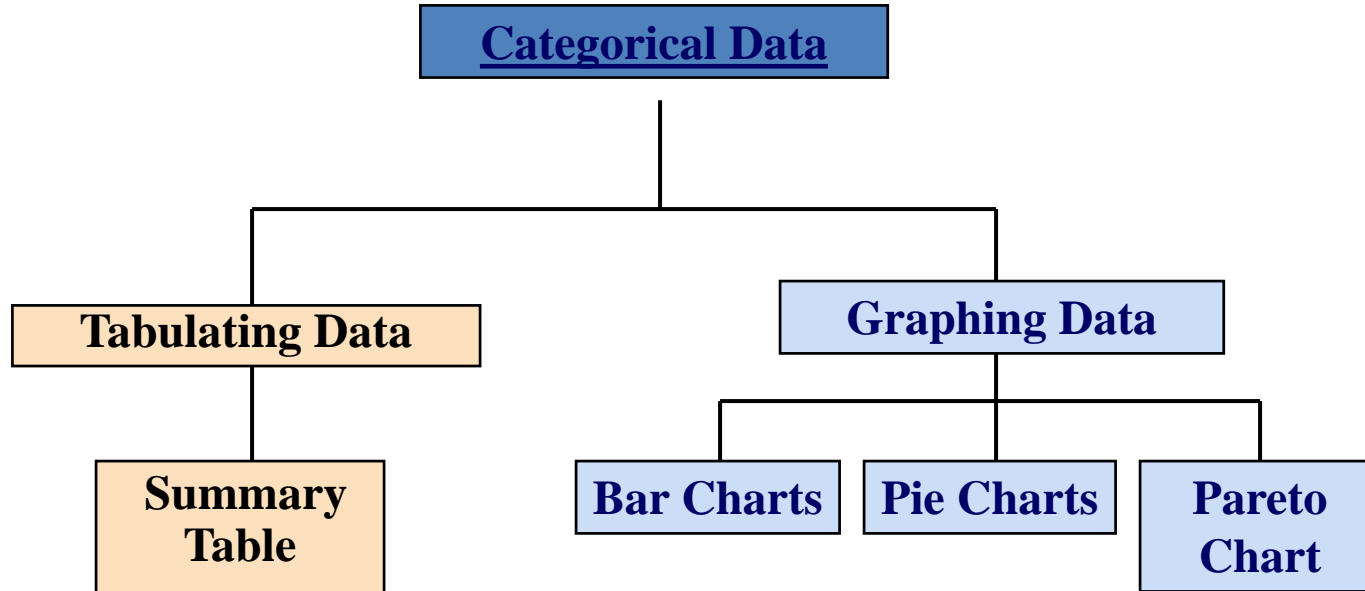# Presenting Data in Tables and Charts

Categorical Data Are Summarized By Tables & Graphs

Categorical Data

- **Tabulating Data**
  - **Summary Table**
- **Graphing Data**
  - **Bar Charts**
  - **Pie Charts**
  - **Pareto Chart**

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.
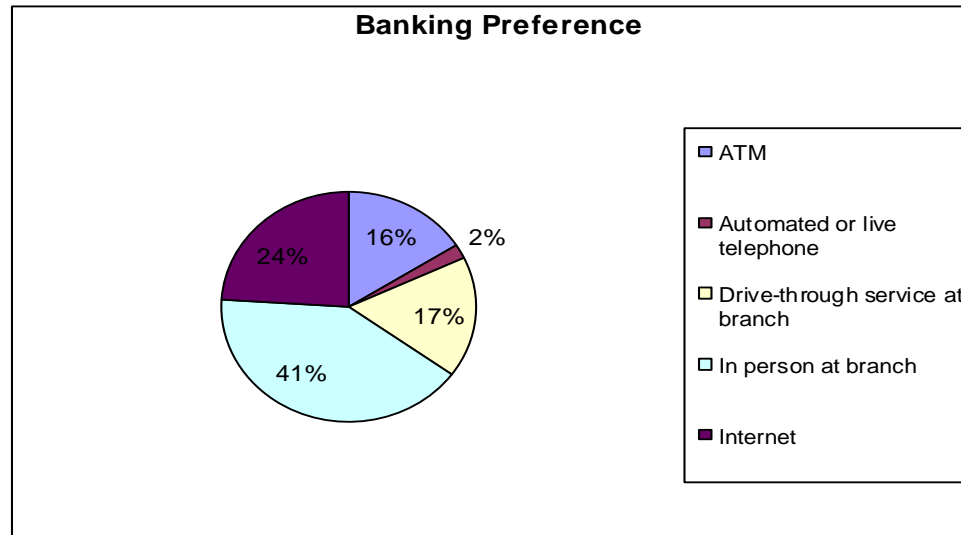
Organizing Categorical Data: Bar Chart

▪ 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.

Banking Preference
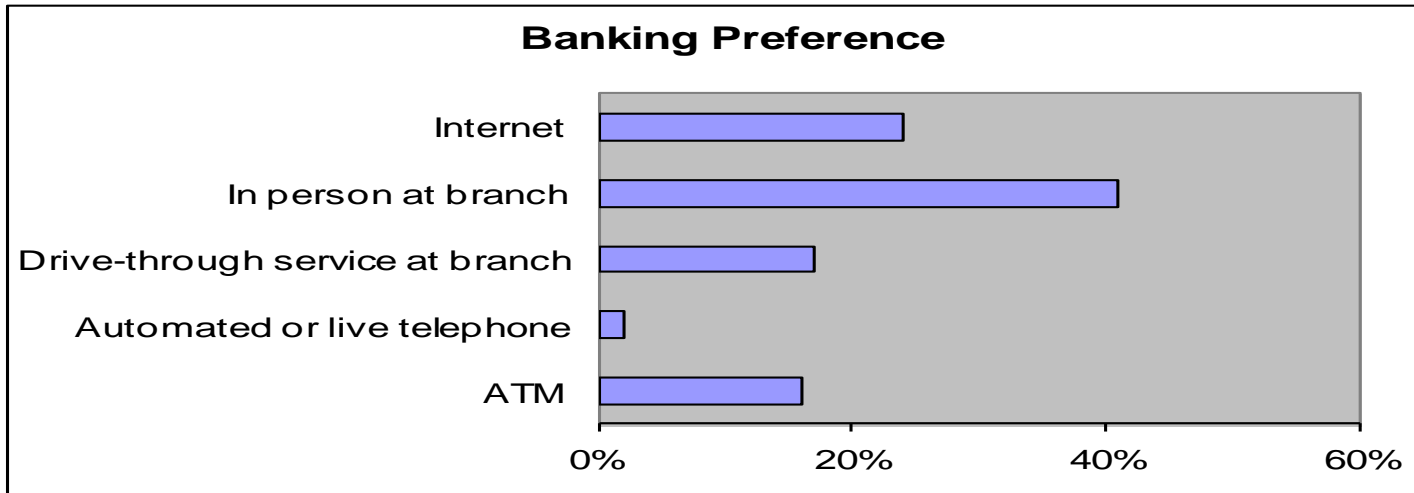
Organizing Categorical Data: Pie Chart

- 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.

Banking Preference

Legend:
- ATM
- Automated or live telephone
- Drive-through service at branch
- In person at branch
- Internet

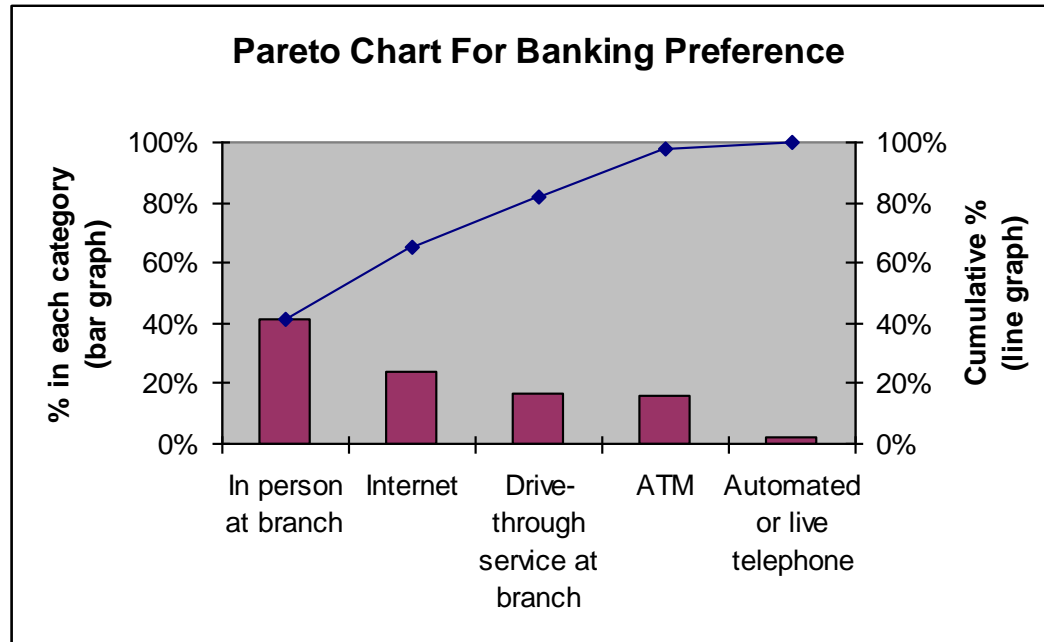Pie chart values: 16%, 2%, 17%, 41%, 24%

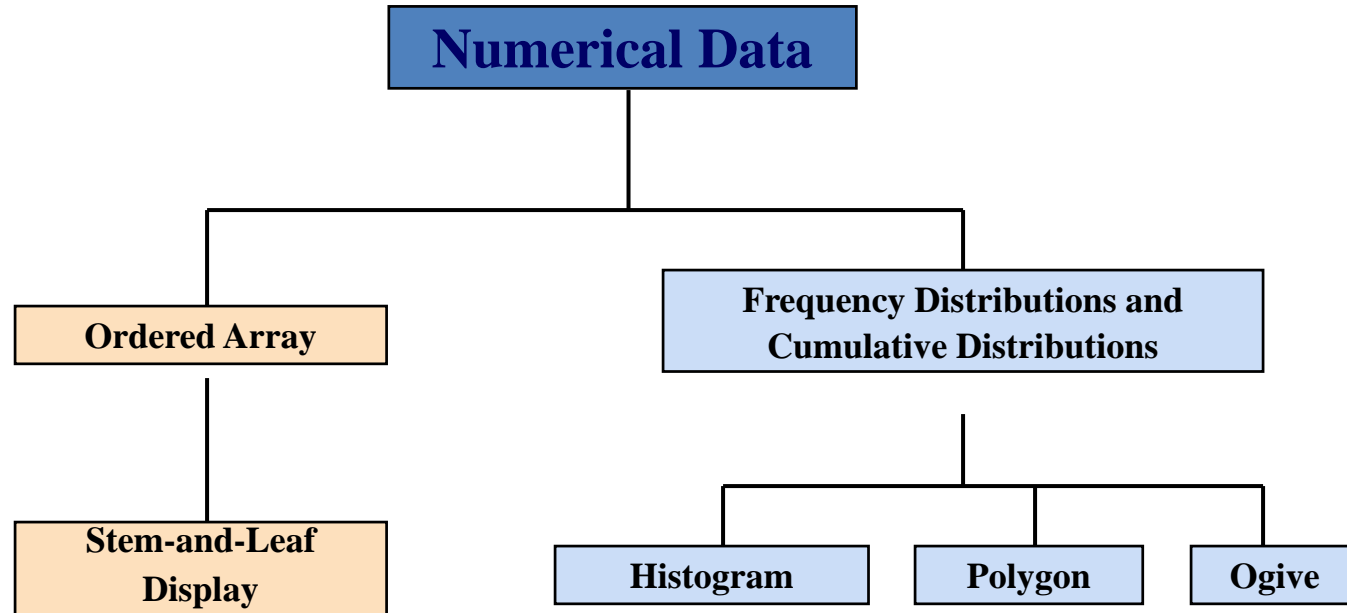# Organizing Categorical Data: Pareto Chart

- Used to portray categorical data (nominal scale)

- **A vertical bar chart, where categories are shown in <u>descending order of frequency</u>**

- A **cumulative polygon** is shown in the same graph

- **Used to separate the "vital few" from the "trivial many"**

Organizing Categorical Data: Pareto Chart

Tables and Charts for <u>Numerical</u> 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)
- Which values appear **more than one**
- Divide data in **sections** (Day students- 1/3rd of data below 18, 2/3rd below 22,etc)

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

Girls		Boys
7, 8, 2, 2, 1	1	5, 8
3, 3, 3, 2	2	2, 2, 3, 6
5, 4, 3	3	4, 5, 5, 5
7, 5, 4	4	0, 0, 2, 7, 9
1, 1, 0	5	0, 0, 1

Stems	Leaves			
10	4	7		
11	2	5	5	6
12	3			
13	0	4		
14	5	7		

Means 145

Stem and Leaf plot for decimal numbers

Stem							
8.	0	0					
9.	0						
10.	0	0					
11.	0	0	5				
12.	0	0	0	2			
13.	2	5	8	8			
14.	0	0	0	0	4	6	8
15.	0	0	5				
16.	0	2	6	8			
17.	0	0	5				
18.	0	2	5				
19.	0	5					
20.	0	5					

Organizing Numerical Data: Frequency Distribution

▪ The **frequency distribution** is a summary table in which **the data are arranged into numerically ordered <u>classes.</u>**

▪ 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 <u>less than</u> 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**

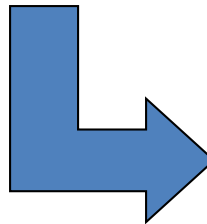# Organizing Numerical Data: The Histogram

- 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 <u>percentage histogram</u> the vertical axis would be defined to show the percentage of observations per class)

Histogram: Daily High Temperature
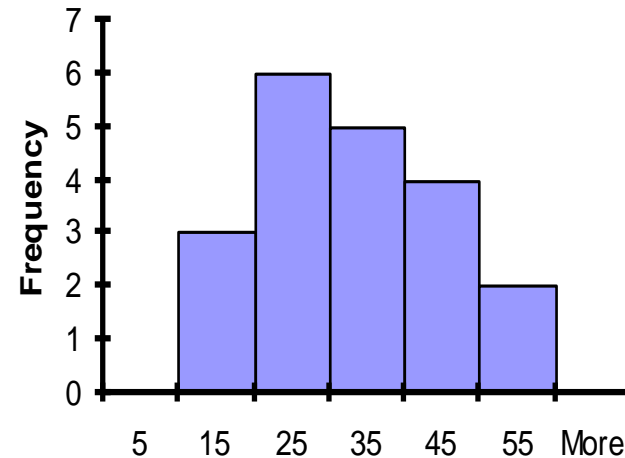
Organizing Numerical Data: The Polygon

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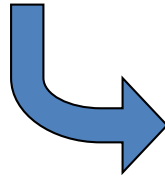
(In a percentage polygon the **vertical axis** would be defined to show the **percentage of observations per class**)

Frequency Polygon: Daily High Temperature