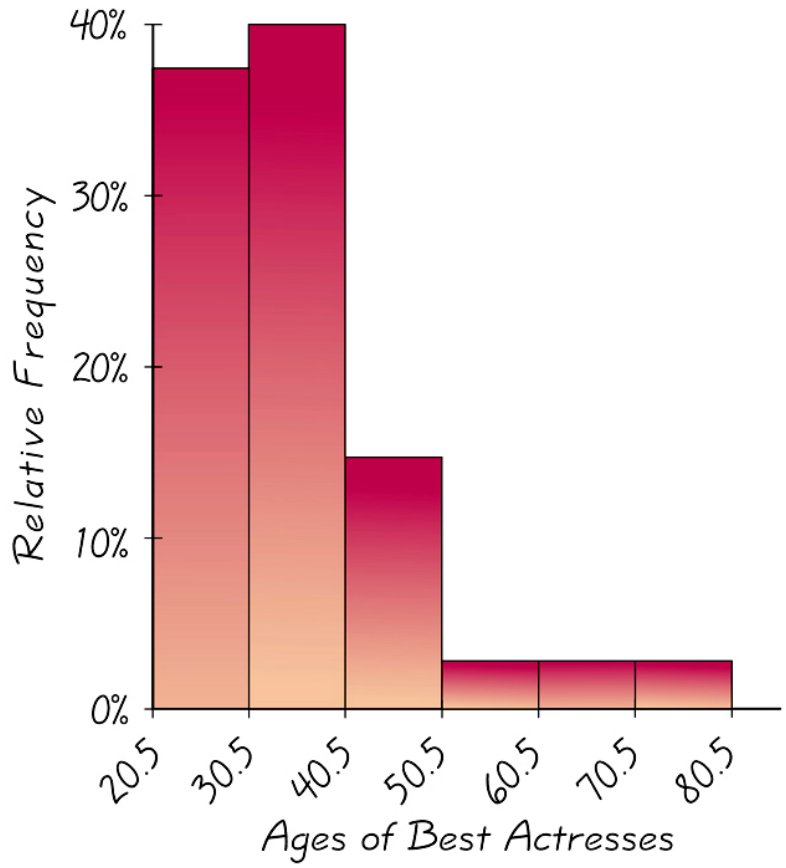
Relative Frequency Histogram

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

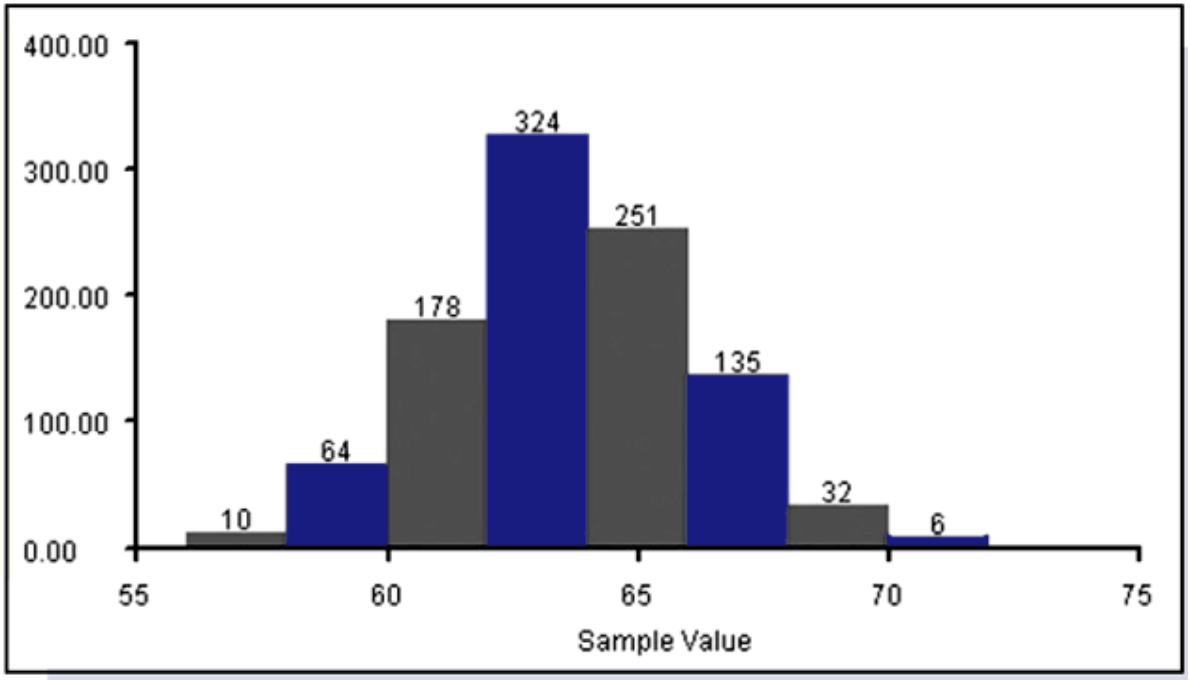
Relative Frequency Distribution of Best Actress Ages

Age of Actress	Relative Frequency
21–30	37%
31–40	39%
41–50	16%
51–60	3%
61–70	3%
71–80	3%



Critical Thinking
Interpreting Histograms

One key characteristic of a normal distribution is that it has a "bell" shape. The histogram below illustrates this.

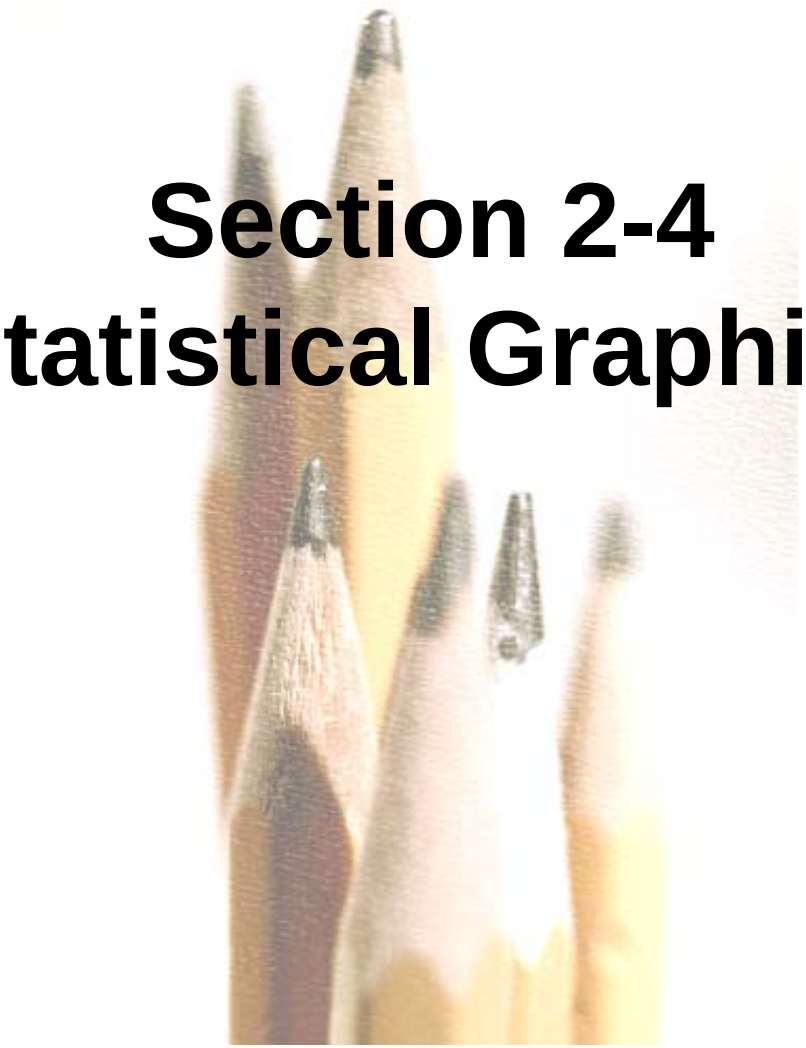


Recap

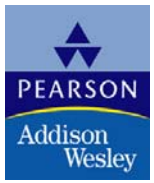
In this Section we have discussed

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

Section 2-4
Statistical Graphics



Created by Tom Wegleitner, Centreville, Virginia



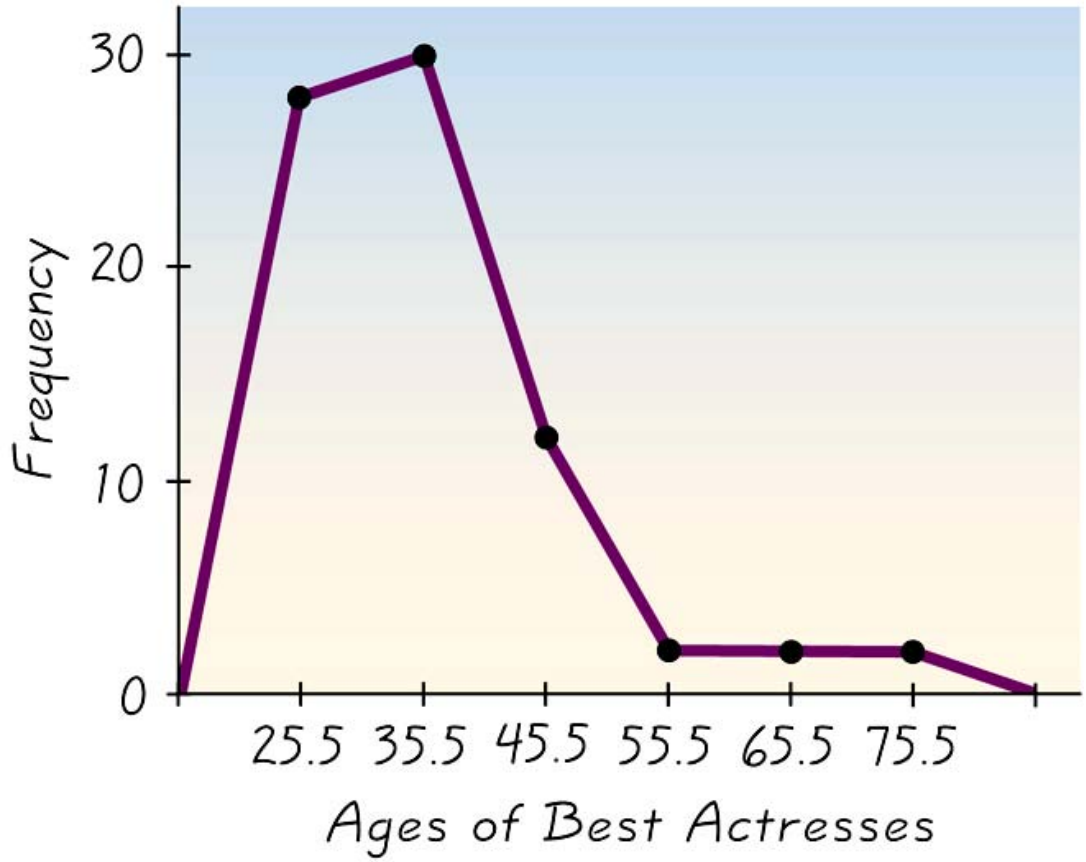
Key Concept

This section presents other graphs beyond histograms commonly used in statistical analysis.

The main objective is to understand a data set by using a suitable graph that is effective in revealing some important characteristic.

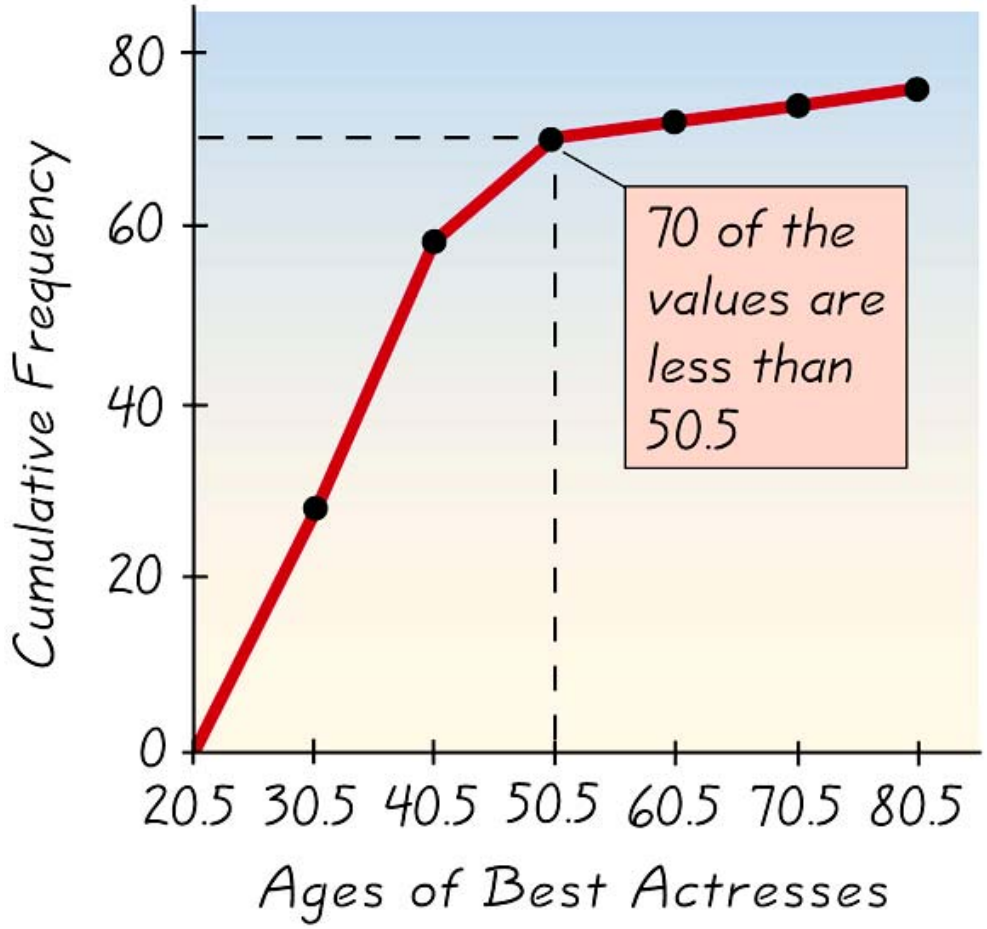
Frequency Polygon

Uses line segments connected to points directly above class midpoint values



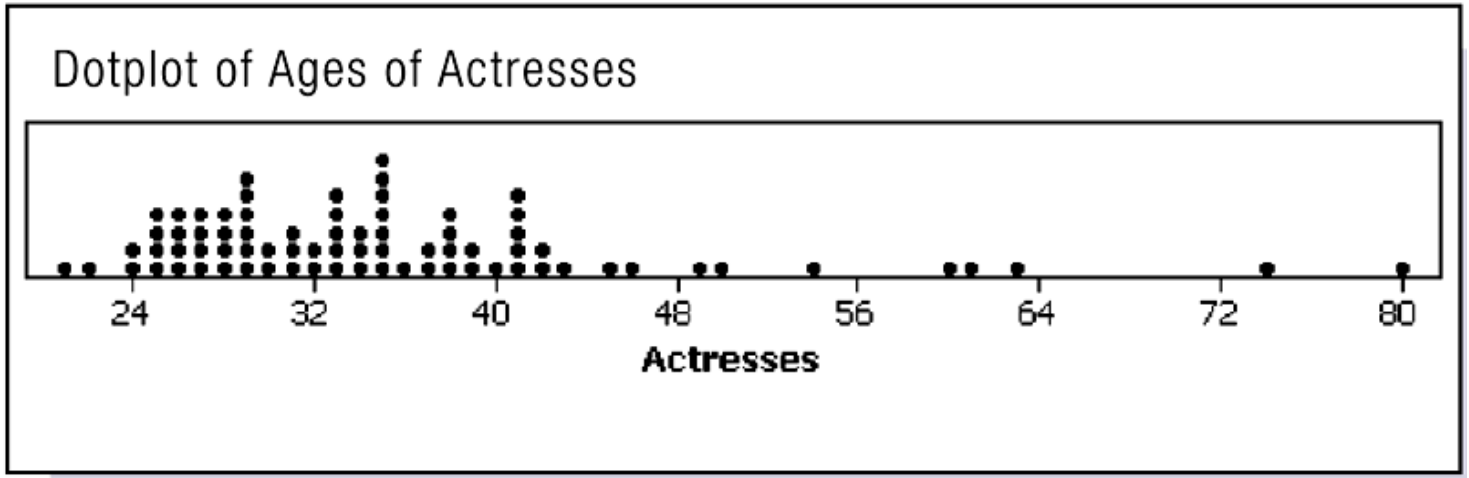
Ogive

A line graph that depicts cumulative frequencies



Dot Plot

Consists of a graph in which each data value is plotted as a point (or dot) along a scale of values



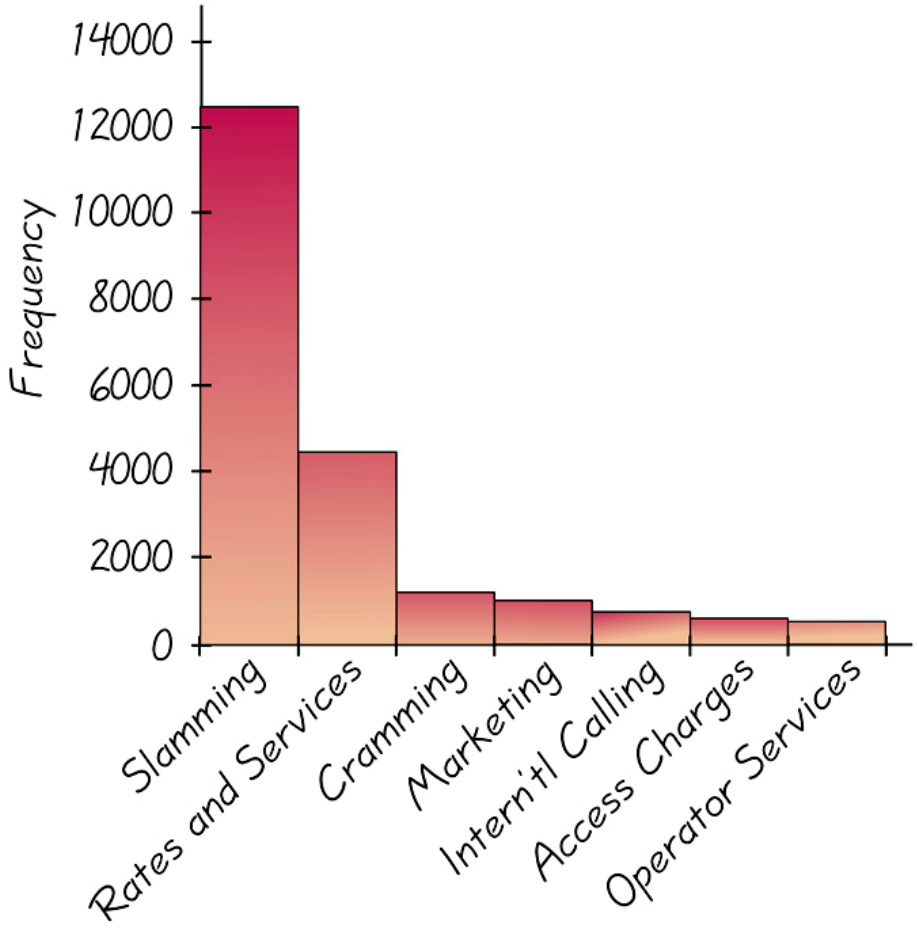
Stemplot (or Stem-and-Leaf Plot)

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

Stem (tens)	Leaves (units)
2	1244555566667777888899999
3	00111223333344455555556778888899
4	011111223569
5	04 ←Values are 50 and 54.
6	013
7	4
8	0 ←Value is 80.

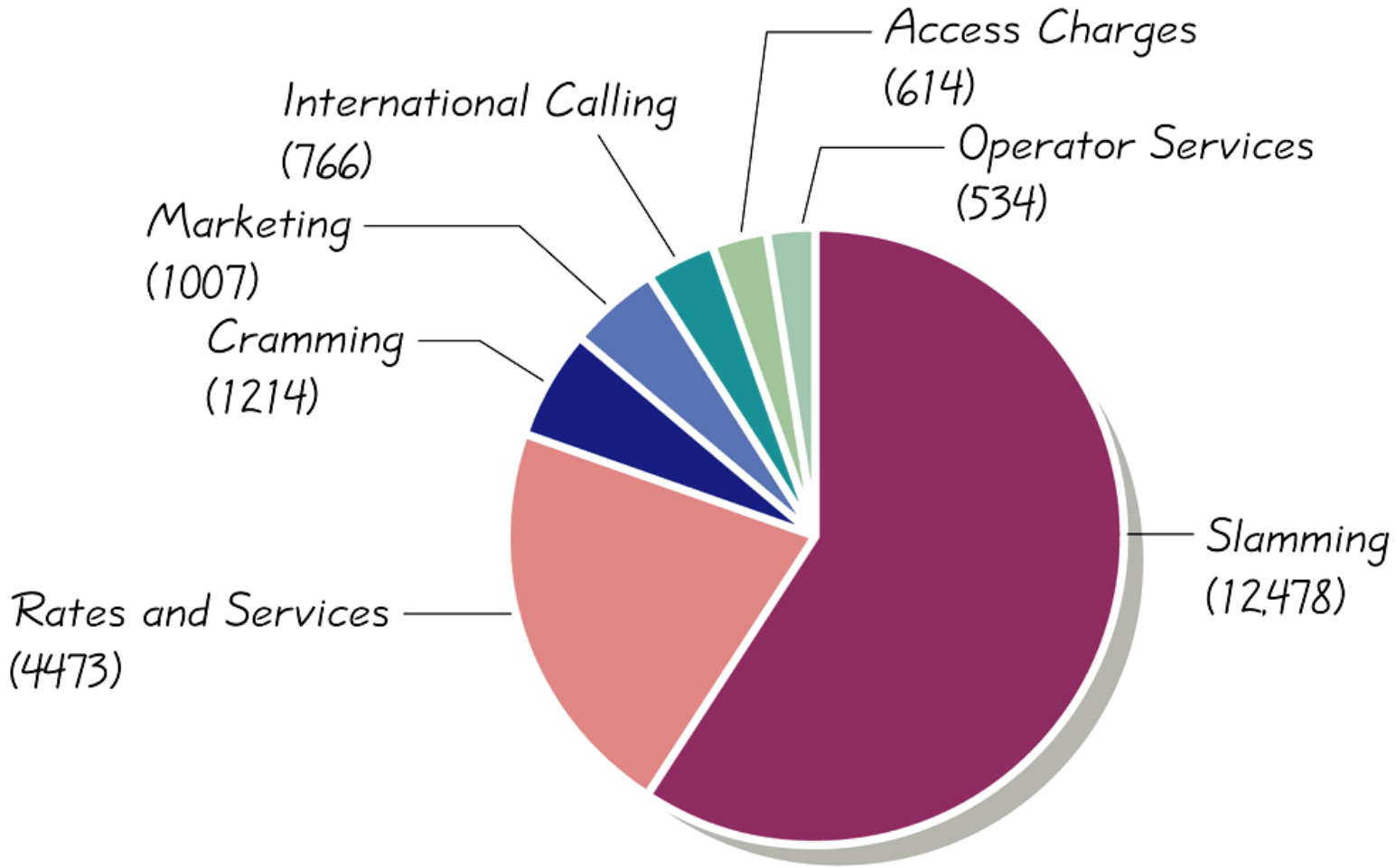
Pareto Chart

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



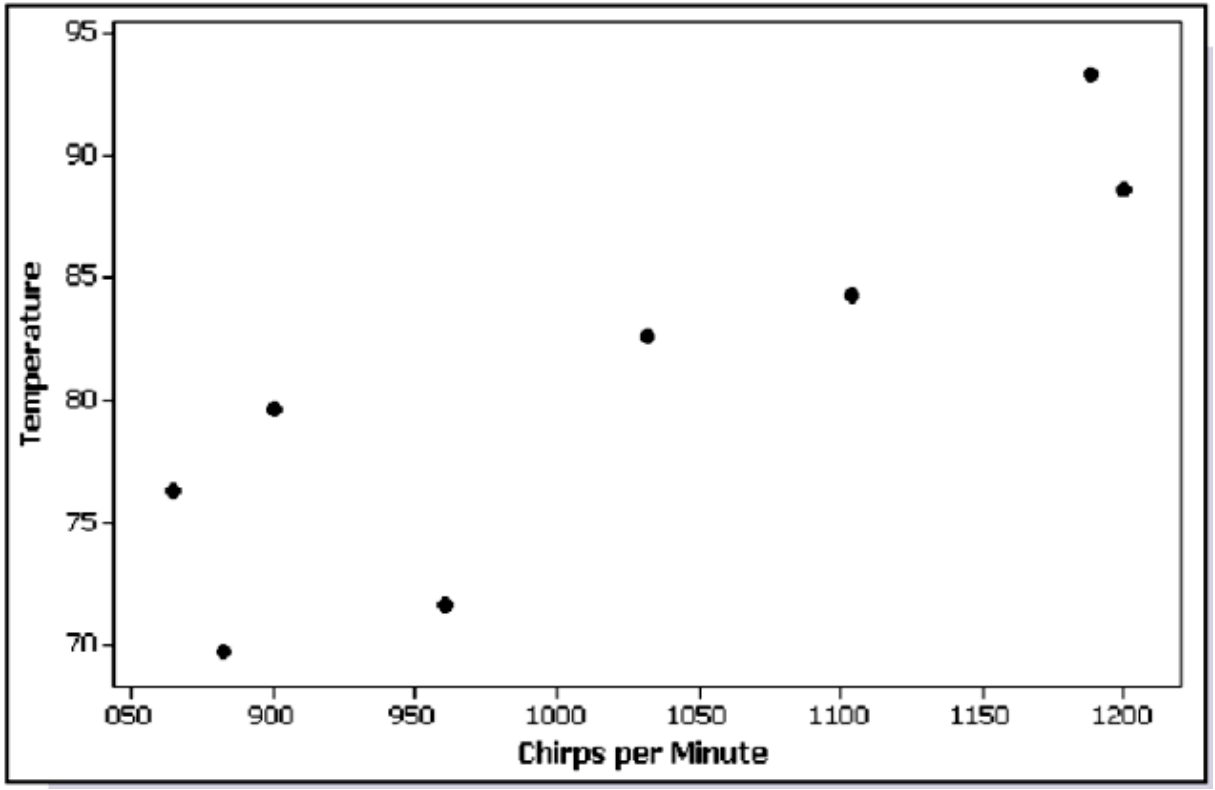
Pie Chart

A graph depicting qualitative data as slices of a pie



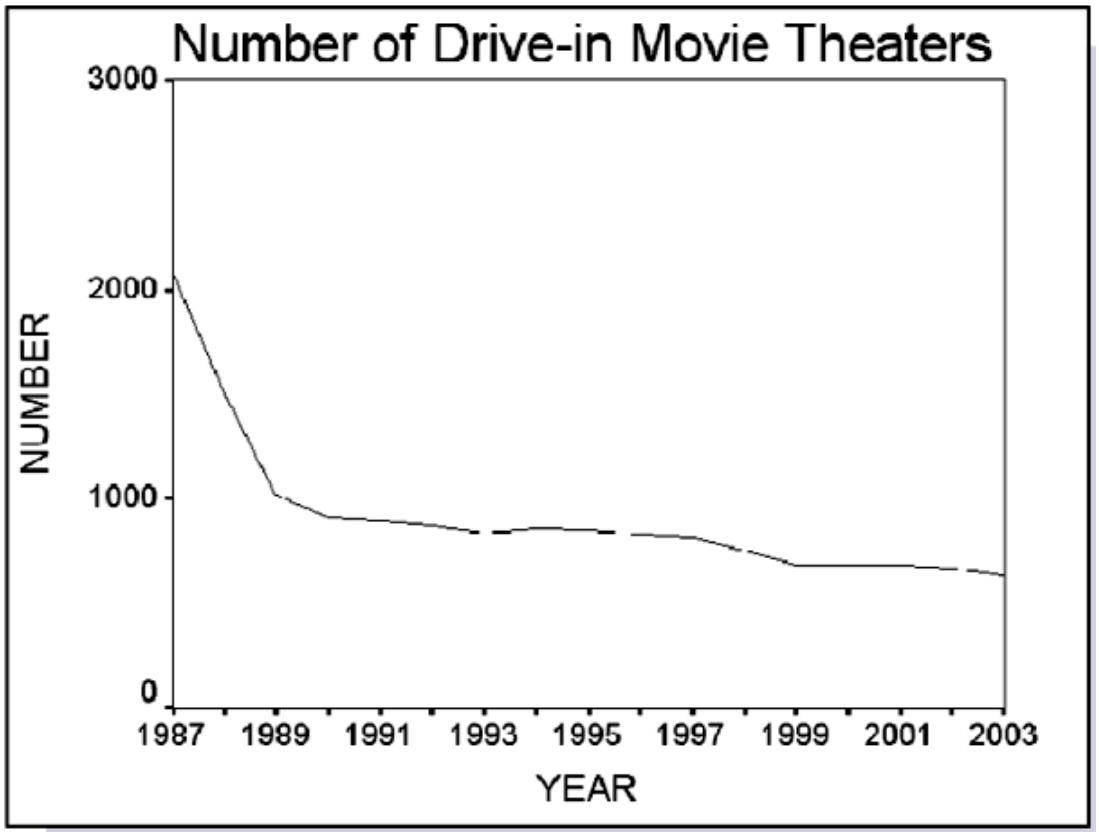
Scatter Plot (or Scatter Diagram)

A plot of paired (*x*,*y*) data with a horizontal x-axis and a vertical y-axis

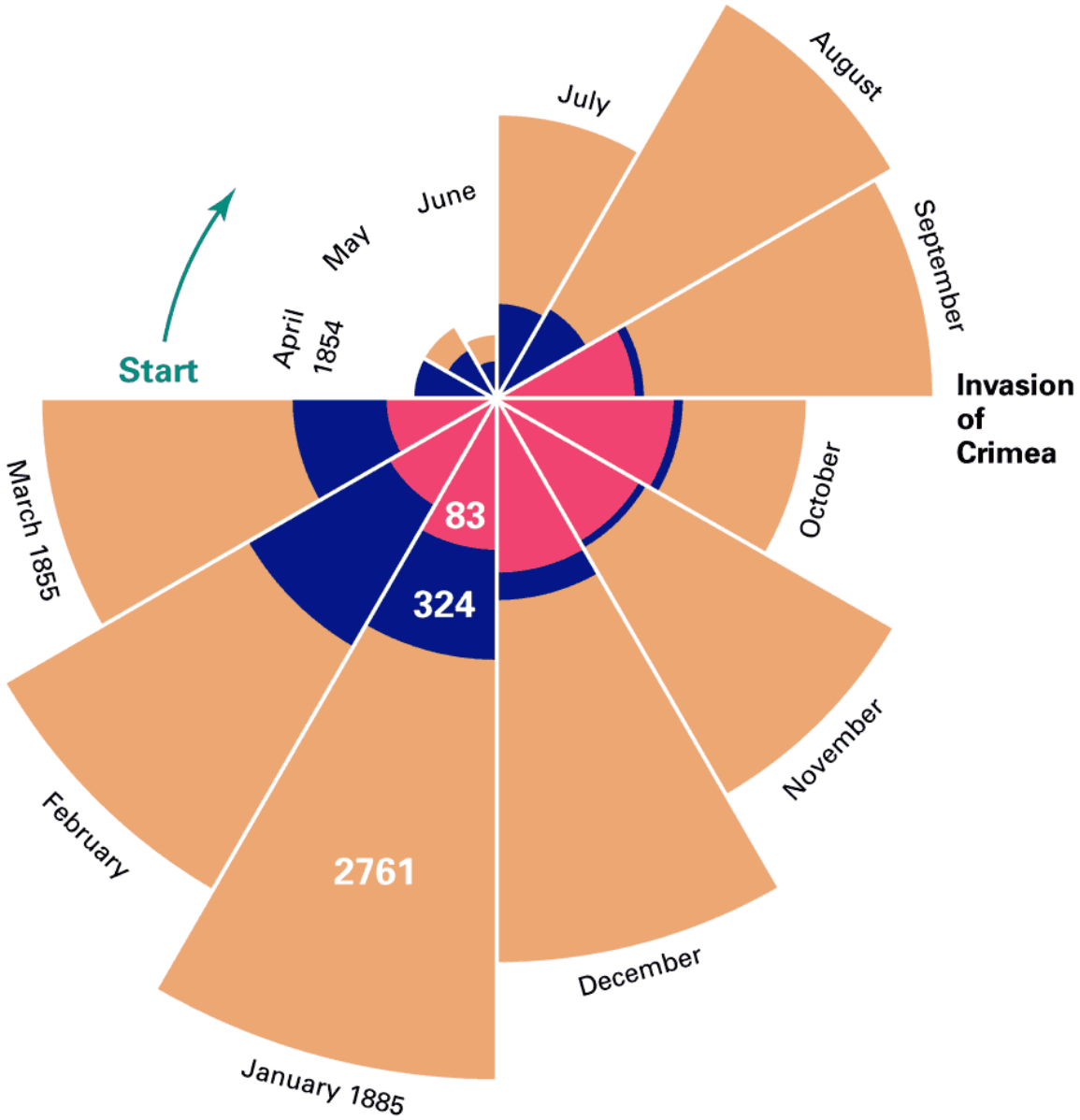


Time-Series Graph

Data that have been collected at different points in time



Other Graphs



Recap

In this section we have discussed graphs that are pictures of distributions.

Keep in mind that a graph is a tool for describing, exploring and comparing data.