



Faculty of Applied Engineering and Urban Planning

Civil Engineering Department

Dr. Eng. Mustafa Maher Al-tayeb

2 Semester 2022/2023

Probabilistic and Statistical

Chapter 2

Summarizing and Graphing Data

2-1 Overview

2-2 Frequency Distributions

2-3 Histograms

2-4 Statistical Graphics

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.

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.

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				

Original Data

Frequency Distribution Ages of Best Actresses

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

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

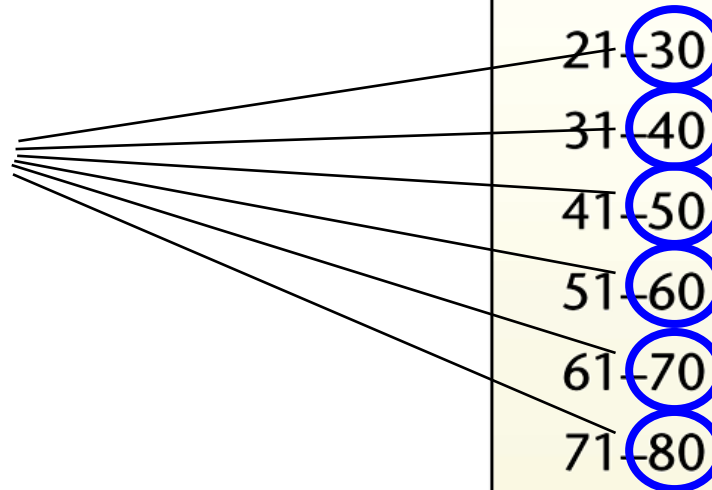


Table 2-2

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**

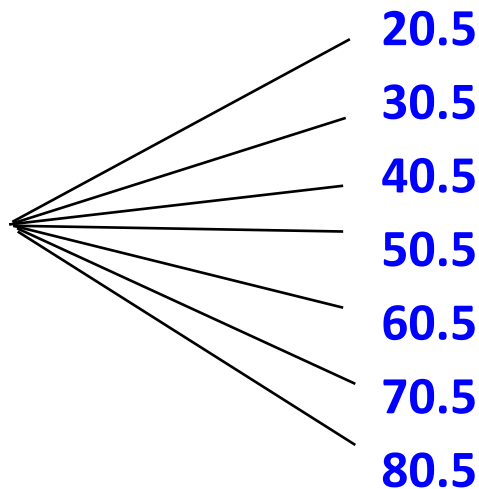


Table 2-2

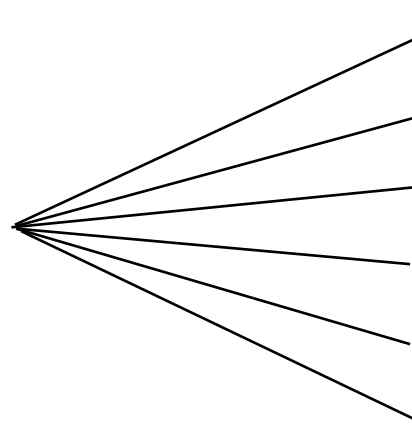
Frequency Distribution:
Ages of Best Actresses

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

**Class
Midpoints**



25.5
35.5
45.5
55.5
65.5
75.5

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 Width

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

**Class
Width**

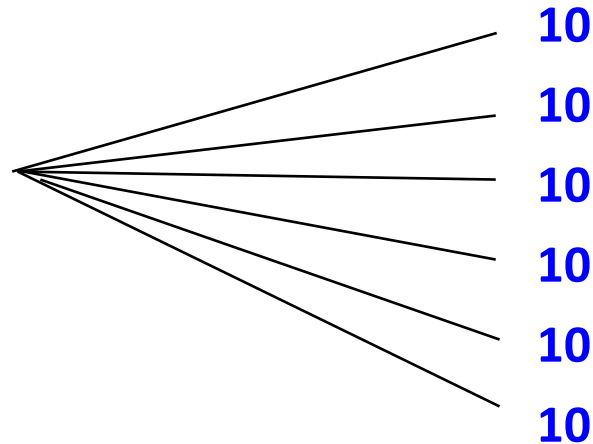


Table 2-2

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 such as Histogram.

Constructing A Frequency Distribution

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

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

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 each data value in its class interval

Constructing A Frequency Distribution

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

A useful recipe to determine the number of classes (K) is the “2 to the k rule”. This guide suggests you select the smallest number k for the number of classes such that 2^k is greater than the number of observations n .

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

Total Frequency = 76

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 Distribution

$$28/76 = 37\%$$

$$30/76 = 39\%$$

etc.

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%

Cumulative Frequency Distribution

Table 2-2

Frequency Distribution:
Ages of Best Actresses

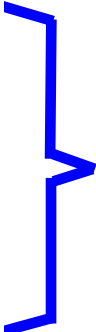
<u>Age of Actress</u>	<u>Frequency</u>
21–30	28
31–40	30
41–50	12
51–60	2
61–70	2
71–80	2

Cumulative Frequency Distribution

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.

Histograms

Key Concept

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

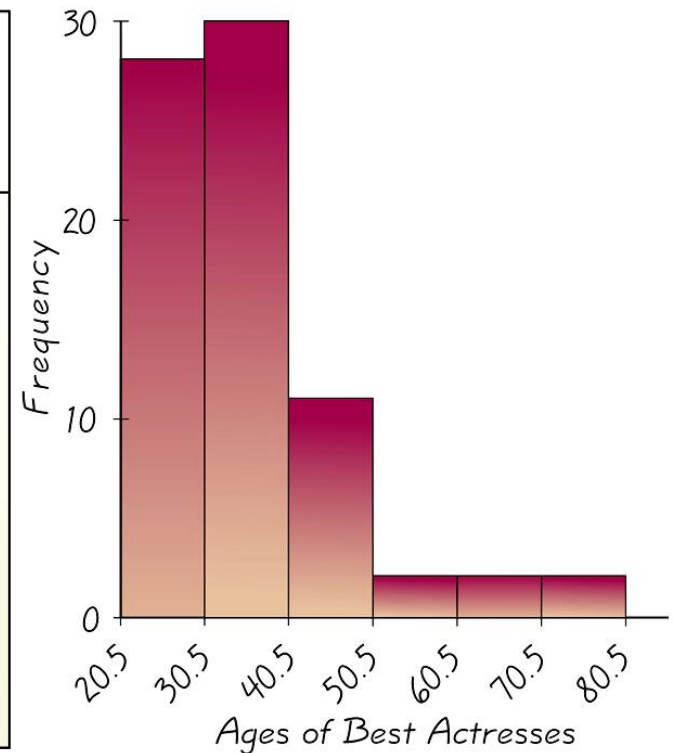
Histogram

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



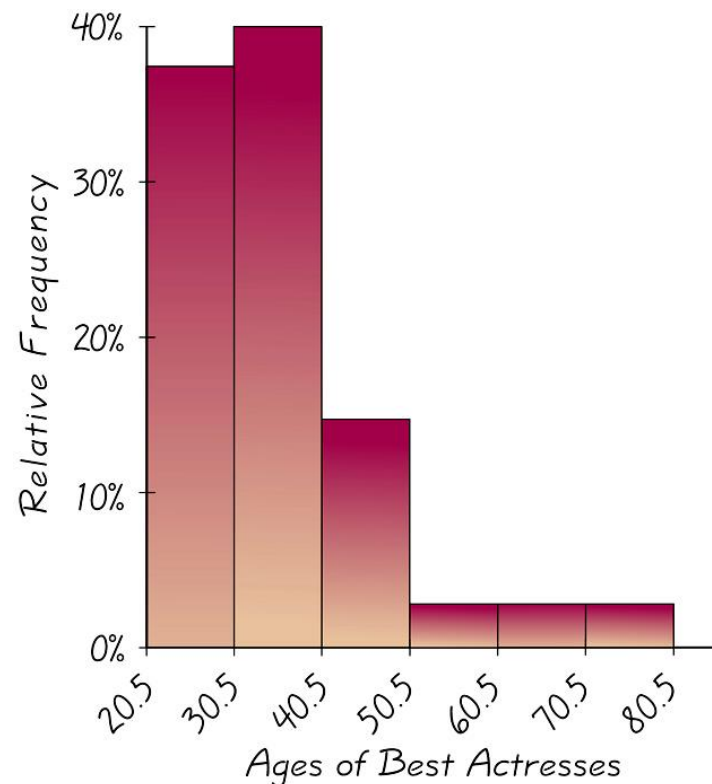
Relative Frequency Histogram

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

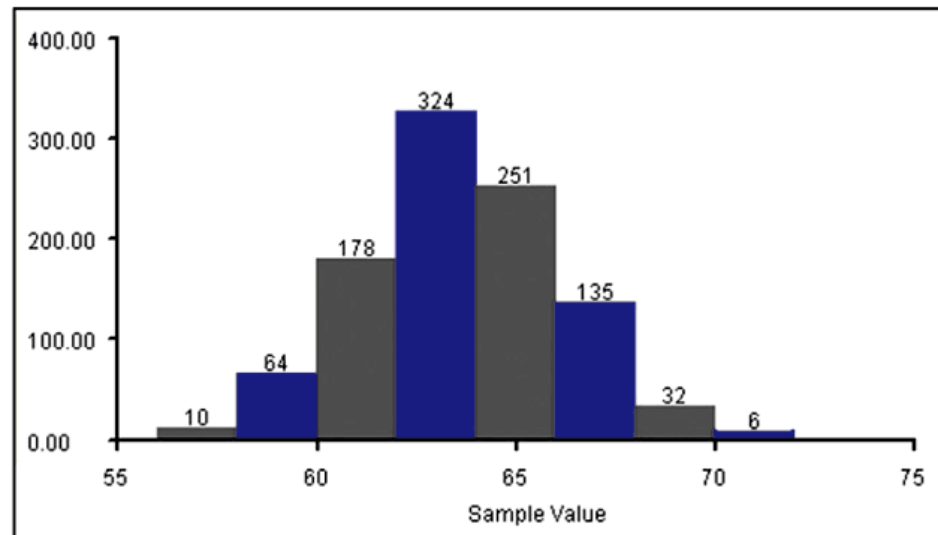
Relative Frequency
Distribution of Best
Actress Ages

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

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

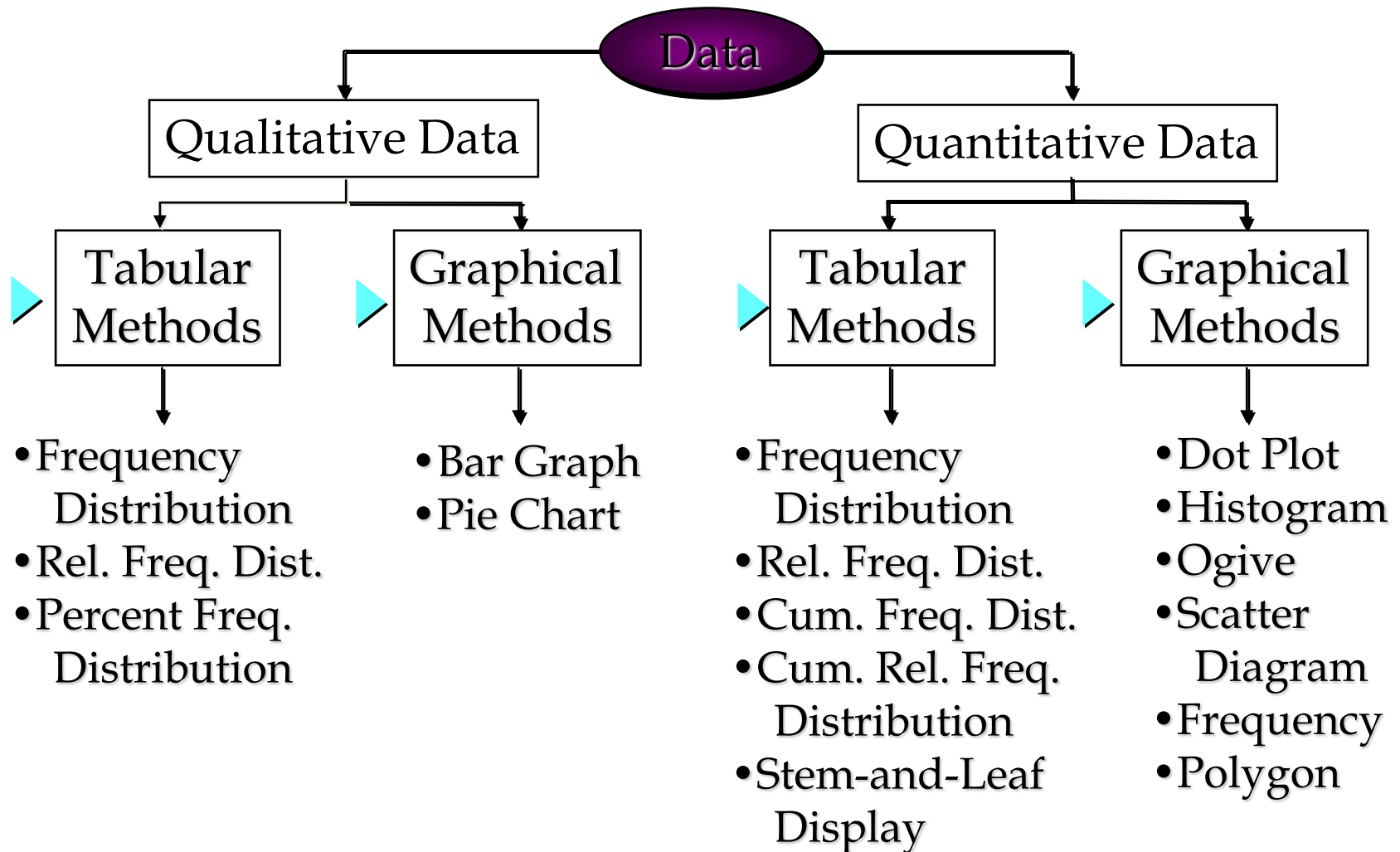


Statistical Graphics

Key Concept

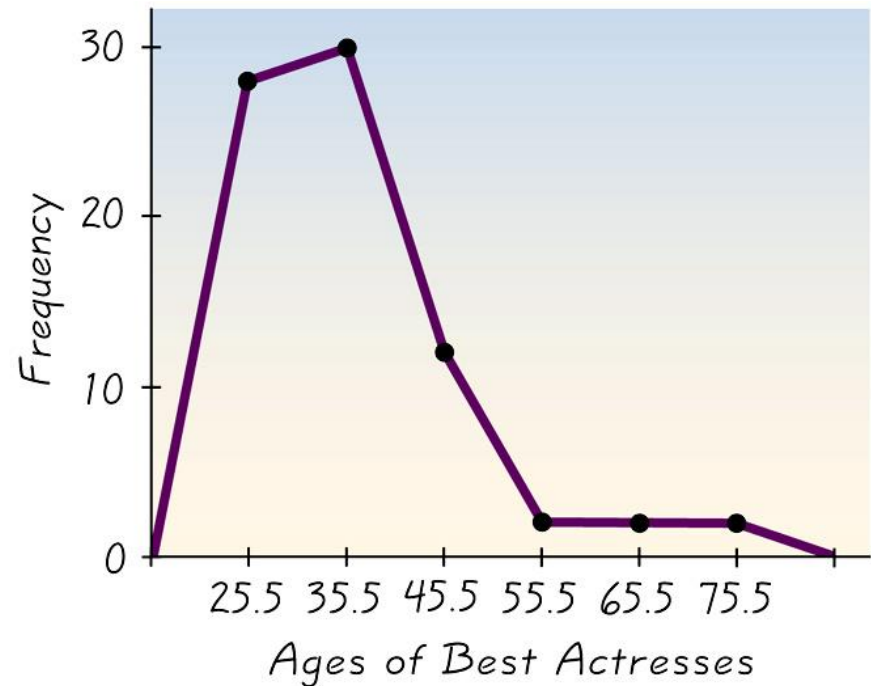
- ✚ This section presents other graphs beyond histograms commonly used in statistical analysis.
- ✚ The main objective is to understand a data set by using a suitable graph that is effective in revealing some important characteristic.

Tabular and Graphical Procedures



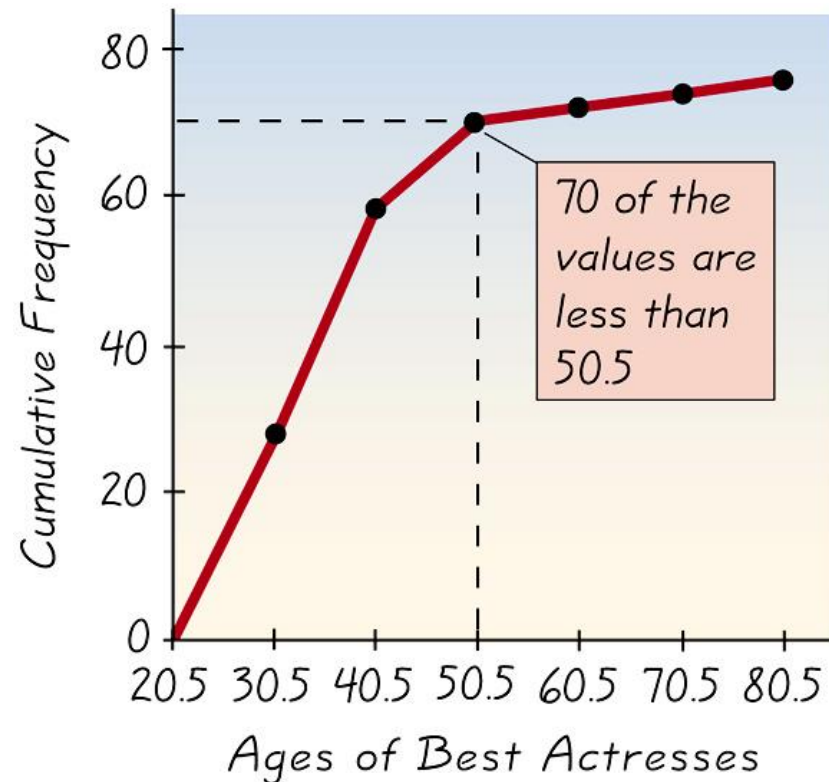
Frequency Polygon

Uses line segments connected to points directly above class midpoint values



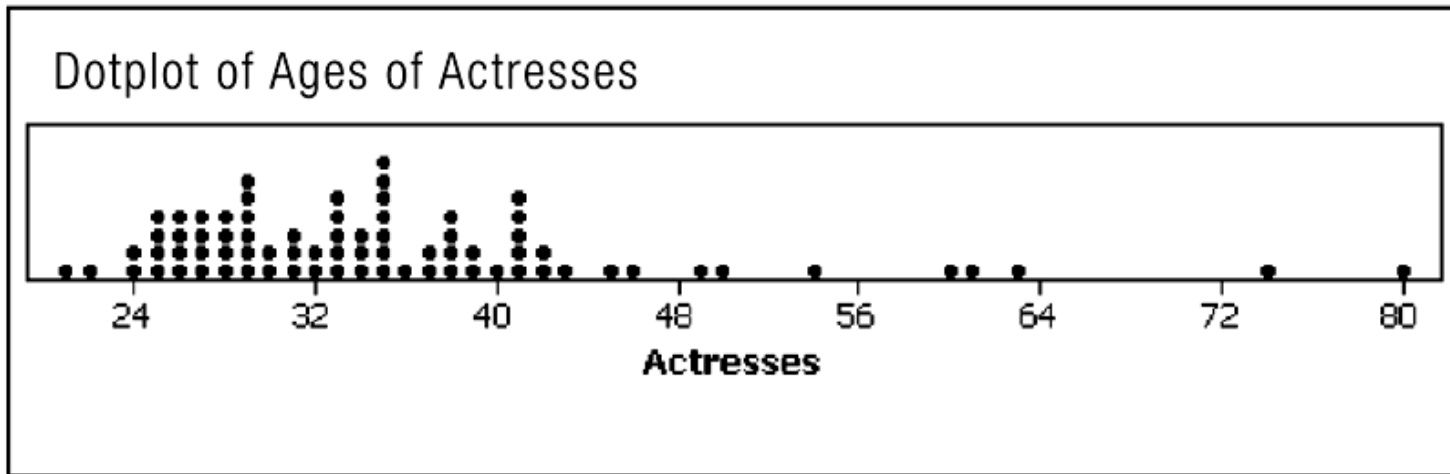
Ogive (S – Curve)

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

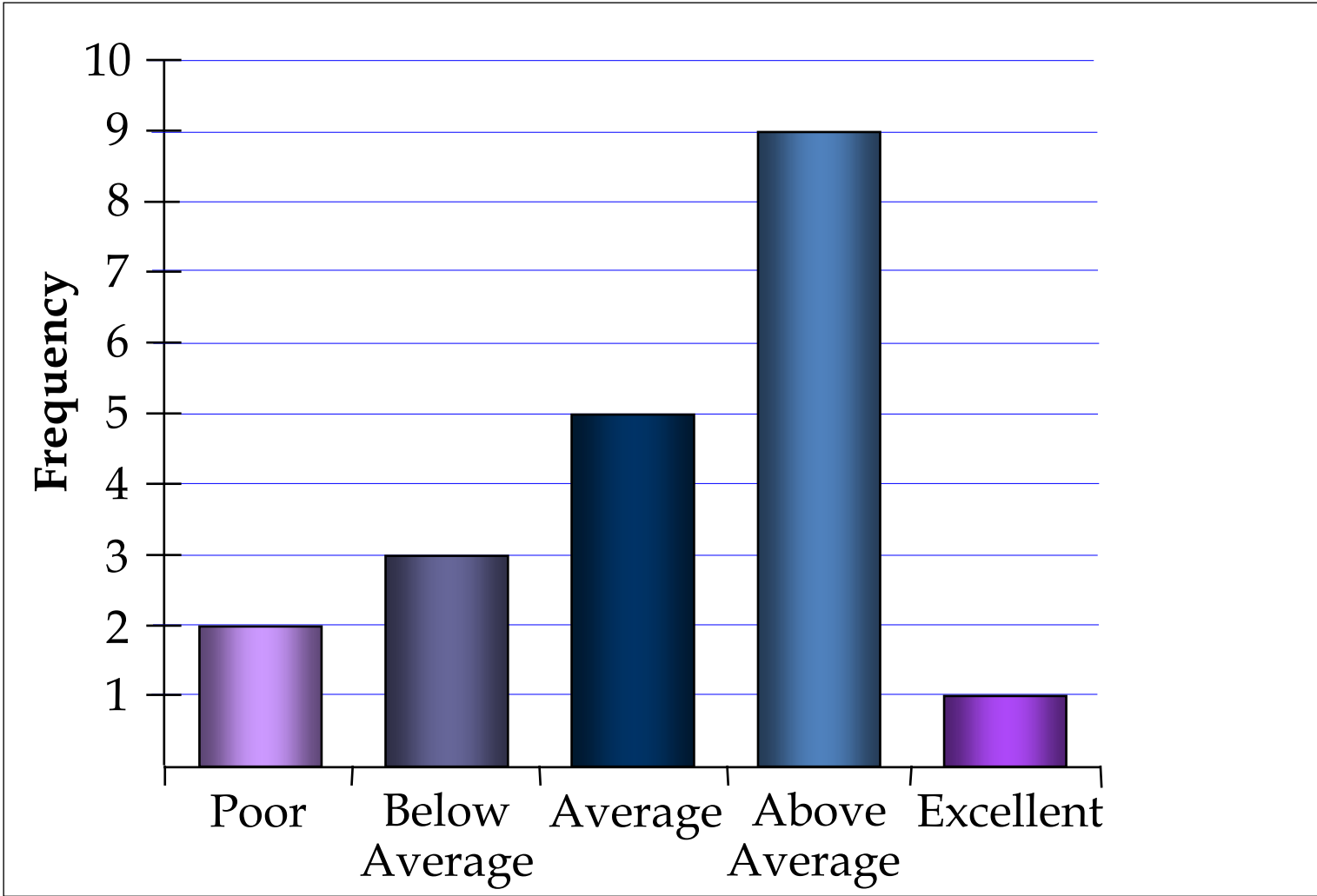


Bar Graph

- A bar graph is a graphical device for depicting qualitative data.
- On one axis (usually the horizontal axis), we specify the labels that are used for each of the classes.
- A frequency, relative frequency, or percent frequency scale can be used for the other axis (usually the vertical axis).
- Using a bar of fixed width drawn above each class label, we extend the height appropriately.
- The bars are separated to emphasize the fact that each class is a separate category.

Probabilistic and Statistical Techniques

Bar Graph



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)

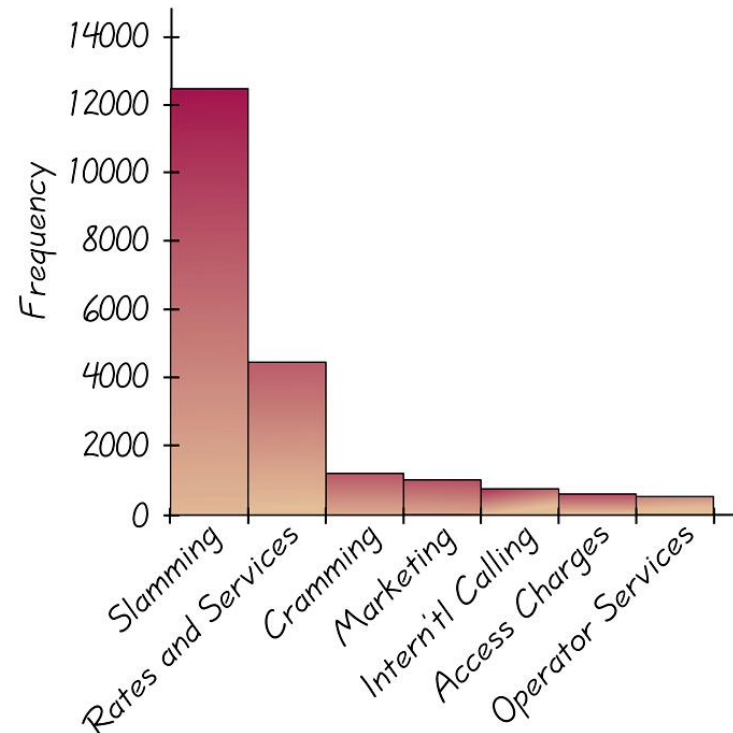
Stem (tens)	Leaves (units)
2	12445555666677778888999999
3	0011122333334445555555677888899
4	011111223569
5	04
6	013
7	4
8	0

←Values are 50 and 54.

←Value is 80.

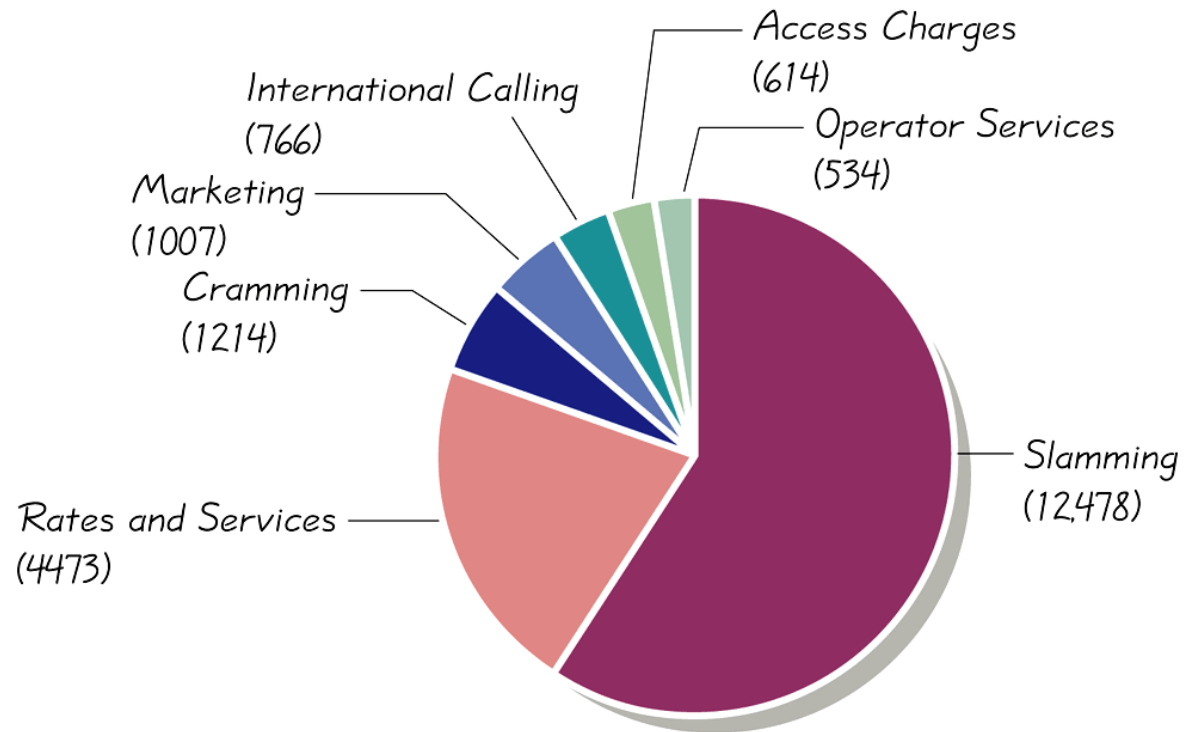
Pareto Chart

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



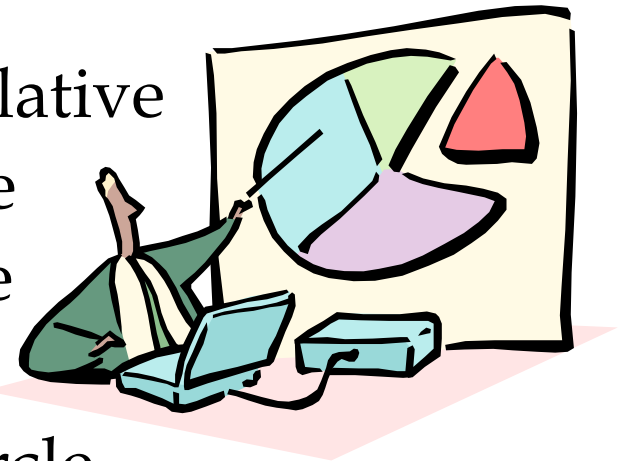
Pie Chart

A graph depicting qualitative data as slices of a pie



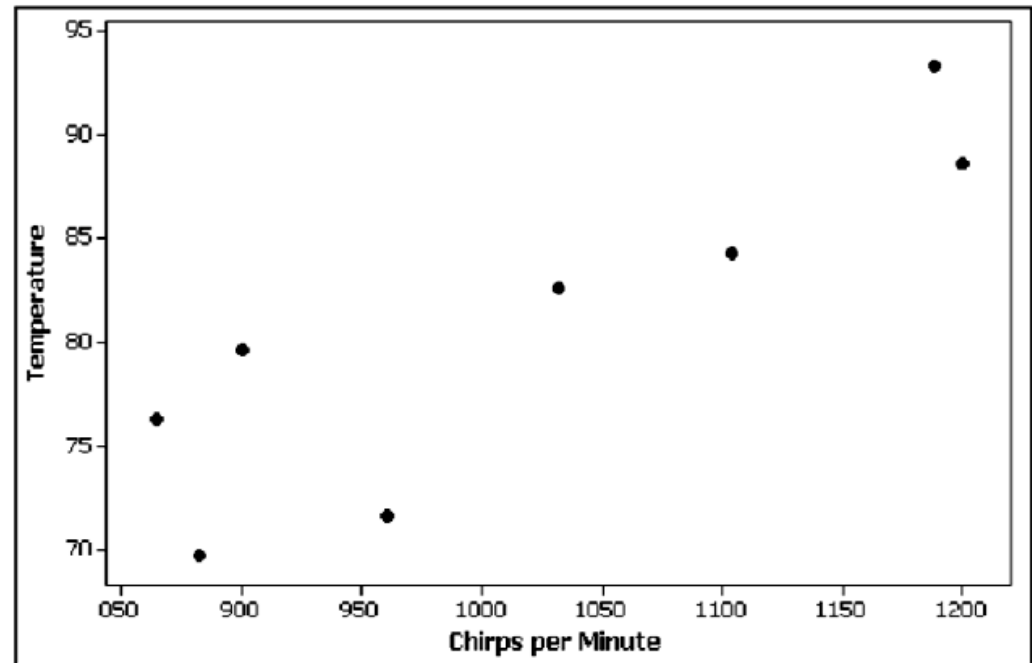
Pie Chart

- ▶ ■ The pie chart is a commonly used graphical device for presenting relative frequency distributions for qualitative data.
- ▶ □ First draw a circle; then use the relative frequencies to subdivide the circle into sectors that correspond to the relative frequency for each class.
- ▶ □ Since there are 360 degrees in a circle, a class with a relative frequency of .25 would consume $.25(360) = 90$ degrees of the circle.



Scatter Plot (or Scatter Diagram)

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



Probabilistic and Statistical Techniques

