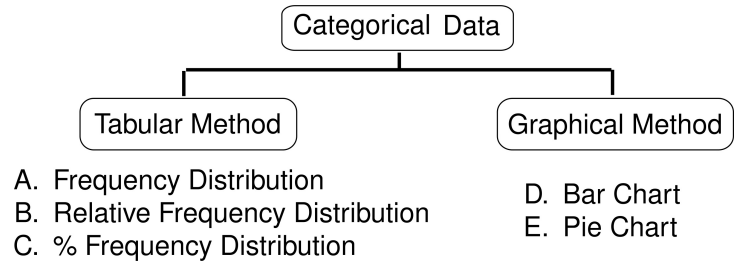


CH 2: Descriptive Statistics: Tabular and Graphical Presentations

Part 1 Organizing Categorical Data

1. Case 1: One Variable Categorical Data (Sec 2.1)



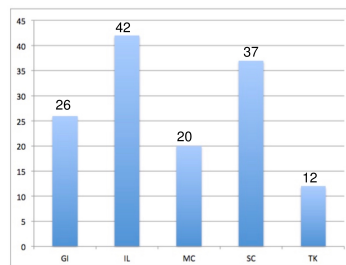
- (A) **Frequency Distribution:** A tabular summary of data showing the number (frequency) of data values in each of several nonoverlapping classes.
- (B) **Relative Frequency Distribution:** A tabular summary of data showing the fraction or proportion of data values in each of several nonoverlapping classes.
 Note: Relative frequency of a class = $\frac{\text{Frequency (in the interval)}}{\text{Total number of observations}}$
- (C) **Percent Frequency Distribution:** A tabular summary of data showing the percentage of data values in each of several nonoverlapping classes.
 Note: Percentage frequency of a class = relative frequency $\times 100\%$
- (D) **Bar chart:** A graphical device for displaying categorical data that have been summarized in a frequency, relative frequency, or percent frequency distribution.
- (E) **Pie chart:** Use pie slices to display the percent frequency distribution of each category.

EX1 Given a summary table of 137 mutual funds:

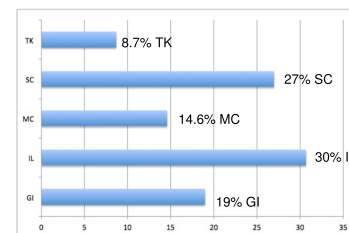
Fund Objective	Number of Funds	Rel. Frequency	% Frequency
Growth and Income (GI)	26		
International (IL)	42		
Midcap (MC)	20		
SmallCap (SC)	37		
Technology(TK)	12		

(a) Provide the relative frequency and the percent frequency distributions.

(b) Construct a bar chart

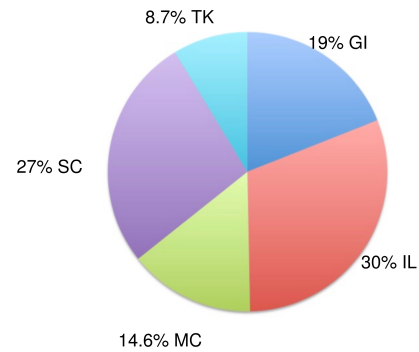


Frequency Bar Chart



Percent Frequency Bar Chart

(c) Construct a pie chart



(d) Based on the pie chart, what percentage of the fund is from $MC + TK$?

2. Case 2: Bivariate Categorical Data (Sec 2.4 Crosstabulations)

Crosstabulation: A tabular summary of data for two variables. The classes for one variable are represented by the rows; the classes for the other variable are presented by the columns.

EX2 Crosstabulation (contingency table) of whether the fund has a sales charge vs. mutual funds

Sales Charge	GI	IL	MC	SC	TK	Total
Y	17	25	6	15	9	72
N	9	17	14	22	3	65
Total	26	42	20	37	12	137

3. **Side-by-side bar chart:** Bar charts arranged side-by-side according to different categories. Useful when looking for patterns or relationships.

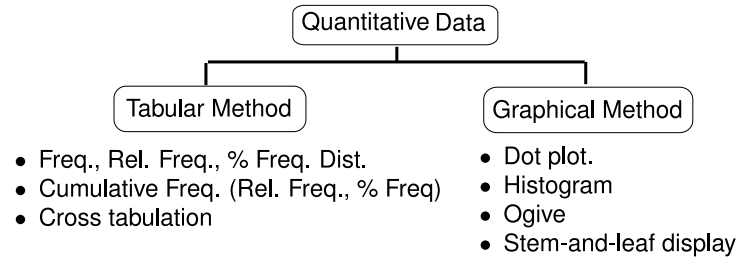
EX 2 (cont.) Construct the side-by-side bar chart.

4. Note: If we divided each cell of Table 1 by the grand total, we obtain a % based overall table.

Sales Charge	GI	IL	MC	SC	TK	Total
Y	17/137=12.4%	18.3%	4.4%	10.9%	6.6%	52.6%
N	9/137=6.6%	12.4%	10.2%	16.1%	2.2%	47.4%
Total	26/137=19%	30.7%	14.6%	27%	8.8%	100%

CH 2: Descriptive Statistics: Tabular and Graphical Presentations

Part 2 Organizing Quantitative Data



1. Dot Plot: A graphical device that summarizes data by the number of dots above each data value on the horizontal axis.

EX3 Given a set of data: 3, 5, 8, 9, 10, 1. Construct a dot plot.

2. Histogram:

(A) **Histogram:** A graphical presentation of a frequency distribution, relative frequency distribution, or percent frequency distribution of quantitative data constructed by placing the class intervals on the horizontal axis and the frequencies, relative frequencies, or percent frequencies on the vertical axis.

(B) Width of the classes:

$$\text{Approximate class width} = \frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}$$

EX3 (Cont.) Construct the distributions (with 2 classes).

EX4 The data set of the monthly finance charges of 50 customers from a store's record is given: \$20, \$5, \$13, The following table is the frequency distribution of the data set.

Class interval	number of customers (Freq)	Relative Frequency
\$0-5	15	
\$ 5-10	20	
\$10-15	10	
\$15-20	5	

a. Construct a histogram (with Relative Frequency)

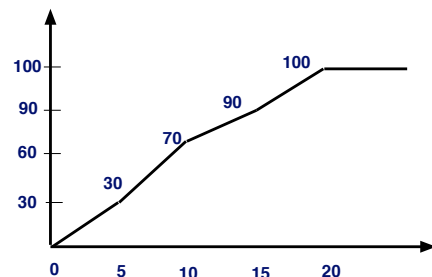
- b. Referring to the histogram, what % of the finance charges is between \$5 to \$15?
- c. Referring to the histogram, what % of the finance charges is below \$10?
- d. Referring to the histogram, 70% of the finance charges is above what amount?

3. Cumulative distribution and Ogive

- (A) **Cumulative frequency distribution:** A tabular summary of quantitative data showing the number of data values that are less than or equal to the upper class limit of each class.
- (B) **Cumulative relative frequency distribution:** A tabular summary of quantitative data showing the fraction or proportion of data values that are less than or equal to the upper class limit of each class.
- (C) **Cumulative percent frequency distribution:** A tabular summary of quantitative data showing the percentage of data values that are less than or equal to the upper class limit of each class.
- (D) **Ogive:** A graph of a cumulative distribution.

EX4 (cont.) Construct a Ogive (cumulative % distribution).

\$	Cumulative %
0	0 %
5	30 %
10	70 %
15	90 %
20	100 %



4. Stem and Leaf Display:

- (A) **Ordered array:** Sorting the observations in rank order from the smallest to the largest.
- (B) **Stem and Leaf:** An exploratory data analysis technique that simultaneously rank orders quantitative data and provides insight about the shape of the distribution. In this case, data are separated into leading digits (stems) and the remaining digits (leaves).

EX5 Given a set of data: 269, 272, 305, 283, 438, and 343. Construct a stem-and-leaf display.

EX6 Given a stem-and-leaf display of the exam scores (0-100). What % of the exam scores is higher than 80?

Stem	Leaf
5	5 7
6	2 3
7	1 5 5
8	4
9	0 8