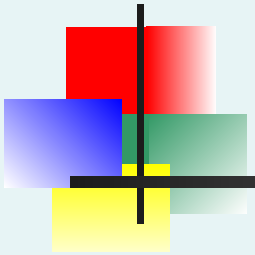
Basic Business Statistics
11th Edition



Chapter 2

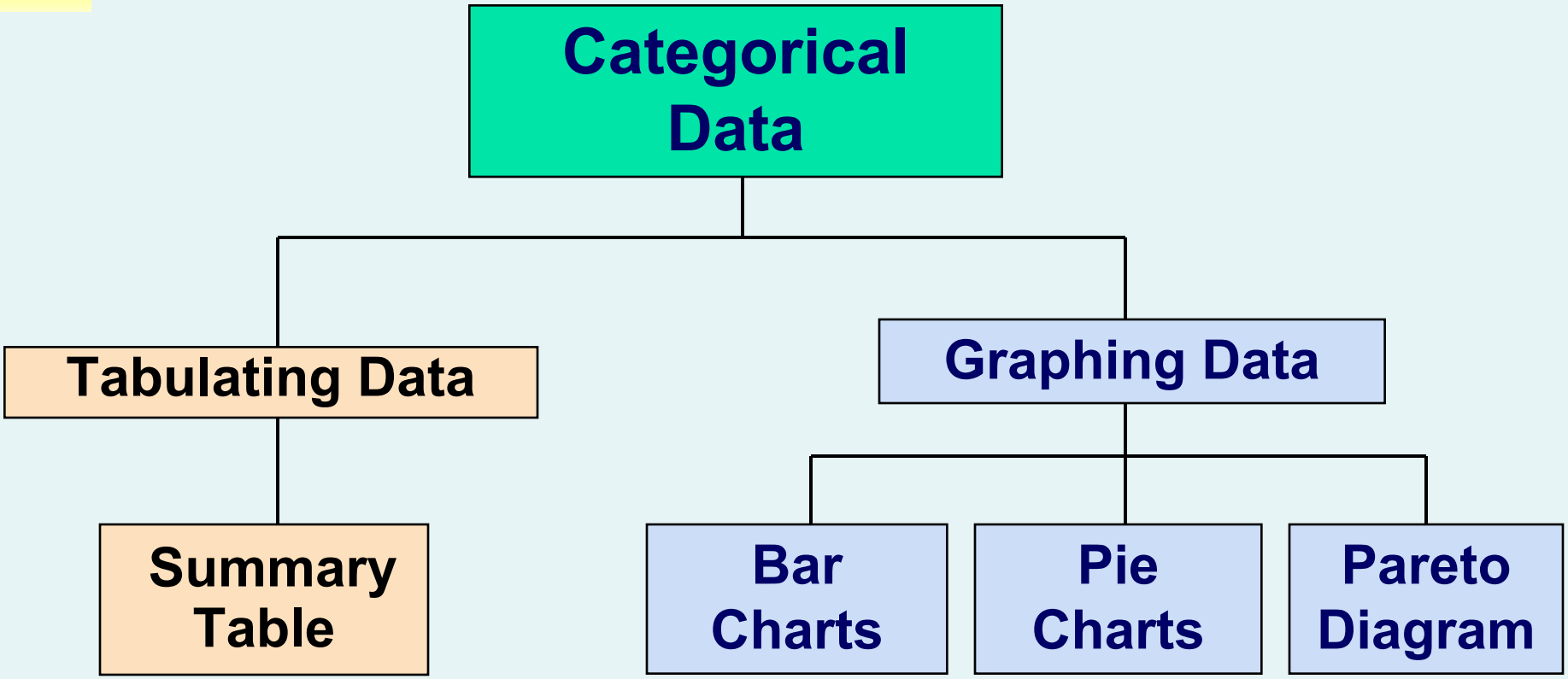
Presenting Data in Tables and Charts

Learning Objectives

In this chapter you learn:

- To develop tables and charts for categorical data
- To develop tables and charts for numerical data
- The principles of properly presenting graphs

Categorical Data Are Summarized By Tables & Graphs



Organizing Categorical Data: Summary Table

- A **summary table** indicates the frequency, amount, or percentage of items in a set of categories so that you can see differences between categories.

Banking Preference?	**Percent**
ATM	16%
Automated or live telephone	2%
Drive-through service at branch	17%
In person at branch	41%
Internet	24%

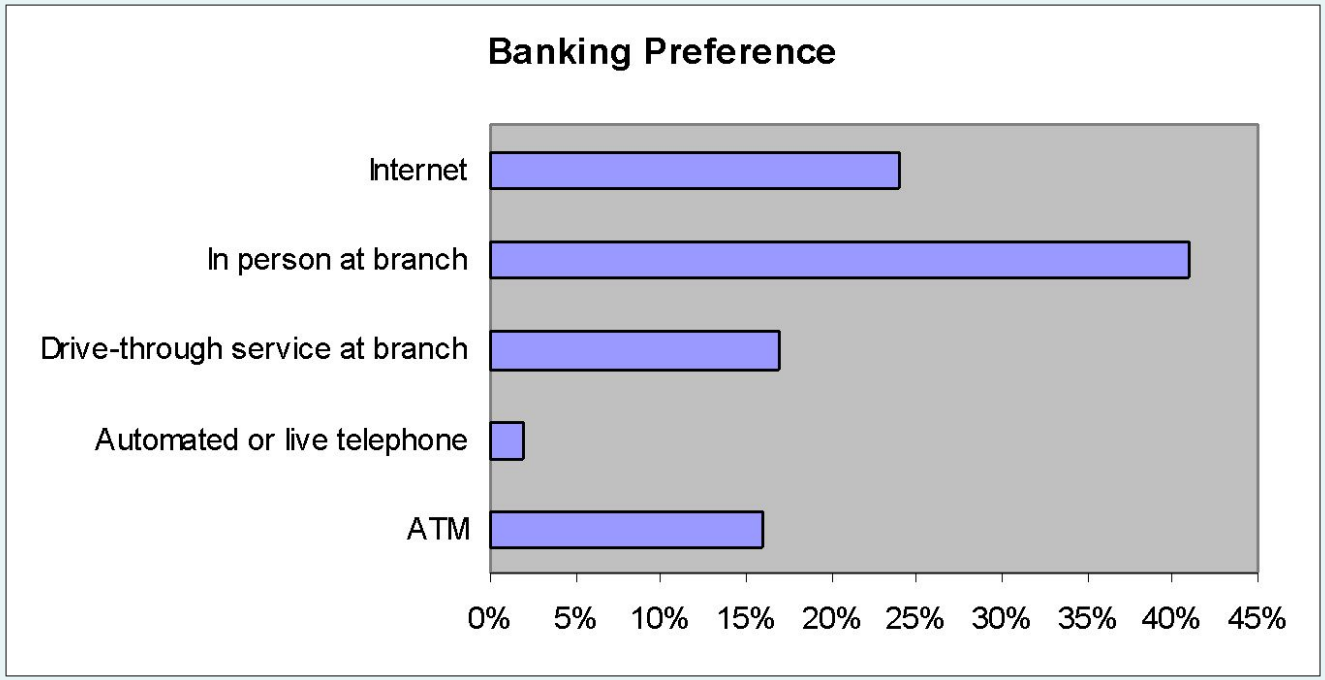
Bar and Pie Charts

- Bar charts and Pie charts are often used for categorical data

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

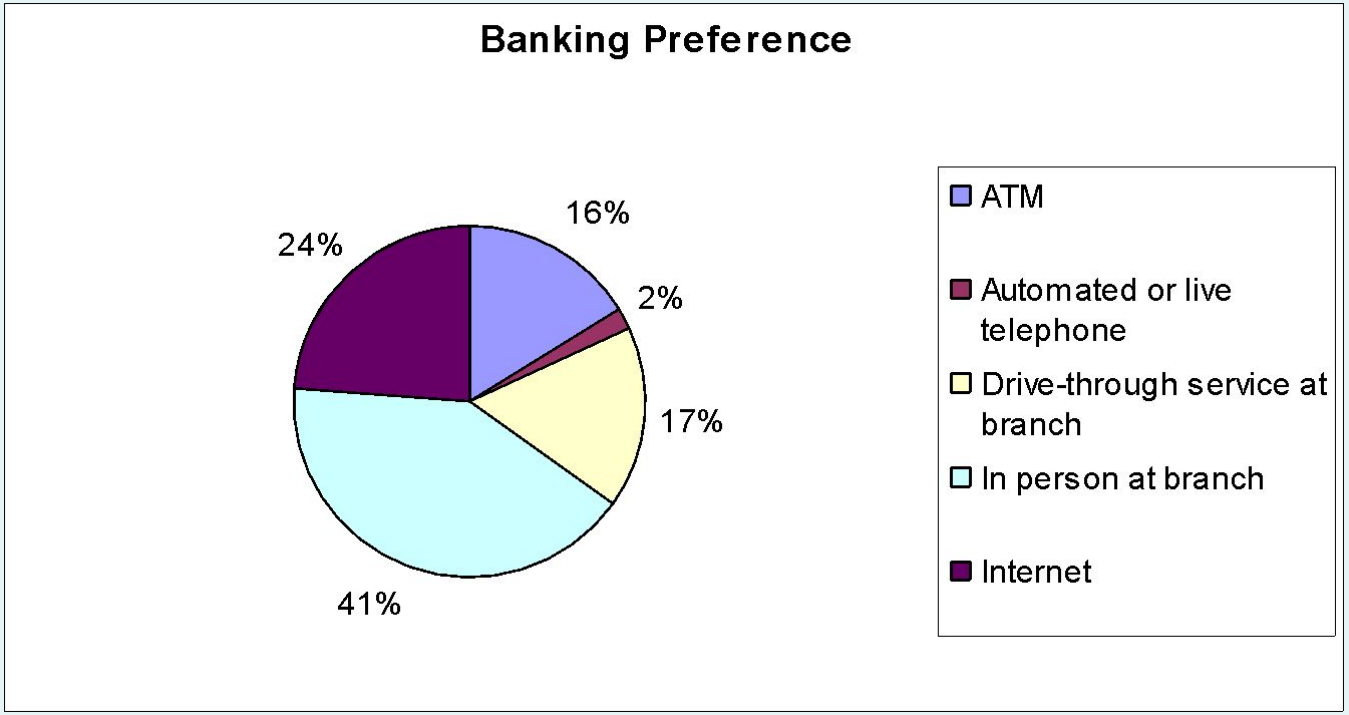
Organizing Categorical Data: Bar Chart

- In a **bar chart,** a bar shows each category, the length of which represents the amount, frequency or percentage of values falling into a category.



Organizing Categorical Data: Pie Chart

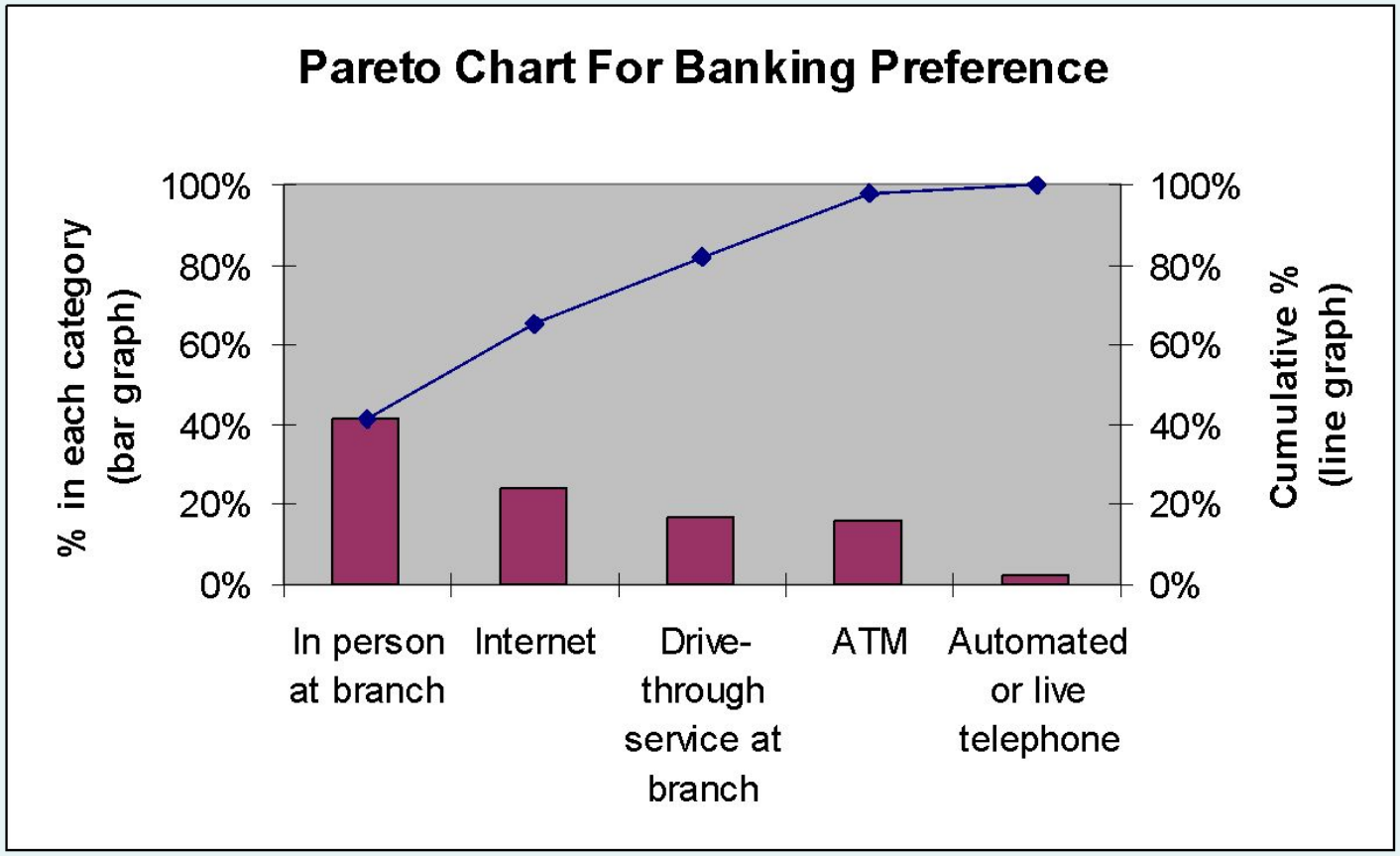
- The **pie chart** is a circle broken up into slices that represent categories. The size of each slice of the pie varies according to the percentage in each category.



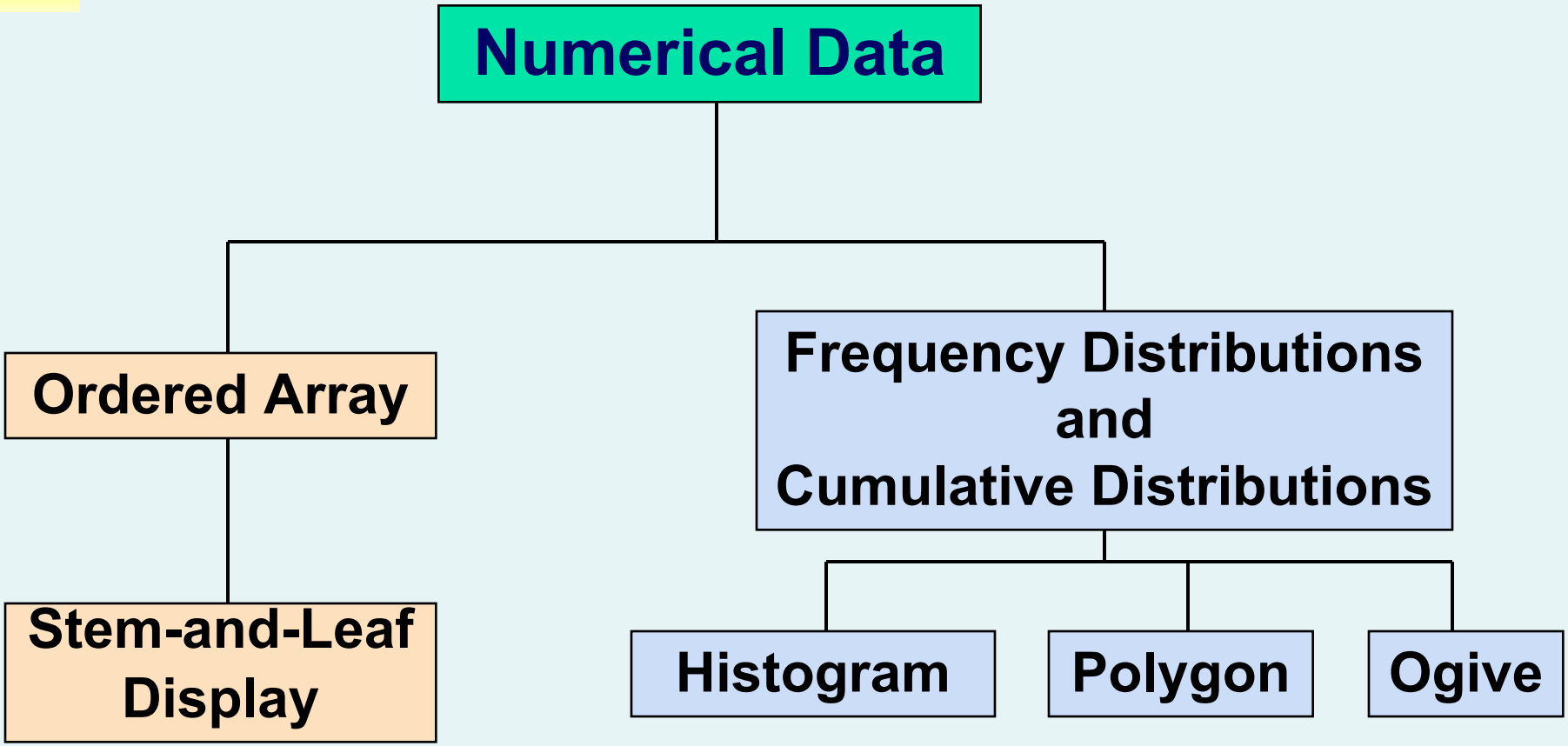
Organizing Categorical Data: Pareto Diagram

- Used to portray categorical data (nominal scale)
- A vertical bar chart, where categories are shown in descending order of frequency
- A cumulative polygon is shown in the same graph
- Used to separate the “vital few” from the “trivial many”

Organizing Categorical Data: Pareto Diagram



Tables and Charts for Numerical Data



Organizing Numerical Data: Ordered Array

- An **ordered array** is a sequence of data, in rank order, from the smallest value to the largest value.
- Shows range (minimum value to maximum value)
- May help identify outliers (unusual observations)

Age of Surveyed College Students	**Day Students**					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Stem-and-Leaf Display

- A simple way to see how the data are distributed and where concentrations of data exist

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

Organizing Numerical Data: Stem and Leaf Display

- A **stem-and-leaf display** organizes data into groups (called stems) so that the values within each group (the leaves) branch out to the right on each row.

Age of Surveyed College Students						
	Day Students					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Age of College Students

Day Students

Stem	Leaf
1	67788899
2	0012257
3	28
4	2

Night Students

Stem	Leaf
1	8899
2	0138
3	23
4	15

Organizing Numerical Data: Frequency Distribution

- The **frequency distribution** is a summary table in which the data are arranged into numerically ordered classes.

- You must give attention to selecting the appropriate *number* of **class groupings** for the table, determining a suitable *width* of a class grouping, and establishing the *boundaries* of each class grouping to avoid overlapping.

- The number of classes depends on the number of values in the data. With a larger number of values, typically there are more classes. In general, a frequency distribution should have at least 5 but no more than 15 classes.

- To determine the **width of a class interval,** you divide the **range** (Highest value–Lowest value) of the data by the number of class groupings desired.

Organizing Numerical Data: 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

Organizing Numerical Data: Frequency Distribution Example

- 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):
  - **Class 1: 10 to less than 20**
  - **Class 2: 20 to less than 30**
  - **Class 3: 30 to less than 40**
  - **Class 4: 40 to less than 50**
  - **Class 5: 50 to less than 60**
- Compute class midpoints: **15, 25, 35, 45, 55**
- Count observations & assign to classes

Organizing Numerical Data: Frequency Distribution Example

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

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	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		

Why Use a Frequency Distribution?

- It condenses the raw data into a more useful form
- It allows for a quick visual interpretation of the data
- It enables the determination of the major characteristics of the data set including where the data are concentrated / clustered

Frequency Distributions: Some Tips

- Different class boundaries may provide different pictures for the same data (especially for smaller data sets)
- Shifts in data concentration may show up when different class boundaries are chosen
- As the size of the data set increases, the impact of alterations in the selection of class boundaries is greatly reduced
- When comparing two or more groups with different sample sizes, you must use either a relative frequency or a percentage distribution

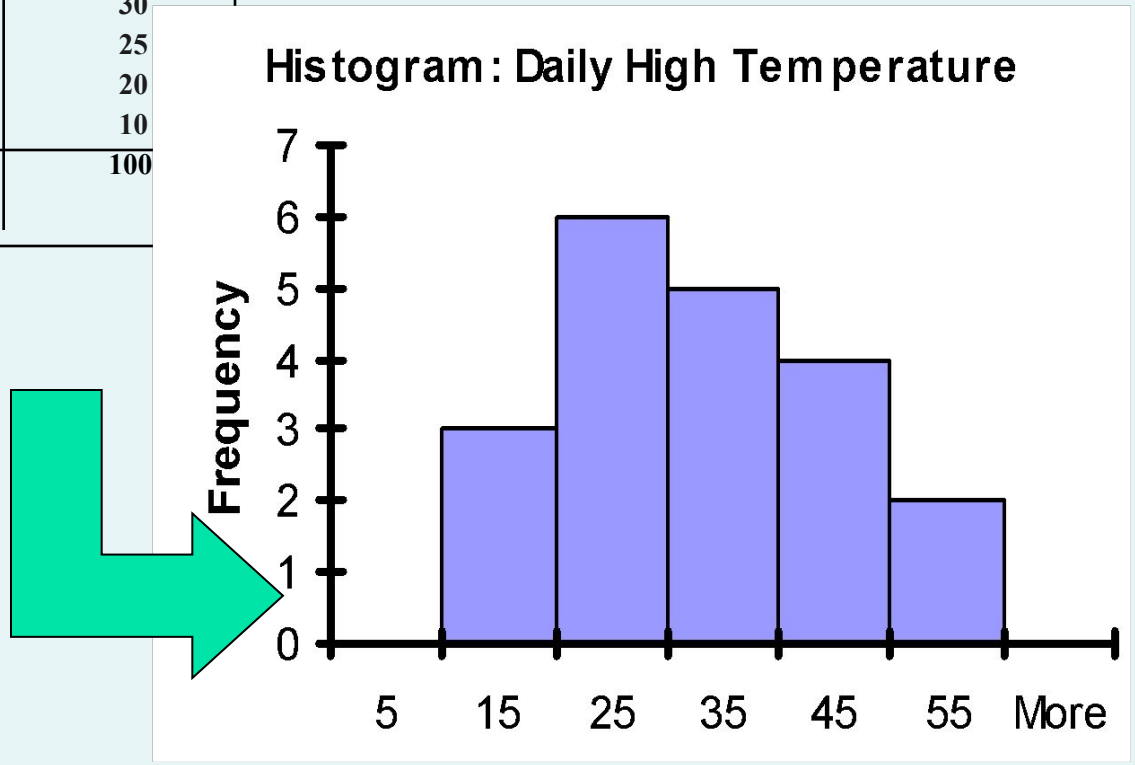
Organizing Numerical Data: The Histogram

- A vertical bar chart of the data in a frequency distribution is called a **histogram.**
- In a histogram there are no gaps between adjacent bars.
- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis.
- The vertical axis is either **frequency, relative frequency,** or **percentage**.
- The height of the bars represent the frequency, relative frequency, or percentage.

Organizing Numerical Data: The Histogram

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

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



Organizing Numerical Data: The Polygon

- A **percentage polygon** is formed by having the midpoint of each class represent the data in that class and then connecting the sequence of midpoints at their respective class percentages.

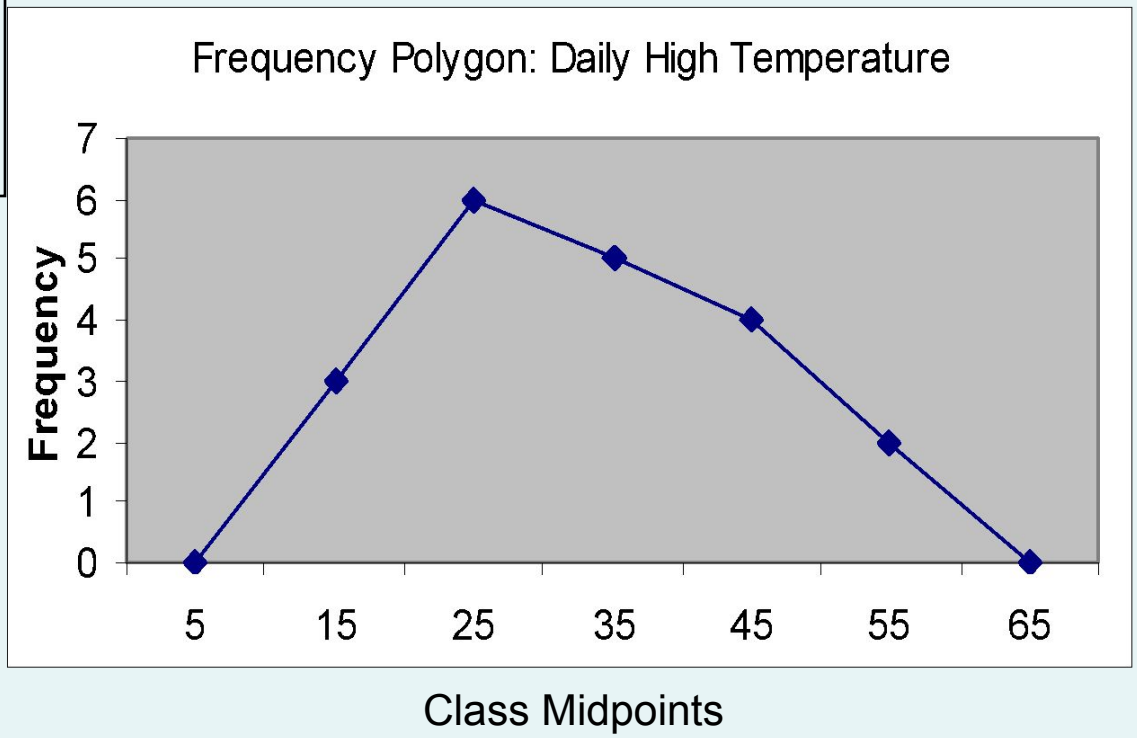
- The **cumulative percentage polygon,** or **ogive,** displays the variable of interest along the *X* axis, and the cumulative percentages along the *Y* axis.

- Useful when there are two or more groups to compare.

Graphing Numerical Data: The Frequency Polygon

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

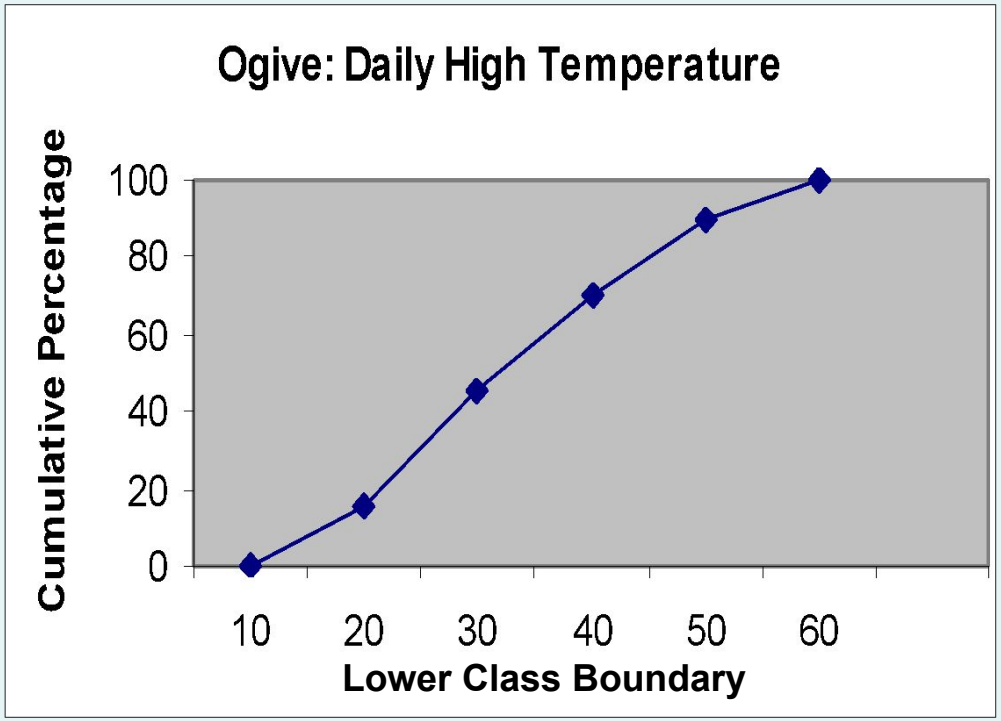
(In a percentage polygon the vertical axis would be defined to show the percentage of observations per class)



Graphing Cumulative Frequencies: The Ogive (Cumulative % Polygon)

Class	Lower class boundary	% less than lower boundary
10 but less than 20	10	15
20 but less than 30	20	45
30 but less than 40	30	70
40 but less than 50	40	90
50 but less than 60	50	100

(In an ogive the percentage of the observations less than each lower class boundary are plotted versus the lower class boundaries.



Cross Tabulations

- Used to study patterns that may exist between two or more categorical variables.

- Cross tabulations can be presented in:
  - Tabular form -- Contingency Tables
  - Graphical form -- Side by Side Charts

Cross Tabulations: The Contingency Table

- A **cross-classification** (or **contingency) table** presents the results of two categorical variables. The joint responses are classified so that the categories of one variable are located in the rows and the categories of the other variable are located in the columns.

- The cell is the intersection of the row and column and the value in the cell represents the data corresponding to that specific pairing of row and column categories.

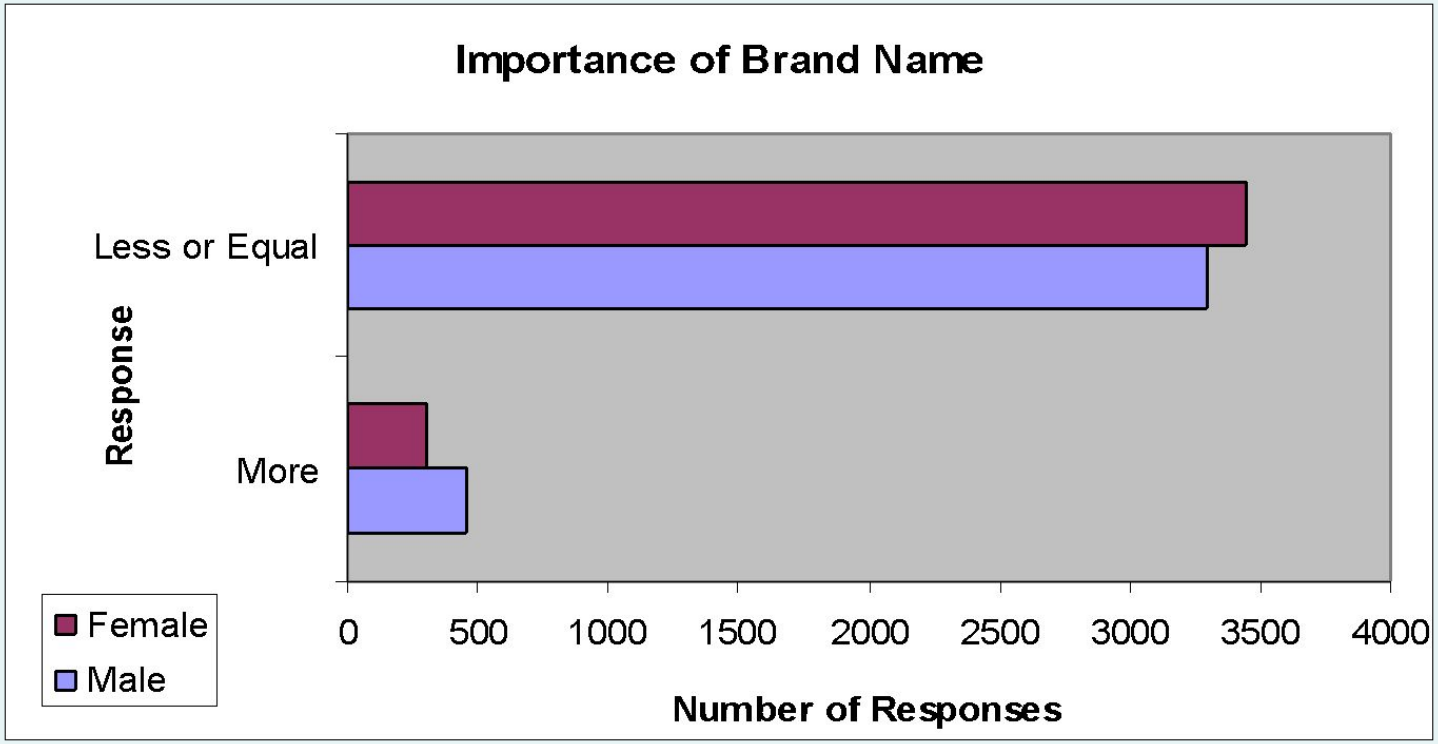
- A useful way to visually display the results of cross-classification data is by constructing a **side-by-side bar chart.**

Cross Tabulations: The Contingency Table

A survey was conducted to study the importance of brand name to consumers as compared to a few years ago. The results, classified by gender, were as follows:

Importance of Brand Name	Male	Female	Total
More	450	300	750
Equal or Less	3300	3450	6750
Total	**3750**	**3750**	**7500**

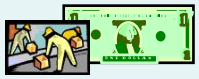
Cross Tabulations: Side-By-Side Bar Charts

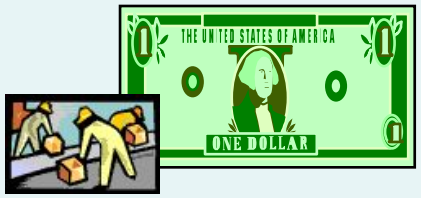


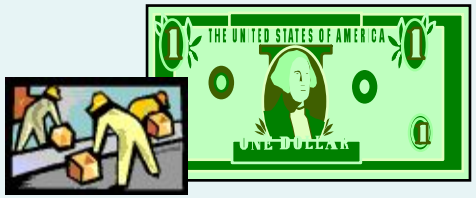
Graphical Errors: Chart Junk

Bad Presentation

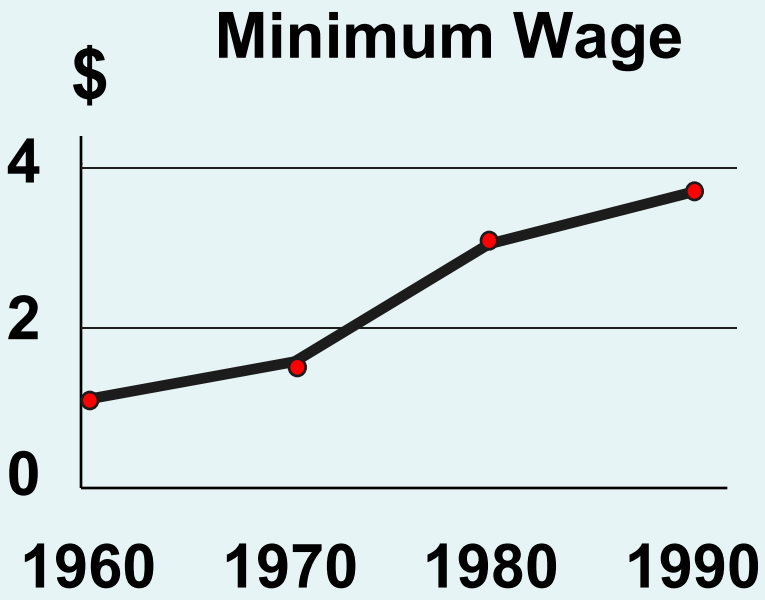
Minimum Wage





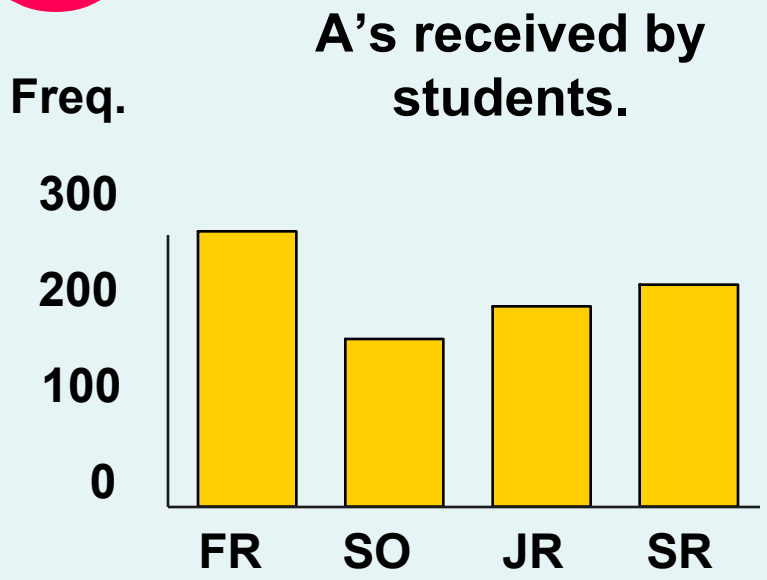


Good Presentation

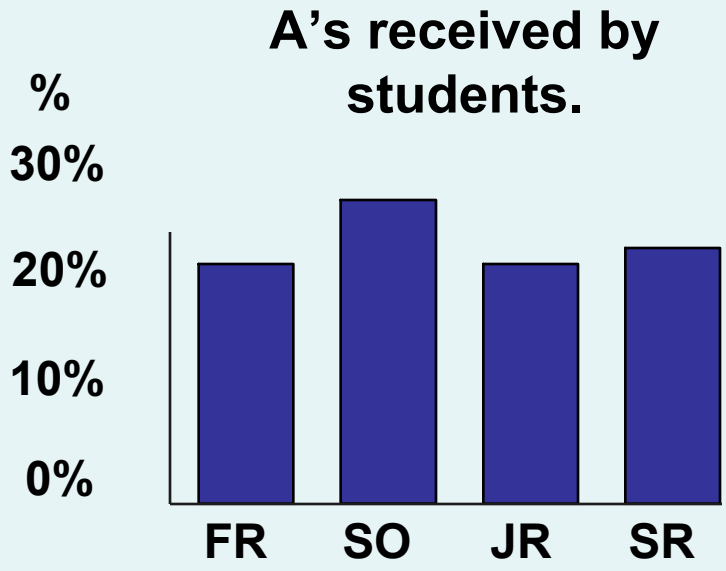


Graphical Errors: No Relative Basis

Bad Presentation



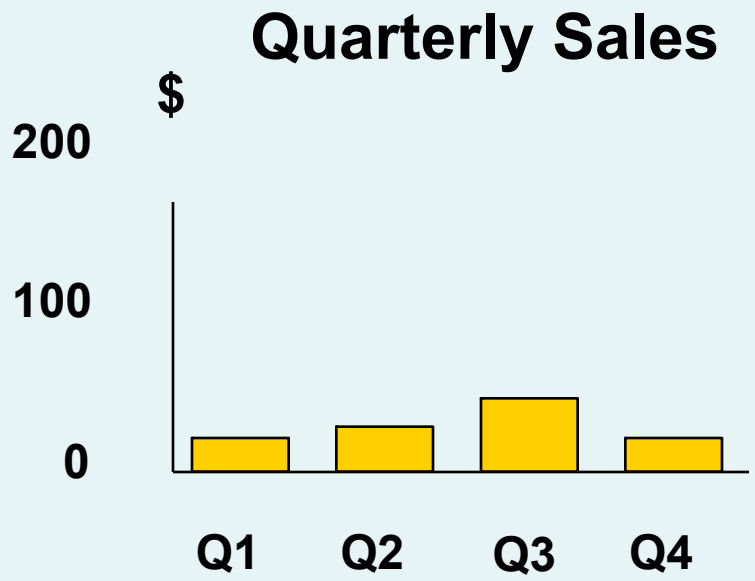
Good Presentation



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

Graphical Errors: Compressing the Vertical Axis

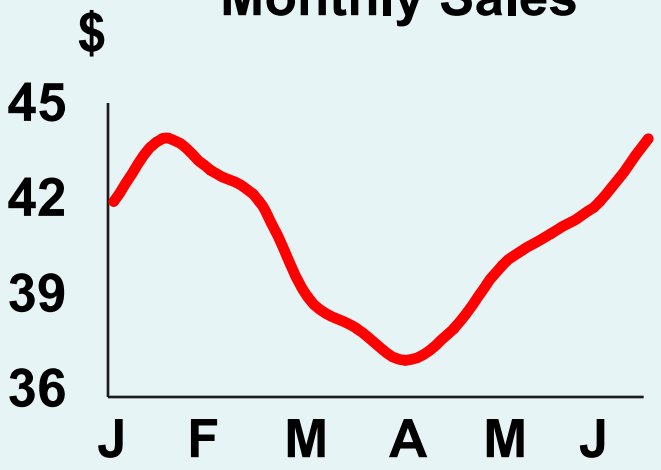
Bad Presentation



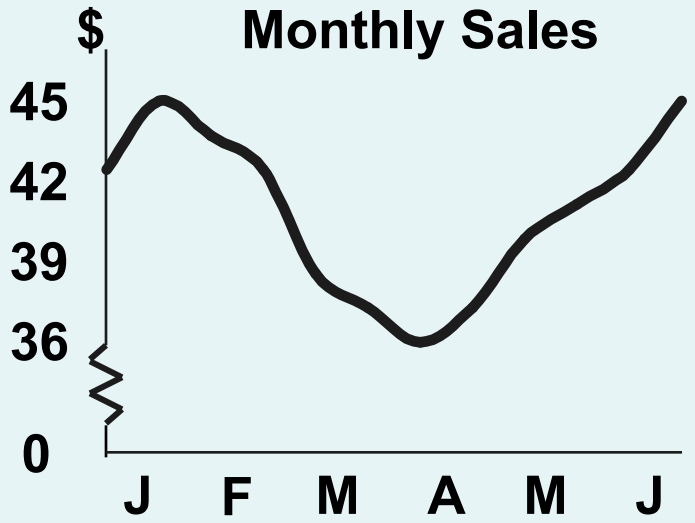
Good Presentation

Graphical Errors: No Zero Point on the Vertical Axis

Bad Presentation



Good Presentations



Graphing the first six months of sales

Chapter Summary

In this chapter, we have

- Organized categorical data using the summary table, bar chart, pie chart, and Pareto diagram.
- Organized numerical data using the ordered array, stem-and-leaf display, frequency distribution, histogram, polygon, and ogive.
- Examined cross tabulated data using the contingency table and side-by-side bar chart.
- Developed scatter plots and time series graphs.
- Examined the do's and don'ts of graphically displaying data.