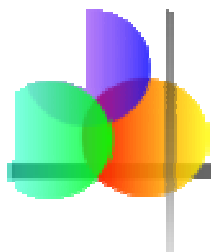
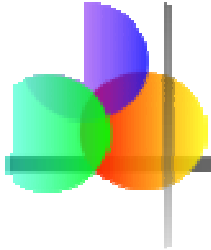


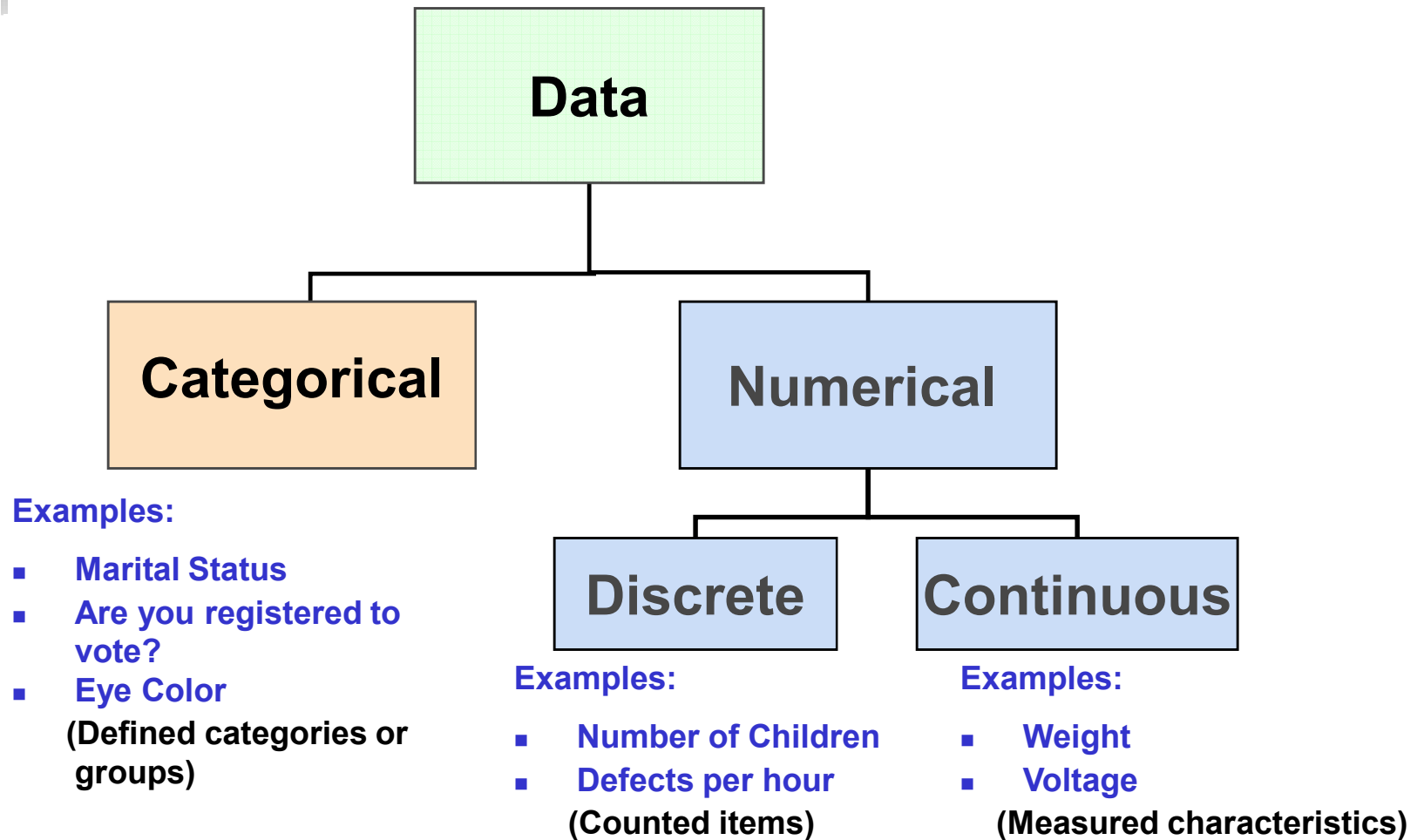
Describing Data: Graphical

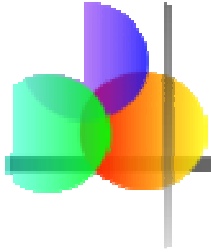


Dr. Md. Israt Rayhan
Associate Professor
Institute of Statistical Research and Training (ISRT)
University of Dhaka
Email: israt@isrt.ac.bd



Types of Data





Measurement Levels

Differences between measurements, true zero exists

Ratio Data

Quantitative Data

Differences between measurements but no true zero

Interval Data

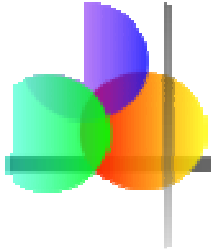
Ordered Categories (rankings, order, or scaling)

Ordinal Data

Qualitative Data

Categories (no ordering or direction)

Nominal Data



Graphical Presentation of Data

(continued)

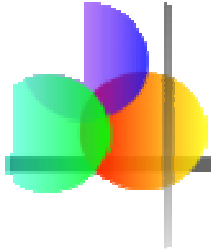
- Techniques reviewed in this chapter:

Categorical Variables

- Frequency distribution
- Bar chart
- Pie chart

Numerical Variables

- Line chart
- Frequency distribution
- Histogram and ogive
- Stem-and-leaf display



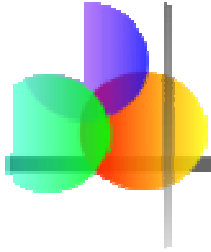
The Frequency Distribution Table

Summarize data by category

Example: Hospital Patients by Unit

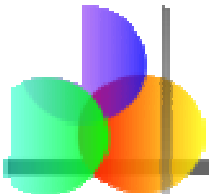
Hospital Unit	Number of Patients
Cardiac Care	1,052
Emergency	2,245
Intensive Care	340
Maternity	552
Surgery	4,630

(Variables are categorical)



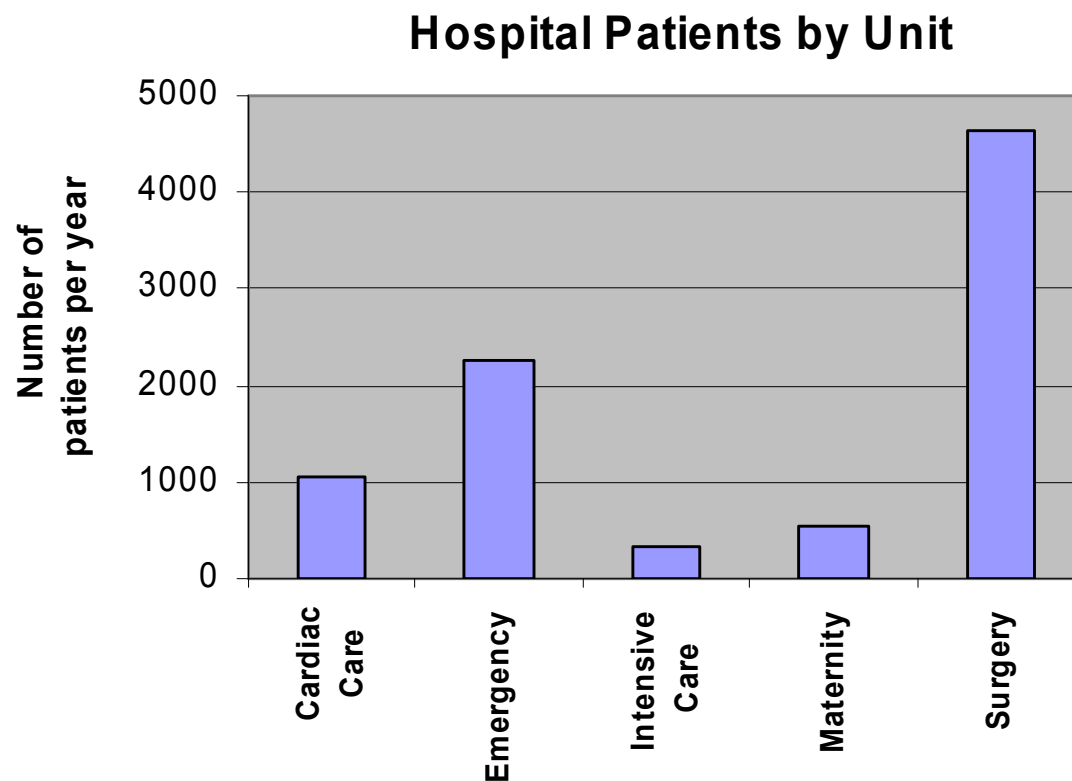
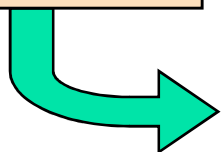
Bar and Pie Charts

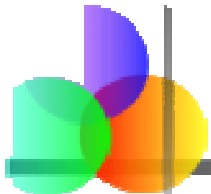
- **Bar charts** and **Pie charts** are often used for qualitative (category) data
- Height of bar or size of pie slice shows the frequency or percentage for each category



Bar Chart Example

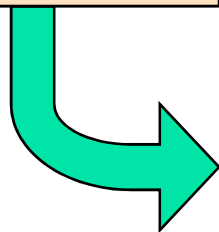
Hospital Unit	Number of Patients
Cardiac Care	1,052
Emergency	2,245
Intensive Care	340
Maternity	552
Surgery	4,630





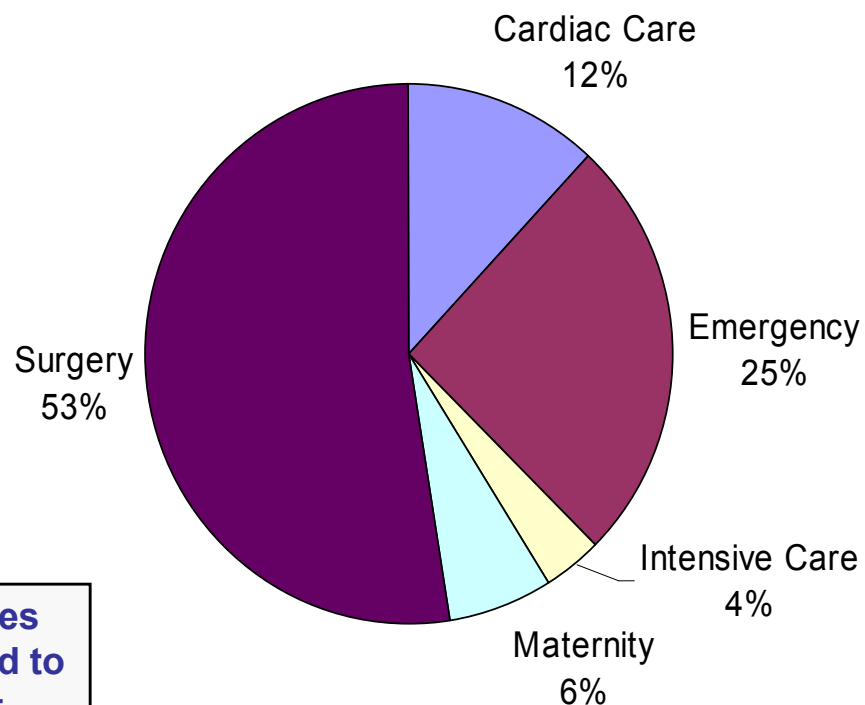
Pie Chart Example

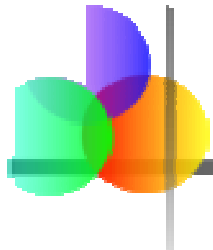
Hospital Unit	Number of Patients	% of Total
Cardiac Care	1,052	11.93
Emergency	2,245	25.46
Intensive Care	340	3.86
Maternity	552	6.26
Surgery	4,630	52.50



(Percentages are rounded to the nearest percent)

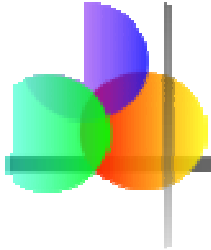
Hospital Patients by Unit



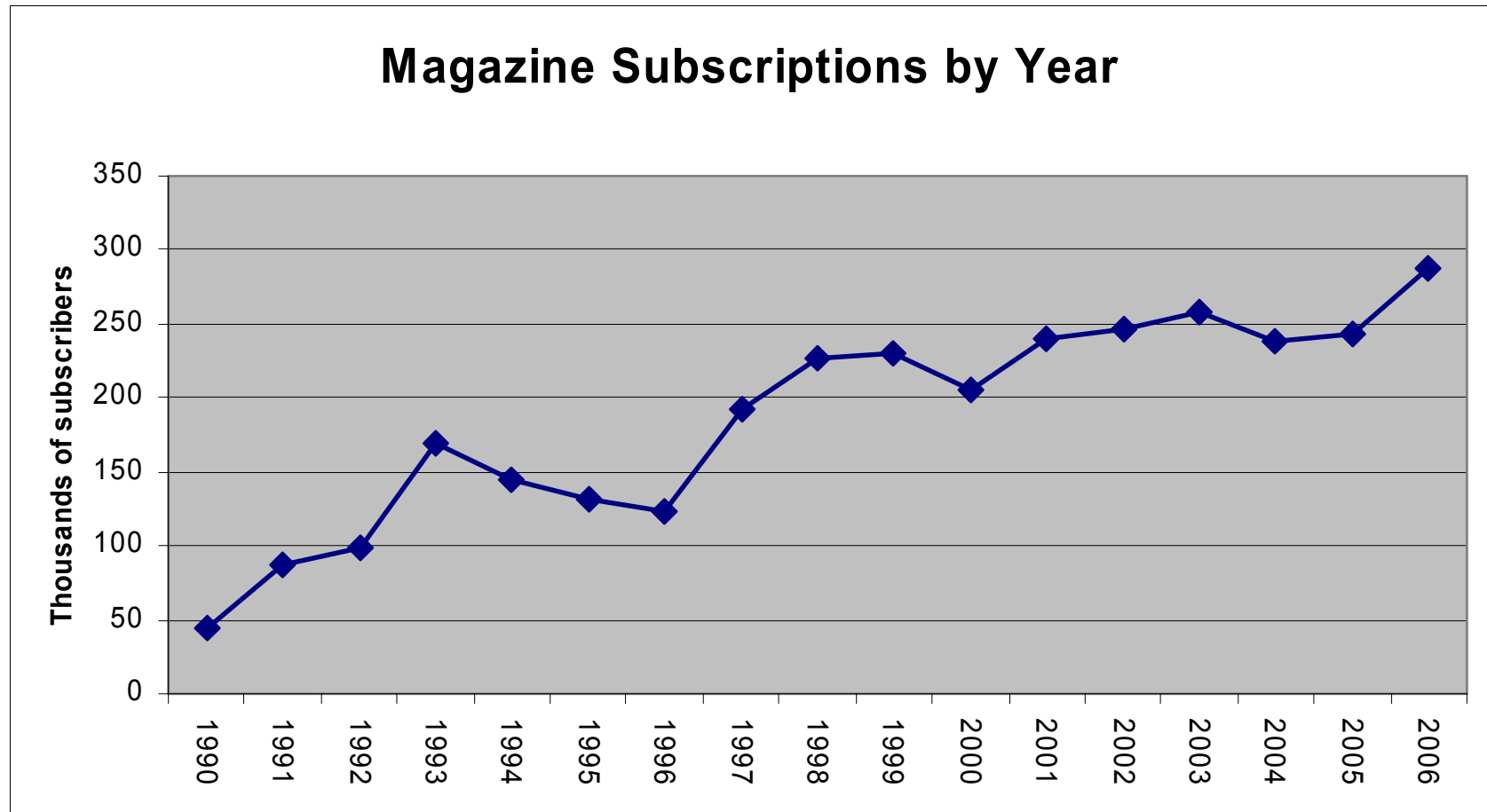


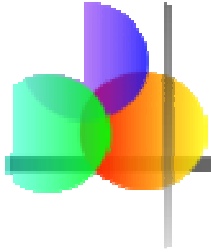
Graphs for Time-Series Data

- A **line chart** (time-series plot) is used to show the values of a variable over time
- Time is measured on the horizontal axis
- The variable of interest is measured on the vertical axis

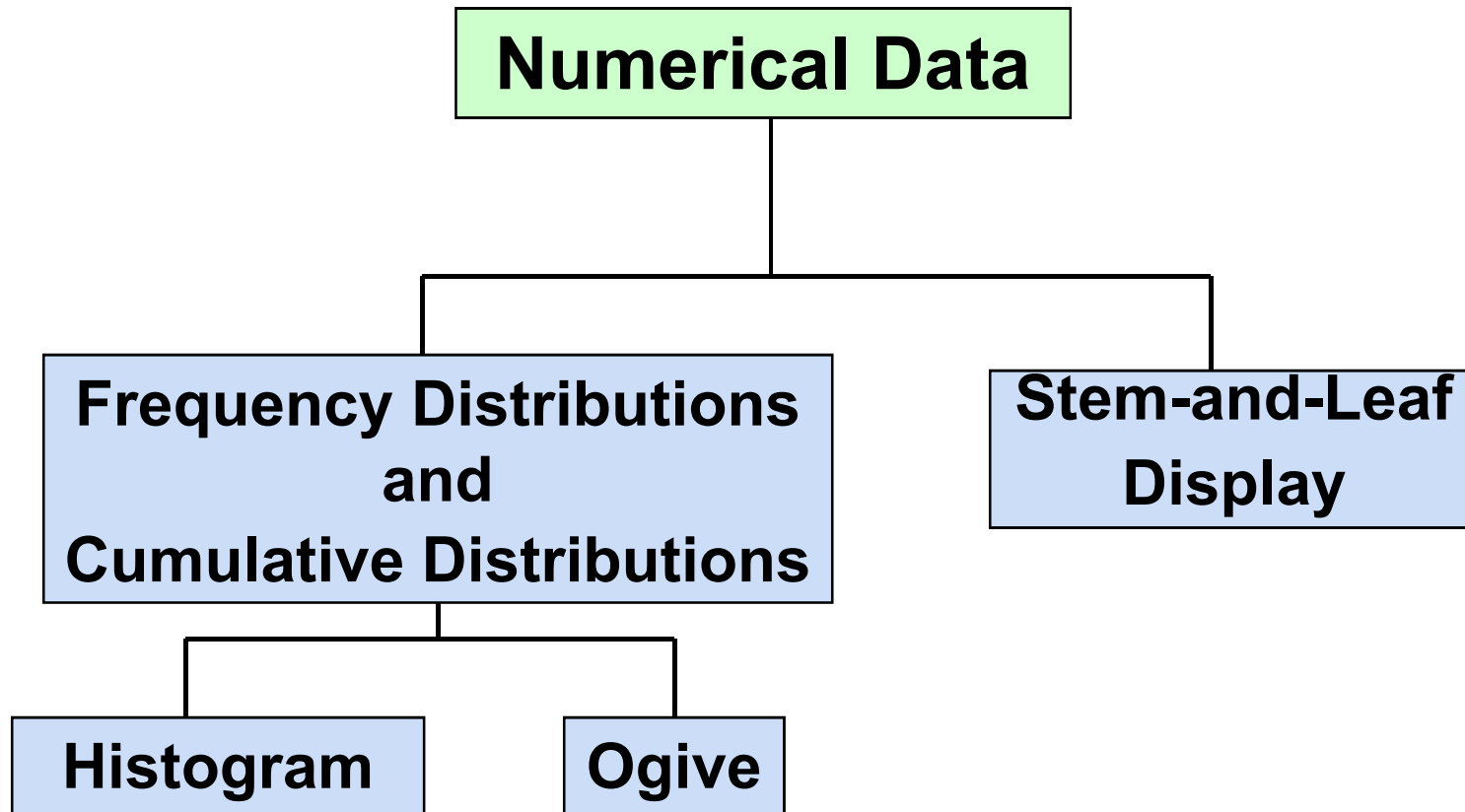


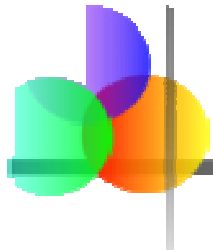
Line Chart Example





Graphs to Describe Numerical Variables

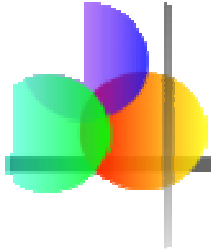




Frequency Distributions

What is a Frequency Distribution?

- A frequency distribution is a **list or a table** ...
- containing **class groupings** (categories or ranges within which the data fall) ...
- and the **corresponding frequencies** with which data fall within each class or category

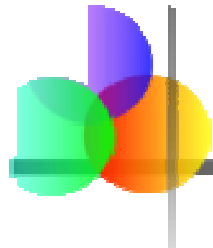


Class Intervals and Class Boundaries

- Each class grouping has the same width
- Determine the width of each interval by

$$w = \text{interval width} = \frac{\text{largest number} - \text{smallest number}}{\text{number of desired intervals}}$$

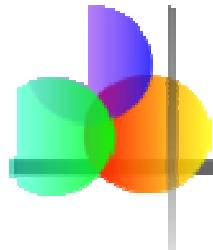
- Use at least 5 but no more than 15-20 intervals
- Intervals never overlap
- Round up the interval width to get desirable interval endpoints



Frequency Distribution Example

Example: A manufacturer of insulation randomly selects 20 winter days and records the **daily high temperature**

**24, 35, 17, 21, 24, 37, 26, 46, 58, 30,
32, 13, 12, 38, 41, 43, 44, 27, 53, 27**



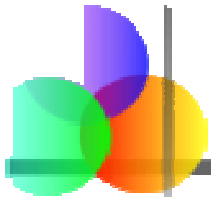
Frequency Distribution Example

(continued)

- Sort raw data in ascending order:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

- Find range: $58 - 12 = 46$
- Select number of classes: **5 (usually between 5 and 15)**
- Compute interval width: **10** ($46/5$ then round up)
- Determine interval boundaries: **10 but less than 20, 20 but less than 30, . . . , 60 but less than 70**
- Count observations & assign to classes



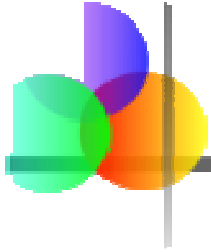
Frequency Distribution Example

(continued)

Data in ordered array:

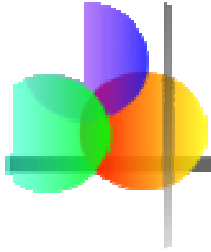
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Interval	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15
20 but less than 30	6	.30	30
30 but less than 40	5	.25	25
40 but less than 50	4	.20	20
50 but less than 60	2	.10	10
Total	20	1.00	100



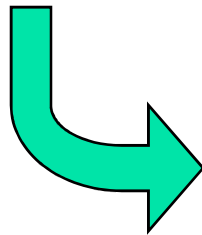
Histogram

- A graph of the data in a frequency distribution is called a **histogram**
- The **interval endpoints** are shown on the horizontal axis
- the vertical axis is either **frequency, relative frequency, or percentage**
- Bars of the appropriate heights are used to represent the number of observations within each class



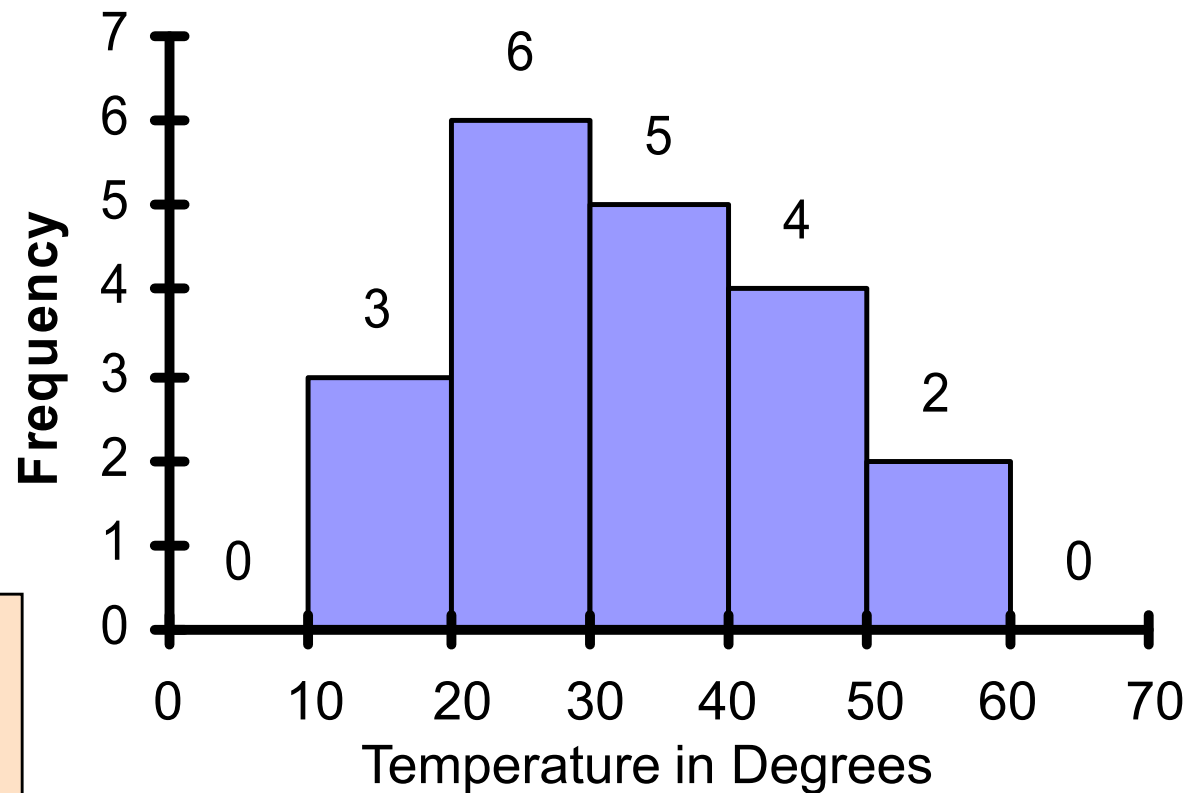
Histogram Example

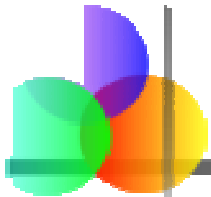
Interval	Frequency
10 but less than 20	3
20 but less than 30	6
30 but less than 40	5
40 but less than 50	4
50 but less than 60	2



(No gaps
between
bars)

Histogram : Daily High Temperature



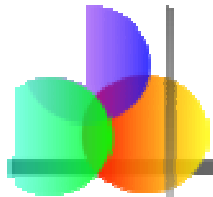


The Cumulative Frequency Distribution

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

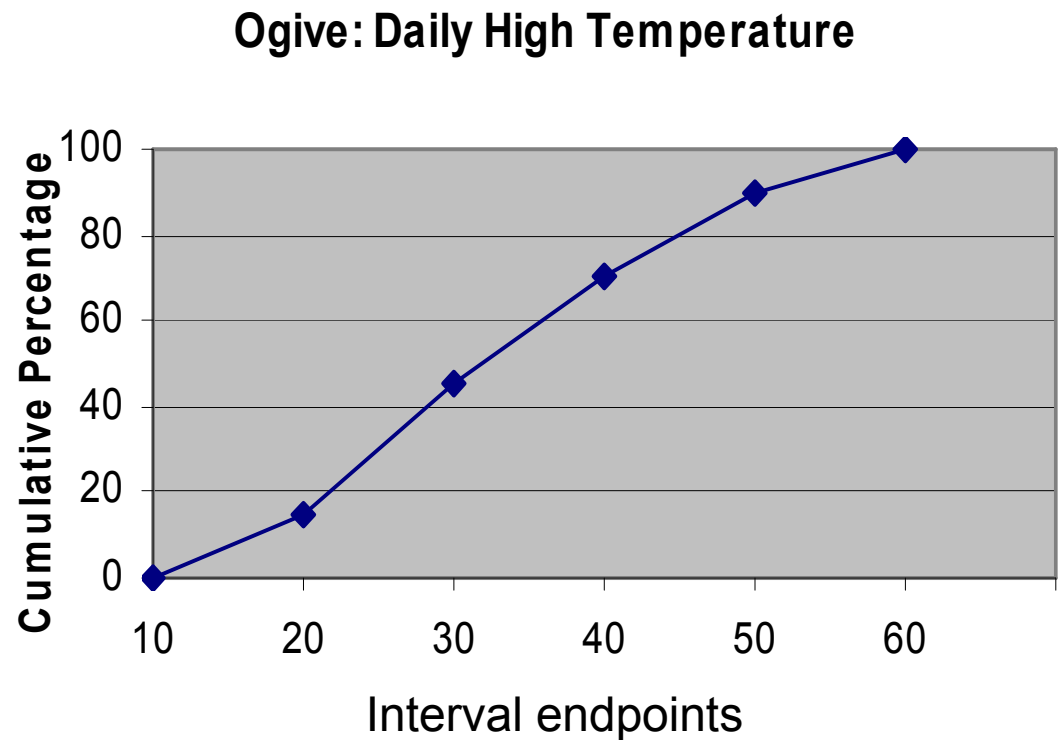
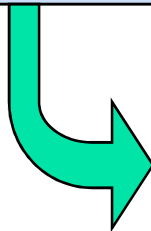
Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15	3	15
20 but less than 30	6	30	9	45
30 but less than 40	5	25	14	70
40 but less than 50	4	20	18	90
50 but less than 60	2	10	20	100
Total	20	100		

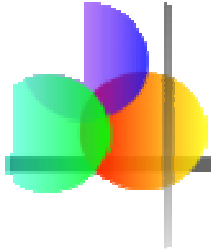


The Ogive

Graphing Cumulative Frequencies

Interval	Upper interval endpoint	Cumulative Percentage
Less than 10	10	0
10 but less than 20	20	15
20 but less than 30	30	45
30 but less than 40	40	70
40 but less than 50	50	90
50 but less than 60	60	100

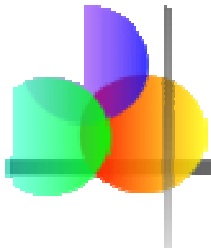




Stem-and-Leaf Diagram

- A simple way to see distribution details in a data set

METHOD: Separate the sorted data series
into leading digits (the **stem**) and
the trailing digits (the **leaves**)



Example

Data in ordered array:

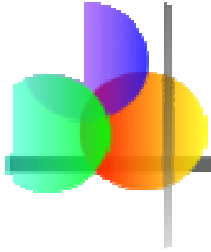
21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Here, use the 10's digit for the stem unit:

- 21 is shown as

- 38 is shown as

Stem	Leaf
2	1
3	8



Example

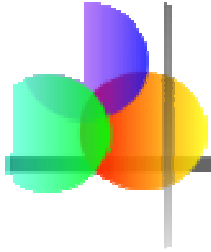
(continued)

Data in ordered array:

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Completed stem-and-leaf diagram:

Stem	Leaves
2	1 4 4 6 7 7
3	0 2 8
4	1

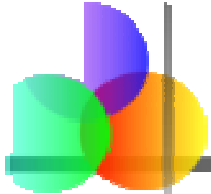


Data Presentation Errors

(continued)

- Unequal histogram interval widths
- Compressing or distorting the vertical axis
- Providing no zero point on the vertical axis
- Failing to provide a relative basis in comparing data between groups





Reference

- Statistics for Business and Economics (2007, 6th edition) by Paul Newbold.