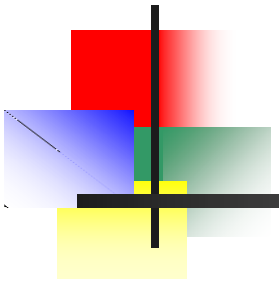
Statistics for Managers Using Microsoft® Excel 4th Edition



Chapter 2

Presenting Data in Tables and Charts

Chapter Goals

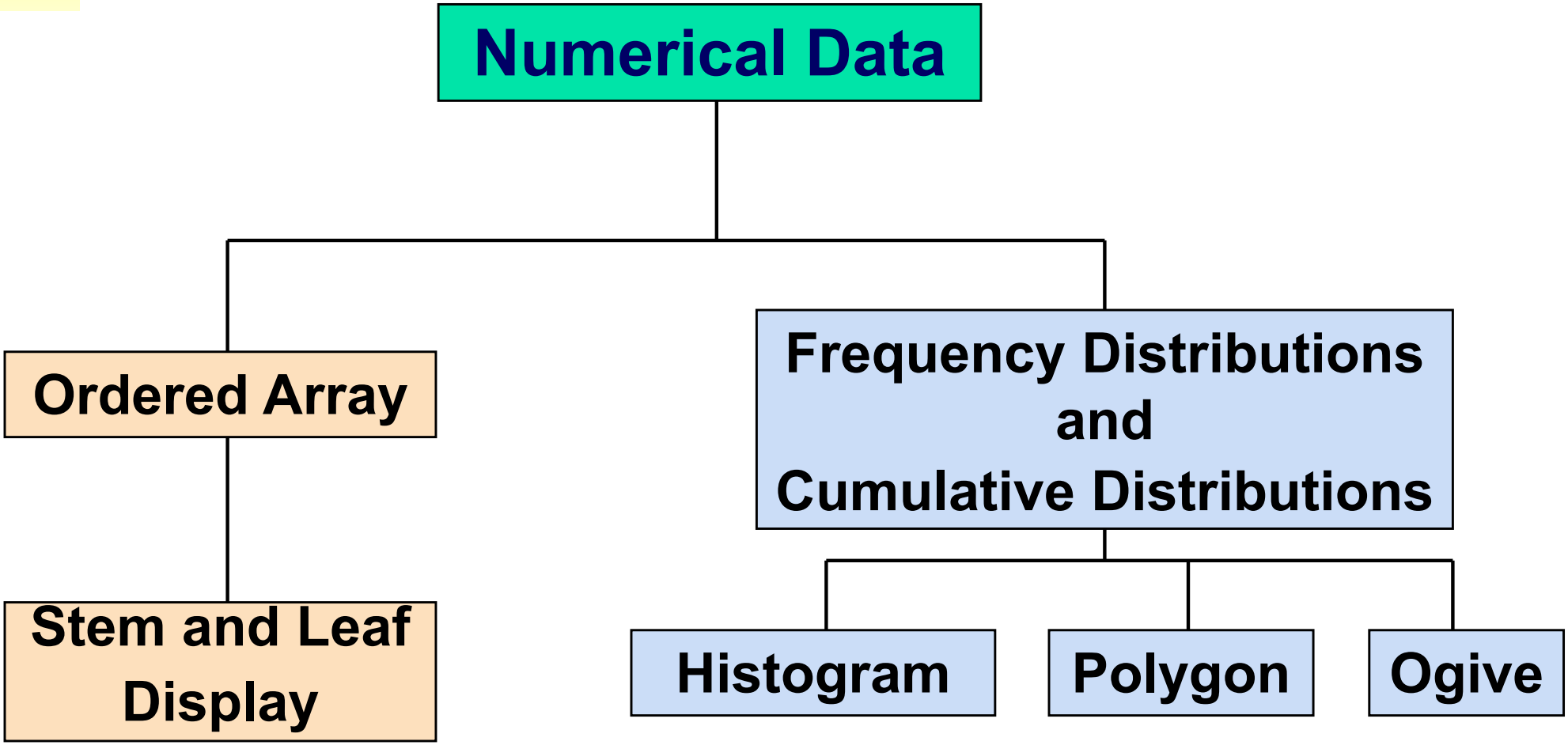
After completing this chapter, you should be able to:

- Create an ordered array and a stem-and-leaf display
- Construct and interpret a frequency distribution, histogram, polygon, and ogive
- Create and interpret bar charts, pie charts, and scatter diagrams for numerical data
- Present and interpret categorical data in bar charts and pie charts
- Describe appropriate and inappropriate ways to display data graphically

Organizing and Presenting Data Graphically

- Data in raw form are usually not easy to use for decision making
  - Some type of organization is needed
    - Table
    - Graph
- Techniques reviewed here:
  - Ordered Array
  - Stem-and-Leaf Display
  - Frequency Distributions and Histograms
  - Bar charts and pie charts
  - Contingency tables

Tables and Charts for Numerical Data



The Ordered Array

Just a **sorted list** of data:

- Shows range (min to max)
- Provides some signals about variability within the range
- May help identify outliers (unusual observations)
- If the data set is large, the ordered array is less useful

The Ordered Array

(continued)

- Data in raw form (as collected):

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

- Data in ordered array from smallest to largest:

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

Stem and Leaf Diagram

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

  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

Using other stem units

- Using the 100's digit as the stem:
  - Round off the 10's digit to form the leaves

- 613 would become
- 776 would become
- . . .
- 1224 becomes

Stem	Leaf
6	1
7	8
12	2

Using other stem units

(continued)

- Using the 100's digit as the stem:
  - The completed stem-and-leaf display:

Data:

613, 632, 658, 717,
722, 750, 776, 827,
841, 859, 863, 891,
894, 906, 928, 933,
955, 982, 1034,
1047,1056, 1140,
1169, 1224

Stem	Leaves
6	1 3 6
7	2 2 5 8
8	3 4 6 6 9 9
9	1 3 3 6 8
10	3 5 6
11	4 7
12	2

Tabulating Numerical Data: Frequency Distributions

A frequency distribution is a list or a table …

- containing class groupings (categories or ranges within which the data falls) ...
- and the corresponding frequencies with which data falls within each grouping or category

Why Use Frequency Distributions?

- A frequency distribution is a way to summarize data
- The distribution condenses the raw data into a more useful form...
- and allows for a quick visual interpretation of the data
- and easy graphical display

Class Intervals and Class Boundaries

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

$$Width\ of\ Interval \cong \frac{range}{number\ of\ desired\ class\ groupings}$$

- Use at least 5 but no more than 15 groupings
- Class boundaries never overlap
- Round up the interval width to get desirable endpoints

Approximate Number of Classes

Observations	Classes
Less than 50	3-6
50-200	6-9
200-1000	8-12
More than 1000	10-15

- Excel uses the square root of n.

Frequency Distribution 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 class interval (width): **10 (46/5 then round up)**
- Determine class boundaries (limits): **10, 20, 30, 40, 50, 60**
- Compute class midpoints: **15, 25, 35, 45, 55**
- 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

Class	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

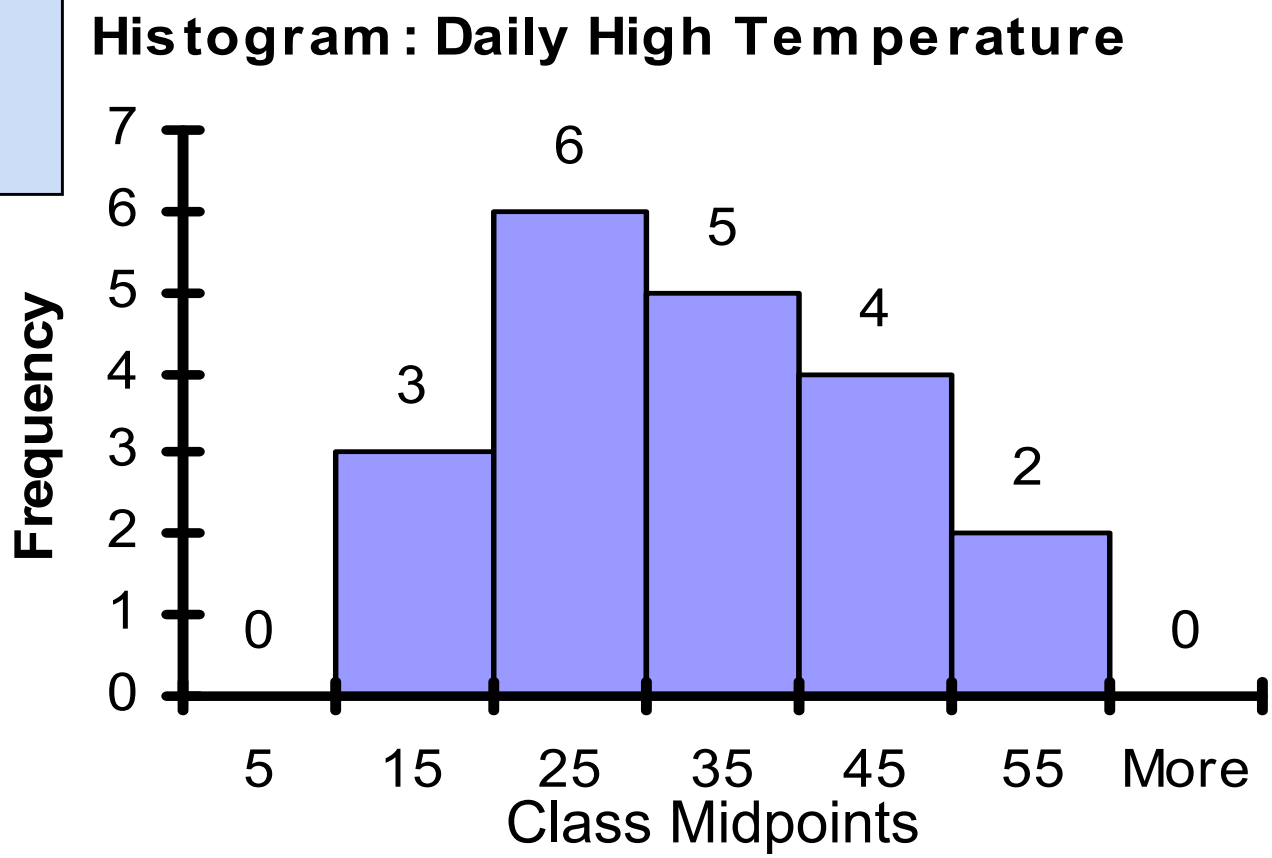
Graphing Numerical Data: The Histogram

- A graph of the data in a frequency distribution is called a **histogram**
- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis
- **frequency** is measured on the vertical axis
- Bars of the appropriate heights can be used to represent the number of observations within each class

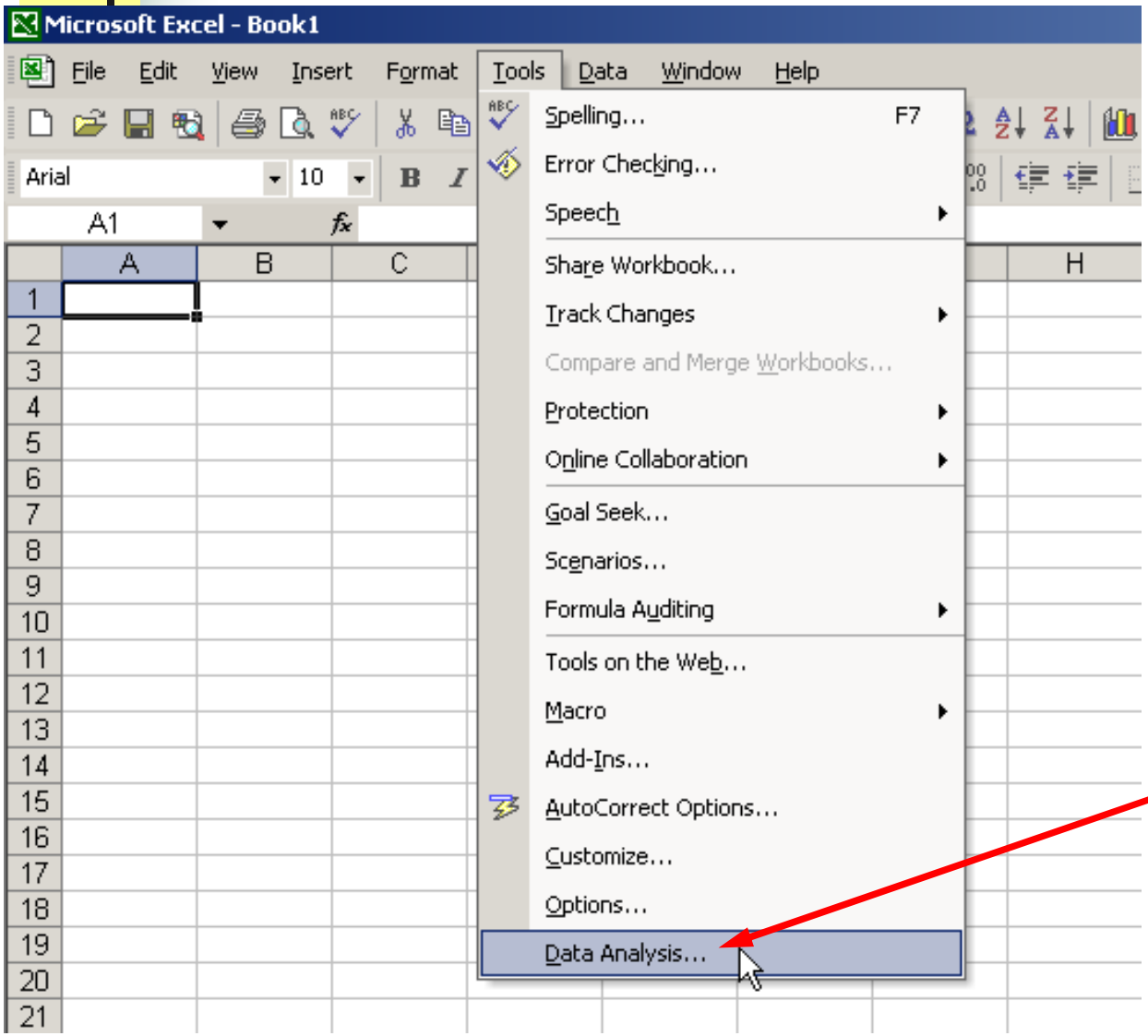
Histogram Example

Class	Class Midpoint	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2

(No gaps between bars)



Histograms in Excel



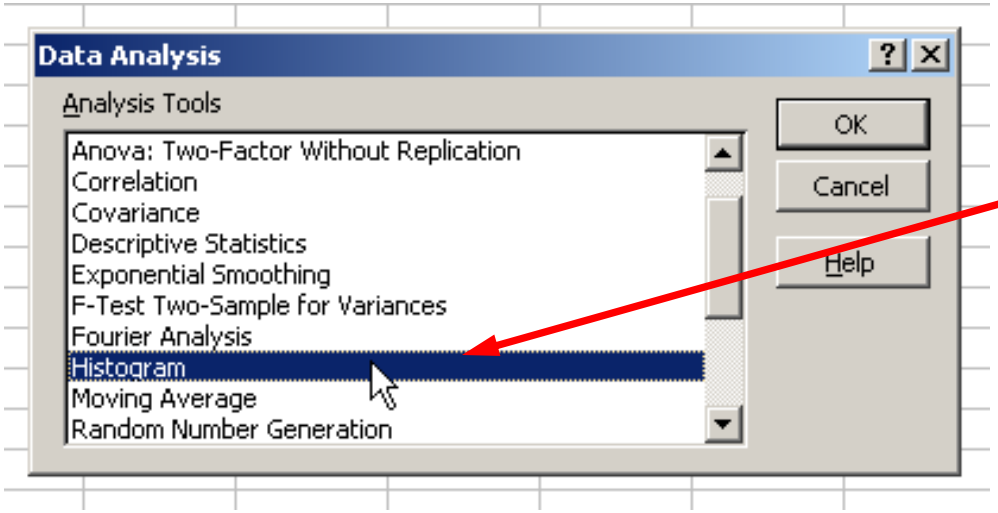
1

Select

Tools/Data Analysis

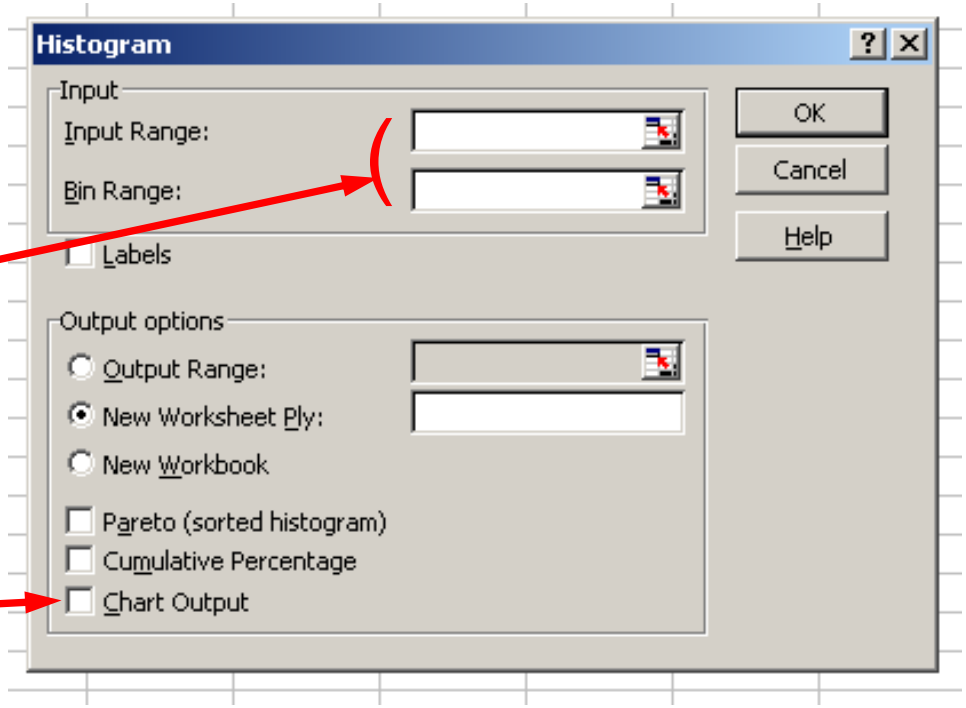
Histograms in Excel

(continued)



2

Choose Histogram



3

Input data range and bin range (bin range is a cell range containing the upper class boundaries for each class grouping)

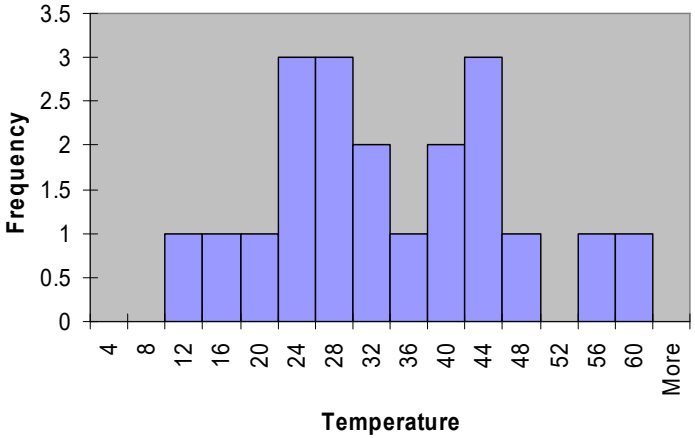
Select Chart Output and click "OK"

Questions for Grouping Data into Classes

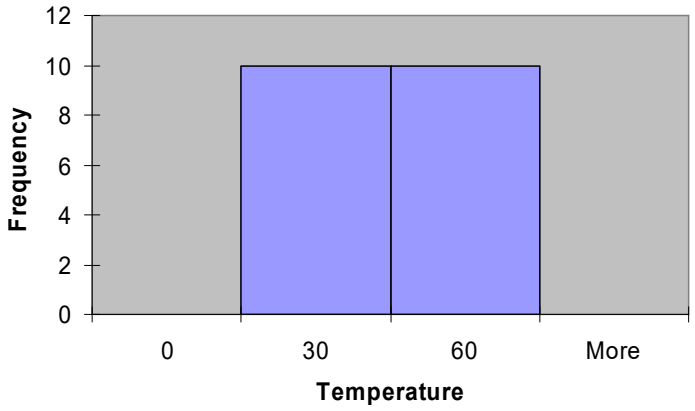
- 1. How wide should each interval be?
  (How many classes should be used?)
- 2. How should the endpoints of the intervals be determined?
  - Often answered by trial and error, subject to user judgment
  - The goal is to create a distribution that is neither too "jagged" nor too "blocky"
  - Goal is to appropriately show the pattern of variation in the data

How Many Class Intervals?

- **Many (Narrow class intervals)**
  - may yield a very jagged distribution with gaps from empty classes
  - Can give a poor indication of how frequency varies across classes



- **Few (Wide class intervals)**
  - may compress variation too much and yield a blocky distribution
  - can obscure important patterns of variation.

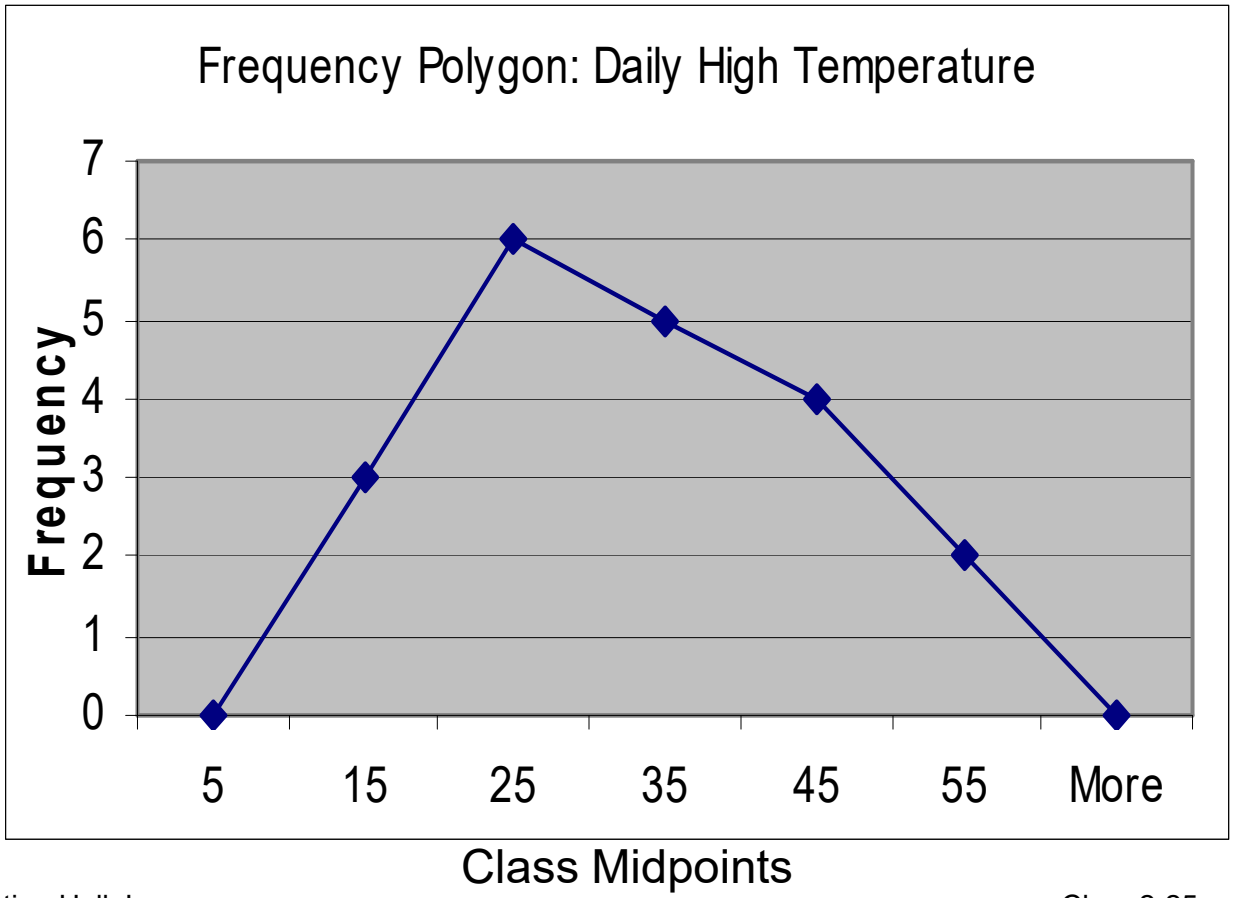


(X axis labels are upper class endpoints)

Graphing Numerical Data: The Frequency Polygon

Class	Class Midpoint	Frequency
1 but less than 10	5	0
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2
More than 60	65	0

(In a percentage polygon the vertical axis would be defined to show the percentage of observations per class)



Tabulating Numerical Data: Cumulative Frequency

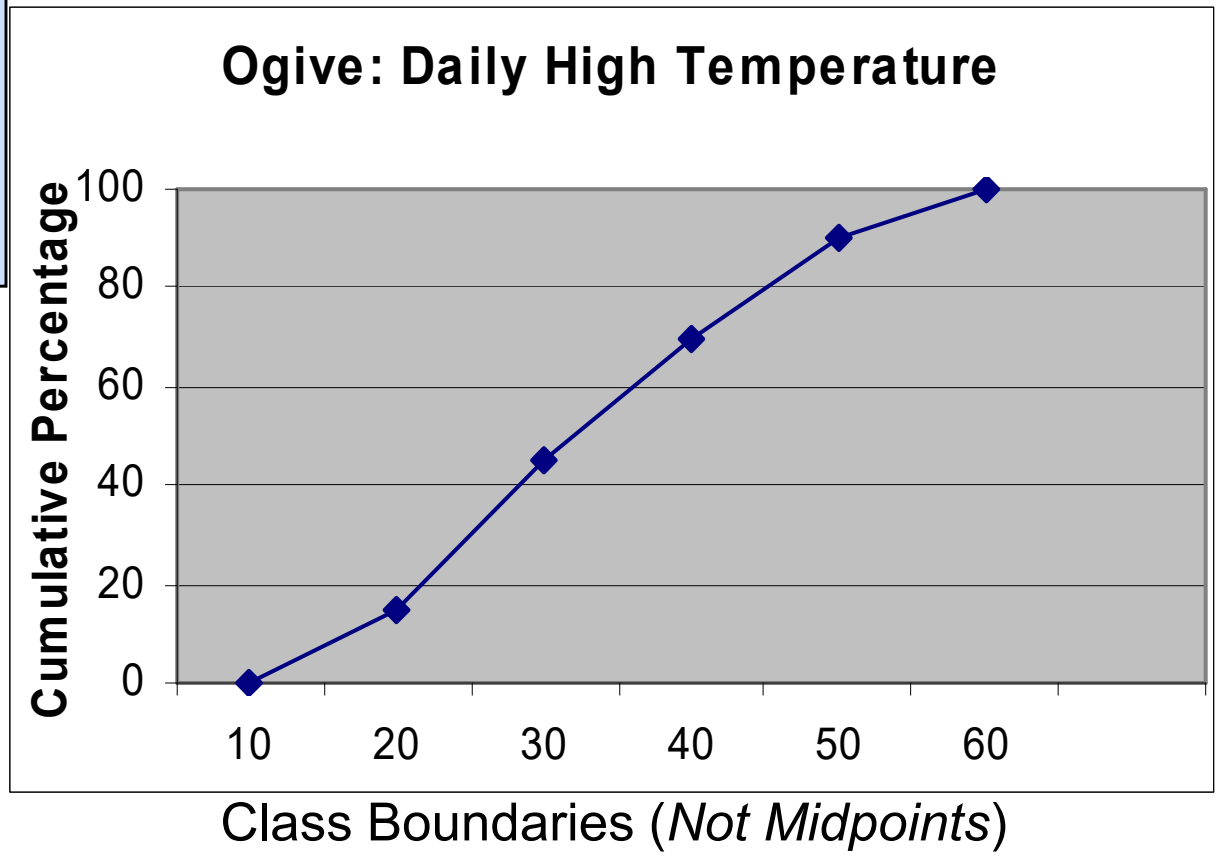
Data in ordered array:

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

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

Graphing Cumulative Frequencies: The Ogive (Cumulative % Polygon)

Class	Lower class boundary	Cumulative Percentage
less than 10	10	0
less than 20	20	15
less than 30	30	45
less than 40	40	70
less than 50	50	90
less than 60	60	100



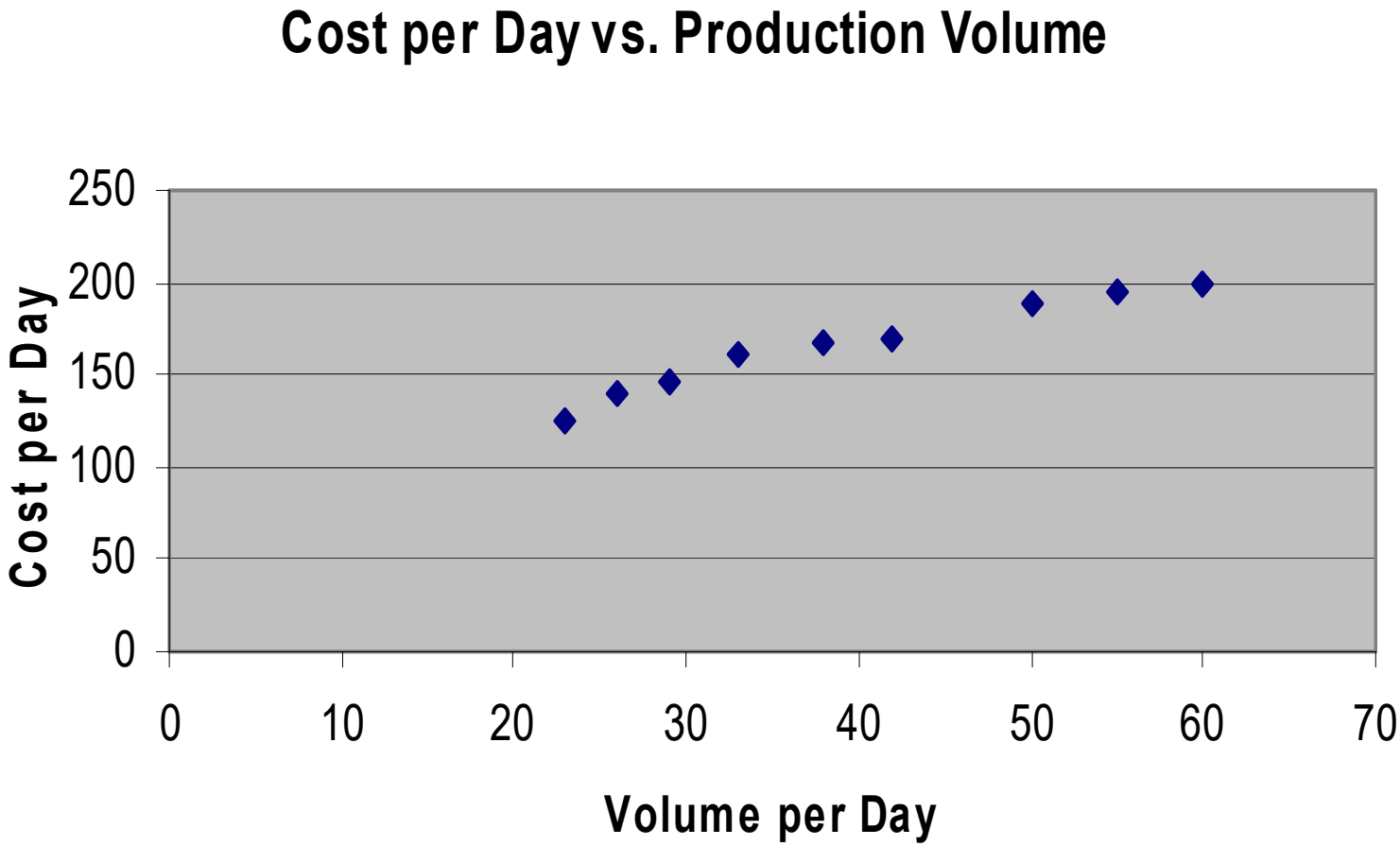
Scatter Diagrams

- Scatter Diagrams are used for bivariate numerical data
  - Bivariate data consists of paired observations taken from two numerical variables

- The Scatter Diagram:
  - one variable is measured on the vertical axis and the other variable is measured on the horizontal axis

Scatter Diagram Example

Volume per day	Cost per day
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

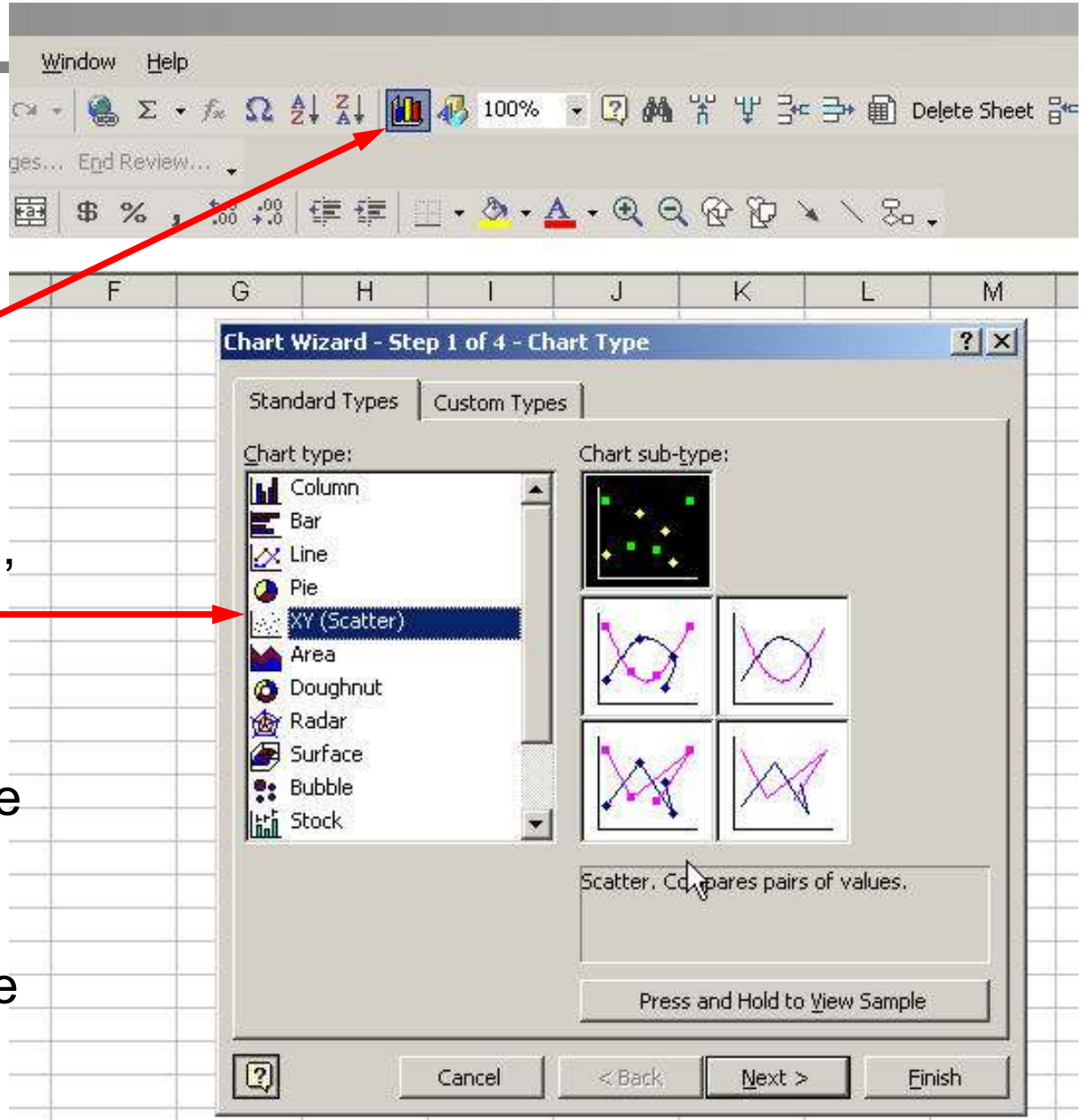


Scatter Diagrams in Excel

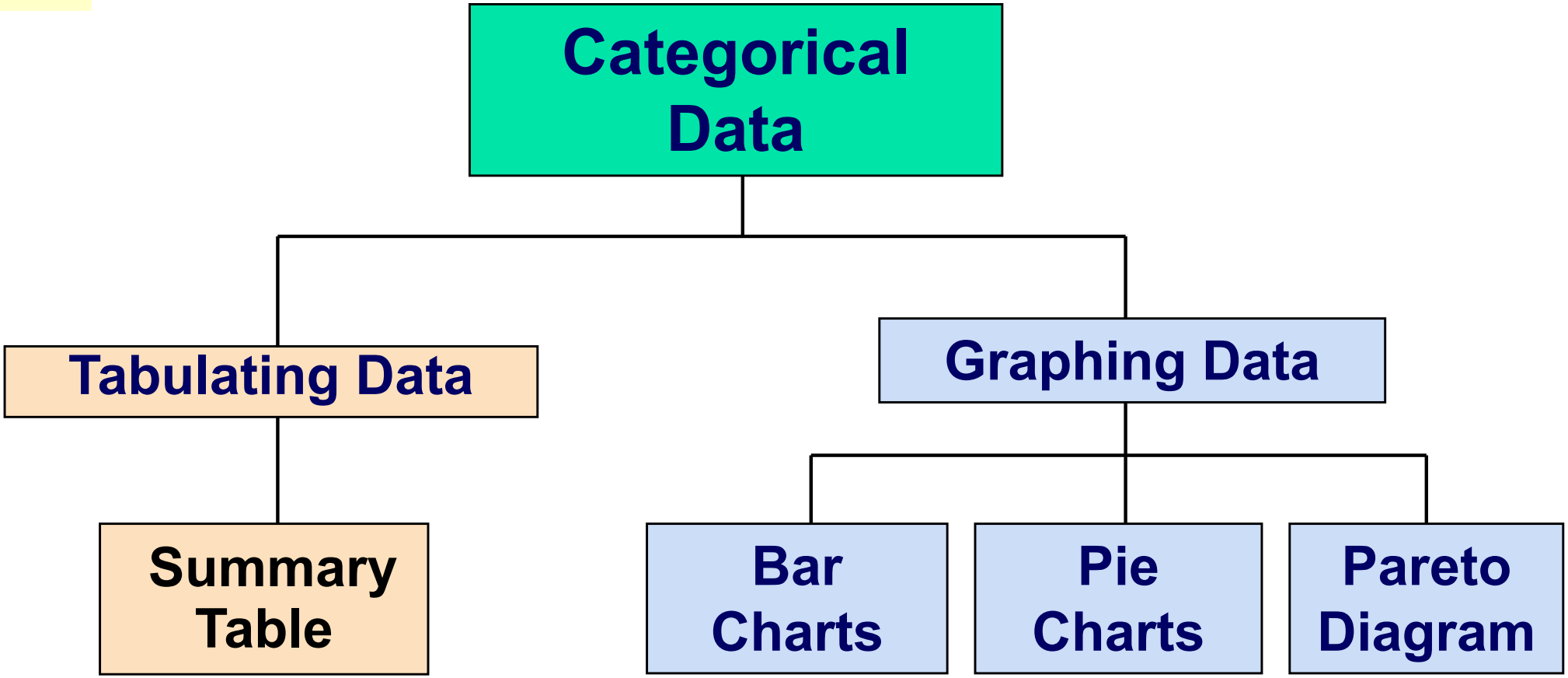
1. Select the chart wizard

2. Select XY(Scatter) option, then click “Next”

3. When prompted, enter the data range, desired legend, and desired destination to complete the scatter diagram



Tables and Charts for Categorical Data: Univariate Data



The Summary Table

Summarize data by category

Example: Current Investment Portfolio

Investment Type	Amount (in thousands $)	Percentage (%)
Stocks	46.5	42.27
Bonds	32.0	29.09
CD	15.5	14.09
Savings	16.0	14.55
Total	**110.0**	**100.0**

(Variables are Categorical)

Bar and Pie Charts

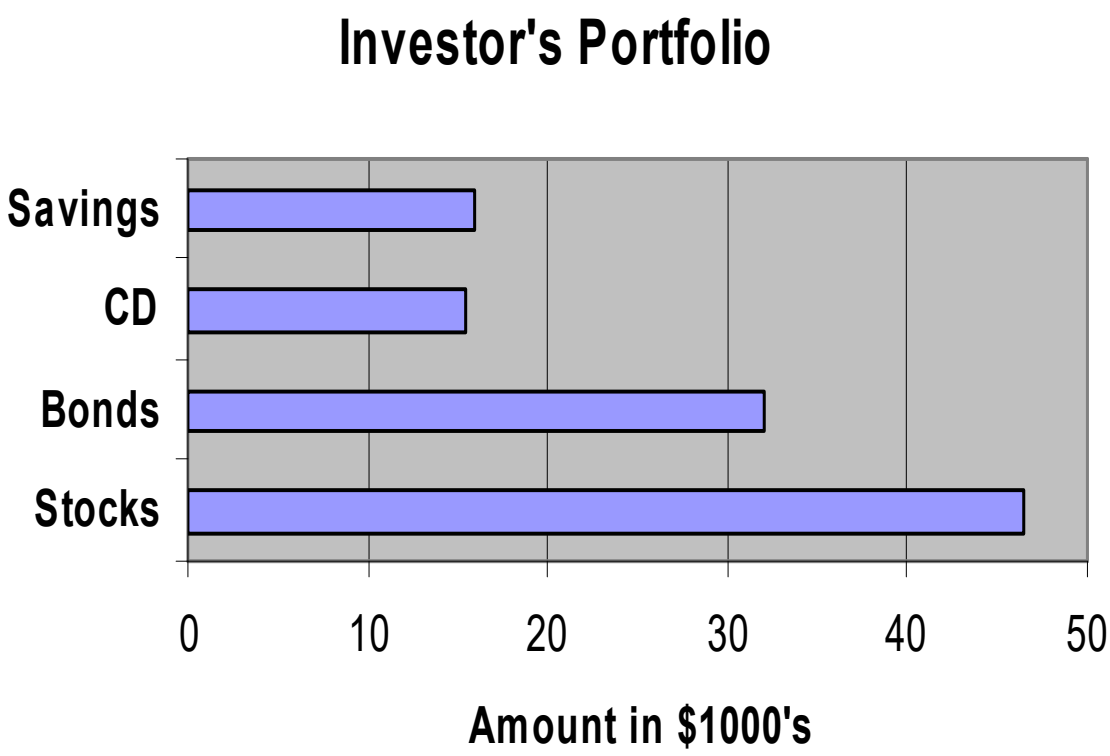
- Bar charts and Pie charts are often used for qualitative (category) data

- Height of bar or size of pie slice shows the frequency or percentage for each category

Bar Chart Example

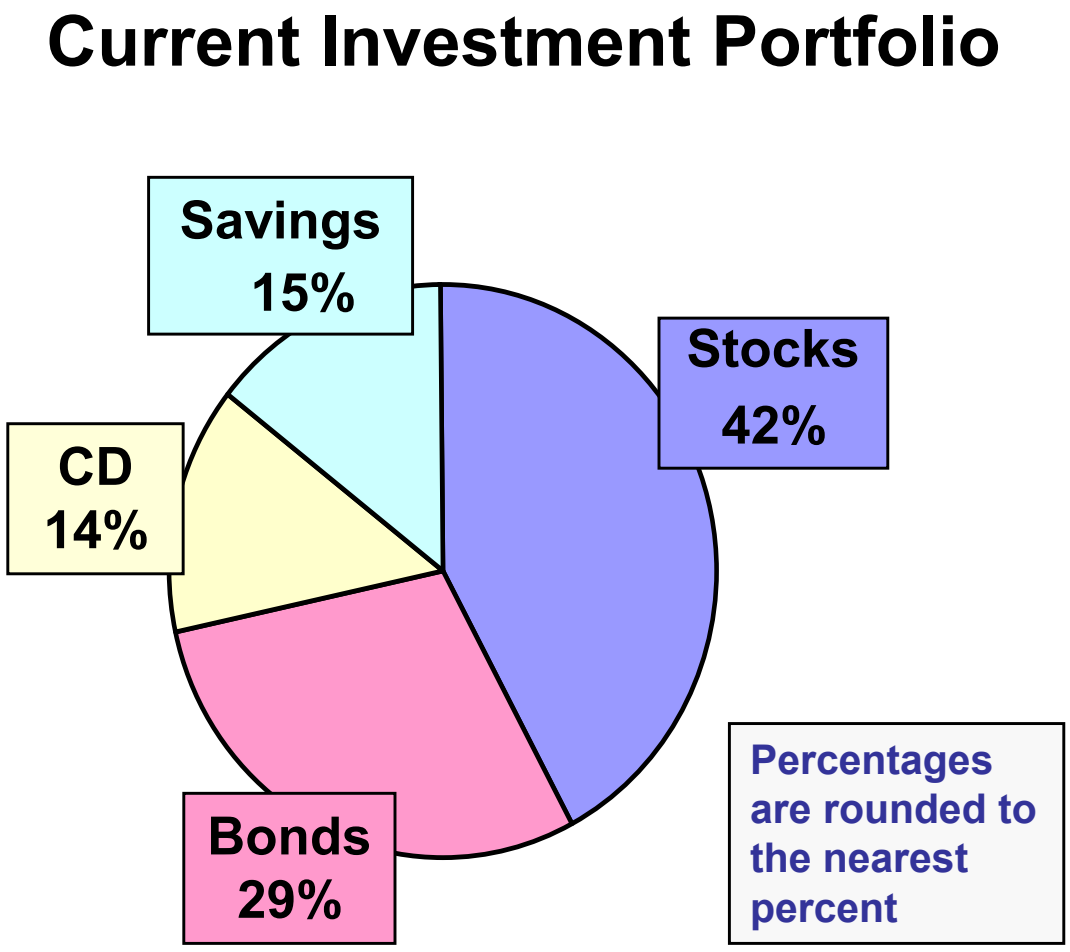
Current Investment Portfolio

Investment Type	Amount (in thousands $)	Percentage (%)
Stocks	46.5	42.27
Bonds	32.0	29.09
CD	15.5	14.09
Savings	16.0	14.55
Total	**110.0**	**100.0**

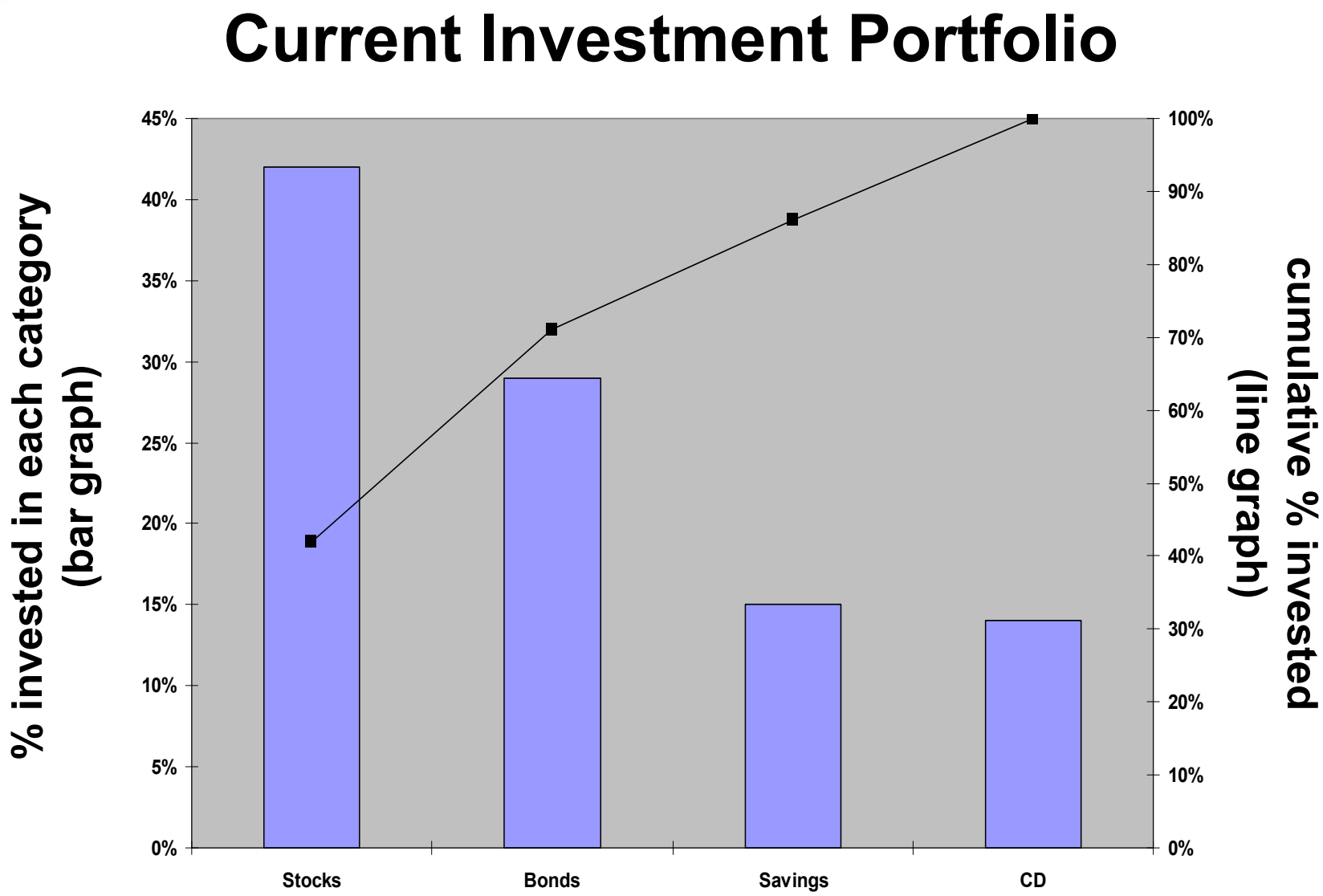


Pie Chart Example

Investment Type	Amount (in thousands $)	Percentage (%)
Stocks	46.5	42.27
Bonds	32.0	29.09
CD	15.5	14.09
Savings	16.0	14.55
Total	**110.0**	**100.0**



Pareto Diagram Example



Tabulating and Graphing Multivariate Categorical Data

- Contingency Table for Investment Choices ($1000's)

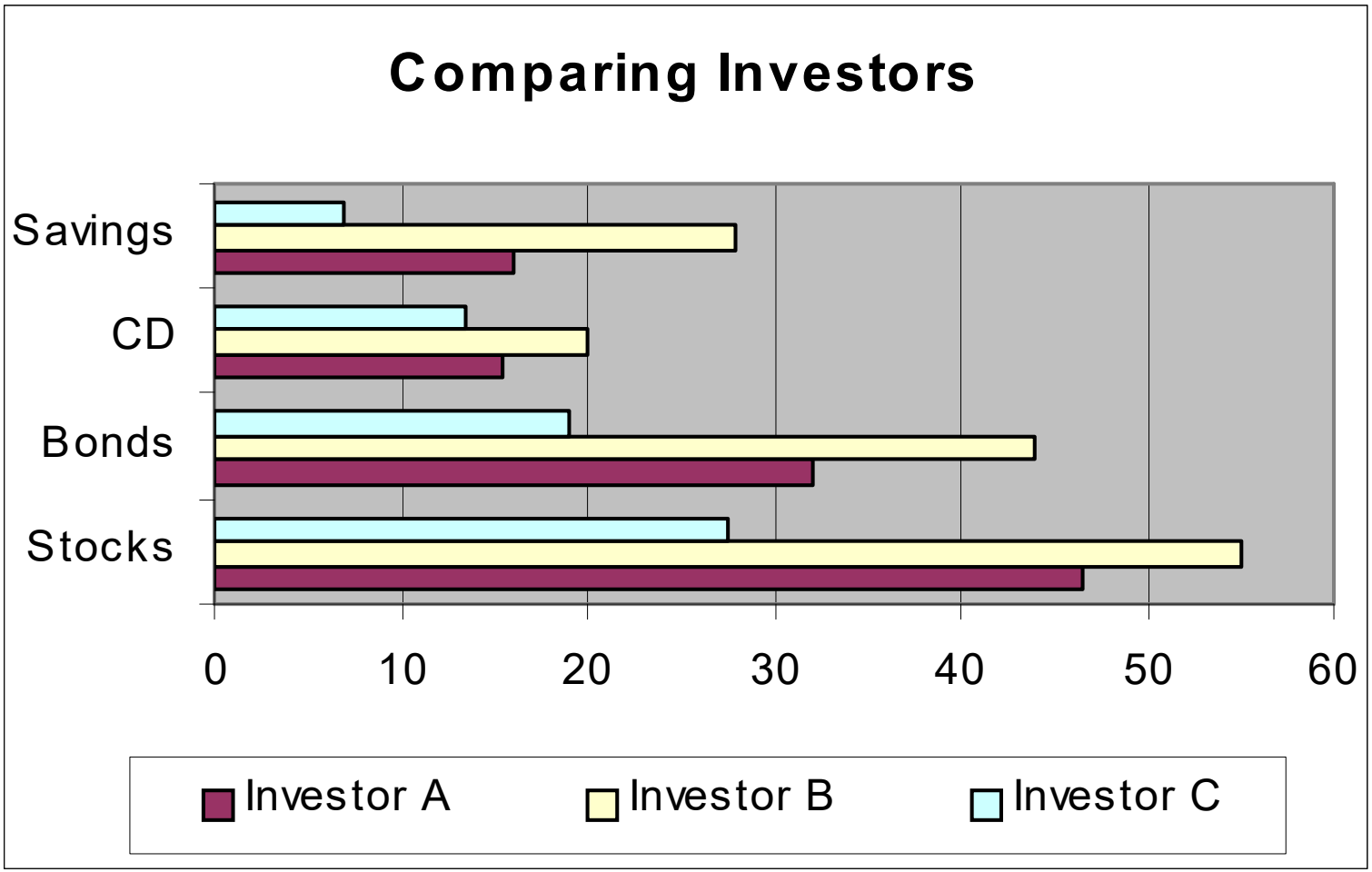
Investment Category	Investor A	Investor B	Investor C	Total
Stocks	46.5	55	27.5	**129**
Bonds	32.0	44	19.0	**95**
CD	15.5	20	13.5	**49**
Savings	16.0	28	7.0	**51**
Total	**110.0**	**147**	**67.0**	**324**

(Individual values could also be expressed as percentages of the overall total, percentages of the row totals, or percentages of the column totals)

Tabulating and Graphing Multivariate Categorical Data

(continued)

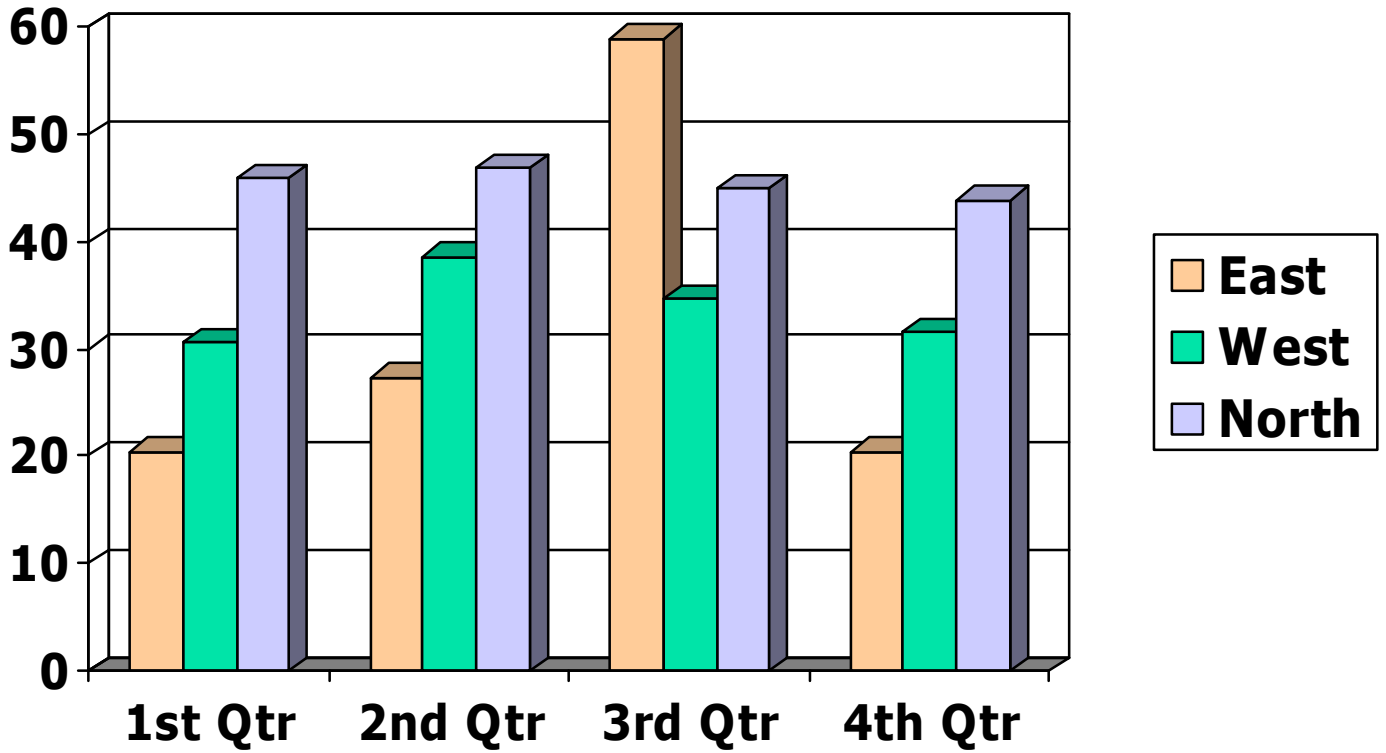
- Side by side bar charts



Side-by-Side Chart Example

- Sales by quarter for three sales territories:

	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East	20.4	27.4	59	20.4
West	30.6	38.6	34.6	31.6
North	45.9	46.9	45	43.9



Principles of Graphical Excellence

- Present data in a way that provides substance, statistics and design
- Communicate complex ideas with clarity, precision and efficiency
- Give the largest number of ideas in the most efficient manner
- Excellence almost always involves several dimensions
- Tells the truth about the data

"Chart Junk" and "Lie Factor"

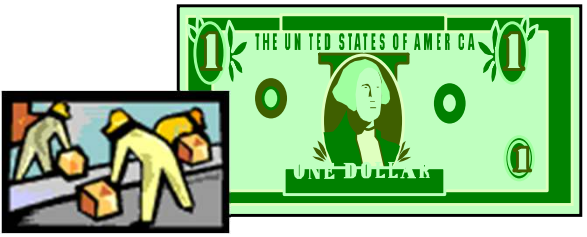
Bad Presentation

Minimum Wage


1960: $1.00

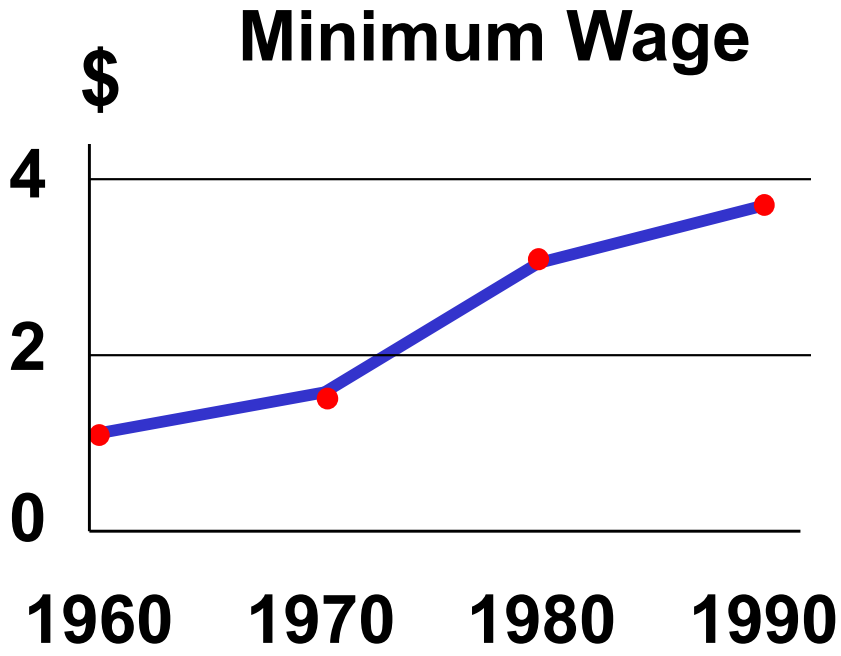

1970: $1.60

1980: $3.10



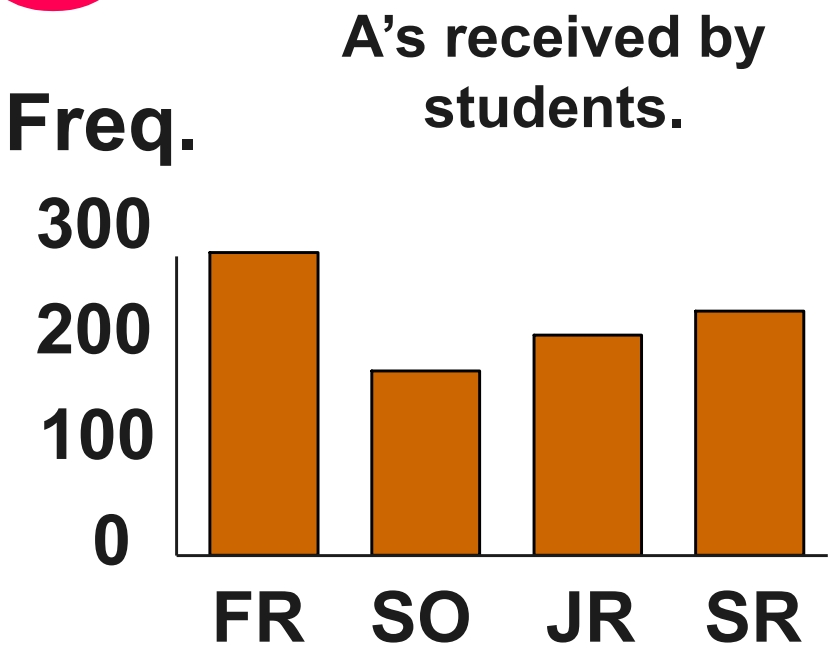
1990: $3.80

✓Good Presentation

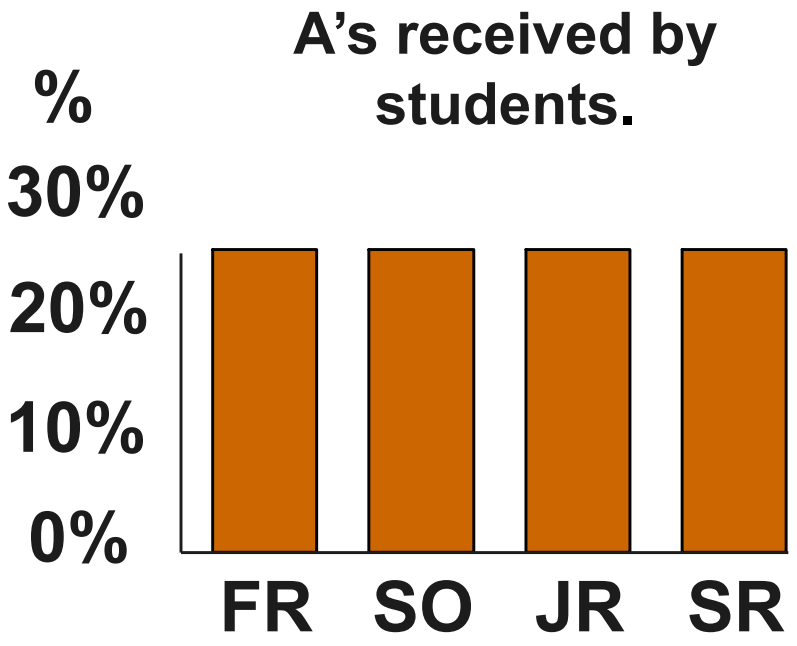


No Relative Basis

Bad Presentation



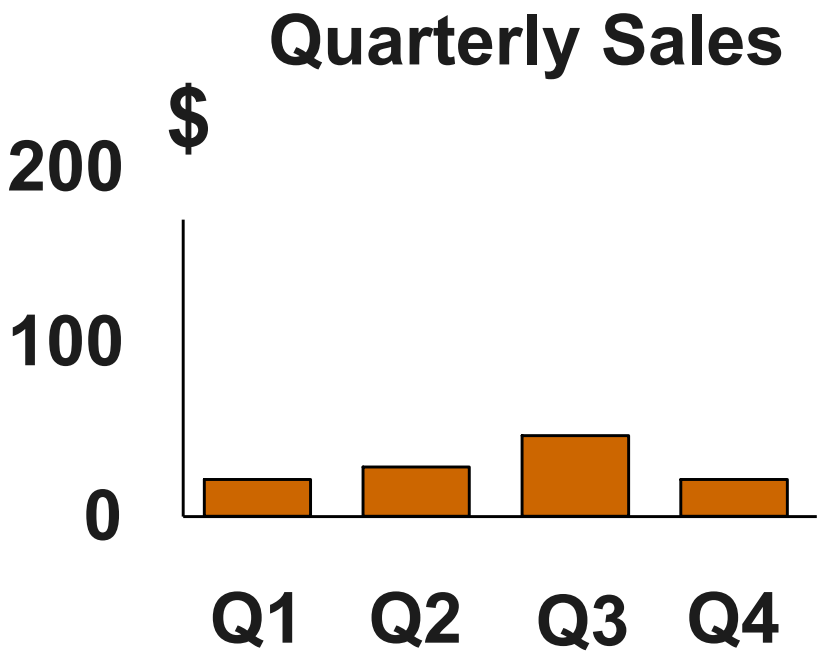
✓ **Good Presentation**



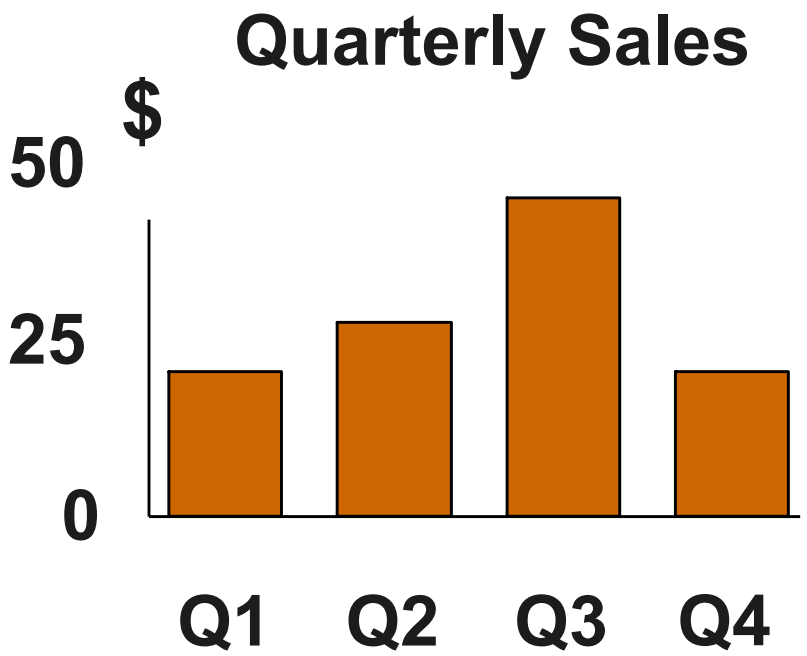
FR = Freshmen, SO = Sophomore, JR = Junior, SR = Senior

Compressing Vertical Axis

Bad Presentation

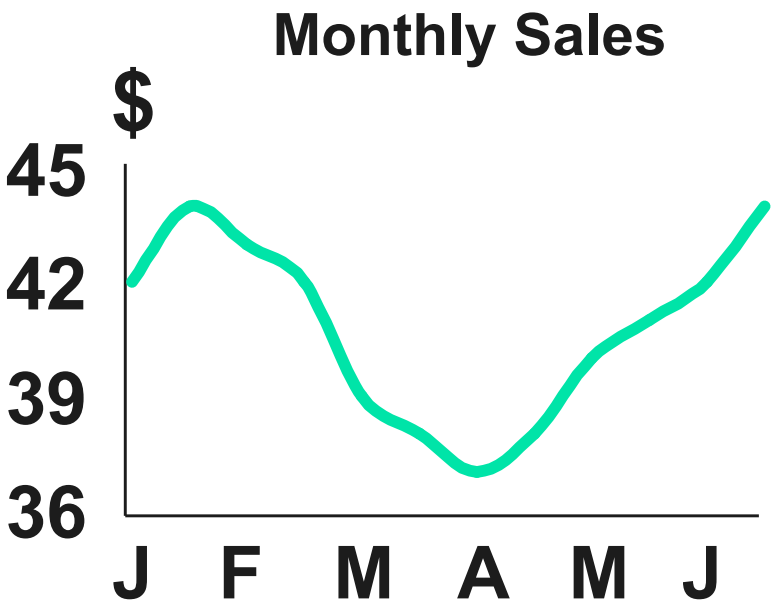


✓Good Presentation

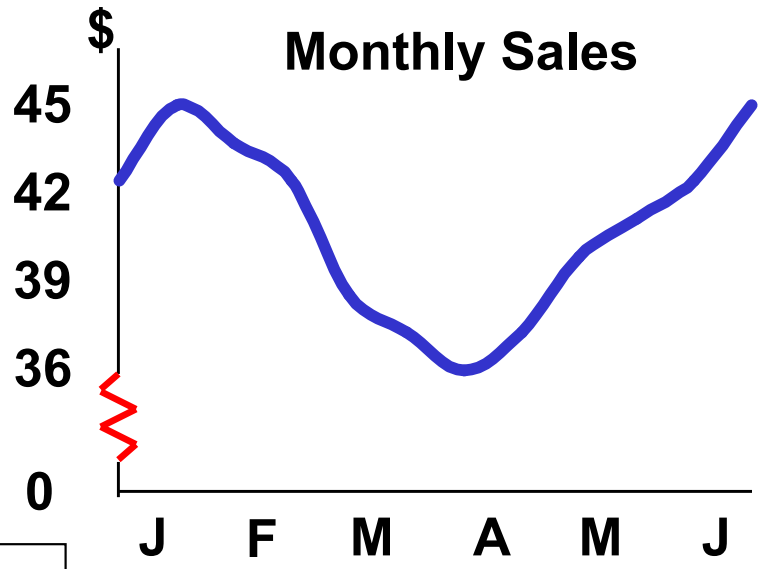


No Zero Point On Vertical Axis

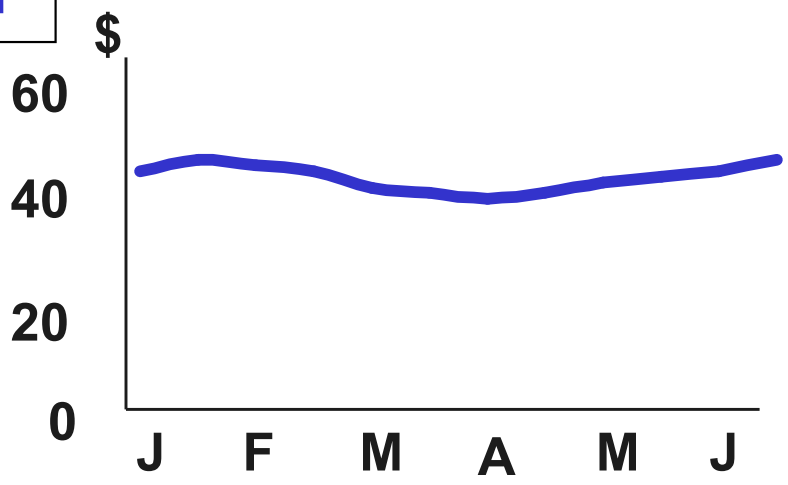
Bad Presentation



✓ Good Presentations



or



Graphing the first six months of sales

Chapter Summary

- Data in raw form are usually not easy to use for decision making -- Some type of organization is needed:
  - Table
  - Graph
- Techniques reviewed in this chapter:
  - Ordered array and stem-and-leaf display
  - Frequency distributions and histograms
  - Polygons and ogives
  - Scatter diagrams for bivariate data
  - Bar charts, pie charts, and Pareto diagrams
  - Contingency tables and side-by-side bar charts