Probability and Statistics

Set # 1 (From Neil Weiss book)

Organizing Data

Variables and Data

Variables

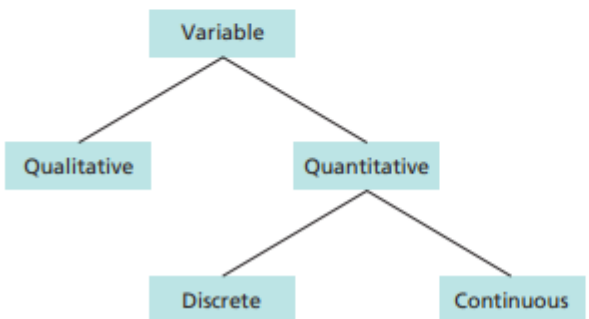
Variable: A characteristic that varies from one person or thing to another.

Qualitative variable: A nonnumerically valued variable.

Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed.

Continuous variable: A quantitative variable whose possible values form some interval of numbers.



Data

Data: Values of a variable.

Qualitative data: Values of a qualitative variable.

Quantitative data: Values of a quantitative variable.

Discrete data: Values of a discrete variable.

Continuous data: Values of a continuous variable.

Organizing Qualitative Data

Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

To Construct a Frequency Distribution of Qualitative Data

Step 1 **List the distinct values of the observations in the data set in the first column of a table.**

Step 2 **For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.**

Step 3 **Count the tallies for each distinct value and record the totals in the third column of the table.**

Relative-Frequency Distributions

In addition to the frequency that a particular distinct value occurs, we are often interested in the **relative frequency,** which is the ratio of the frequency to the total number of observations:

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Number of observations}}.$$

To Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 **Obtain a frequency distribution of the data.**

Step 2 **Divide each frequency by the total number of observations.**

*For each data set in Exercises **2.18–2.23**,*
a. determine a frequency distribution.
b. obtain a relative-frequency distribution.
c. draw a pie chart.
d. construct a bar chart.

2.18 Top Broadcast Shows. The networks for the top 20 television shows, as determined by the *Nielsen Ratings* for the week ending October 26, 2008, are shown in the following table.

CBS	ABC	CBS	ABC	ABC
Fox	CBS	CBS	Fox	CBS
ABC	CBS	CBS	CBS	Fox
Fox	Fox	CBS	Fox	ABC

2.23 Road Rage. The report *Controlling Road Rage: A Literature Review and Pilot Study* was prepared for the AAA Foundation for Traffic Safety by D. Rathbone and J. Huckabee. The authors discuss the results of a literature review and pilot study on how to prevent aggressive driving and road rage. As described in the study, *road rage* is criminal behavior by motorists characterized by uncontrolled anger that results in violence or threatened violence on the road. One of the goals of the study was to determine when road rage occurs most often. The days on which 69 road rage incidents occurred are presented in the following table.

F	F	Tu	Tu	F	Su	F	F	Tu	F
Tu	Sa	Sa	F	Sa	Tu	W	W	Th	Th
Th	Sa	M	Tu	Th	Su	W	Th	W	Tu
Tu	F	Th	Th	F	W	F	Th	F	Sa
F	W	W	F	Tu	W	W	Th	M	M
F	Su	Tu	F	W	Su	W	Th	M	Tu
F	W	Th	M	Su	Sa	Sa	F	F	

Organizing Quantitative Data

Single-Value Grouping

In some cases, the most appropriate way to group quantitative data is to use classes in which each class represents a single possible value. Such classes are called **single-value classes,** and this method of grouping quantitative data is called **single-value grouping.**

Thus, in single-value grouping, we use the distinct values of the observations as the classes, a method completely analogous to that used for qualitative data. Single-value grouping is particularly suitable for discrete data in which there are only a small number of distinct values.

EXAMPLE 2.12 Single-Value Grouping

TVs per Household The Television Bureau of Advertising publishes information on television ownership in *Trends in Television*. Table 2.4 gives the number of TV sets per household for 50 randomly selected households. Use single-value grouping to organize these data into frequency and relative-frequency distributions.

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

Solution

Number of TVs	Frequency	Relative frequency
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	50	1.00

Organizing Quantitative Data

Limit Grouping

A second way to group quantitative data is to use **class limits.** With this method, each class consists of a range of values. The smallest value that could go in a class is called the **lower limit** of the class, and the largest value that could go in the class is called the **upper limit** of the class.

This method of grouping quantitative data is called **limit grouping.** It is particularly useful when the data are expressed as whole numbers and there are too many distinct values to employ single-value grouping.

EXAMPLE 2.13 Limit Grouping

Days to Maturity for Short-Term Investments Table 2.6 displays the number of days to maturity for 40 short-term investments. The data are from *BARRON'S* magazine. Use limit grouping, with grouping by 10s, to organize these data into frequency and relative-frequency distributions.

TABLE 2.6
Days to maturity for 40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Solution

Days to maturity	Tally	Frequency	Relative frequency
30–39	III	3	0.075
40–49	I	1	0.025
50–59	𝍸 III	8	0.200
60–69	𝍸 𝍸	10	0.250
70–79	𝍸 II	7	0.175
80–89	𝍸 II	7	0.175
90–99	IIII	4	0.100
		40	1.000

Organizing Quantitative Data

Terms Used in Limit Grouping

Lower class limit: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class width: The difference between the lower limit of a class and the lower limit of the next-higher class.

Class mark: The average of the two class limits of a class.

Organizing Quantitative Data

Example (limit grouping)

Data:
6 8 12 13 15 16 8 21 23 32 42 52 56 68 23 11 57 58 58 55 65 47 44 43 41 37 38 39 40 41
Construct frequency distribution of above data using 6 classes

Solution

$$width = \frac{range(R)}{\#\ of\ classes(m)} = \frac{68-6}{6} = \frac{62}{6} = 10.333$$

Class Interval (C-I)	Frequency (f)
6-16	8
17-27	3
28-38	3
39-49	8
50-60	6
61-70	2

Organizing Quantitative Data

Cutpoint Grouping

A third way to group quantitative data is to use **class cutpoints.** As with limit grouping, each class consists of a range of values. The smallest value that could go in a class is called the **lower cutpoint** of the class, and the smallest value that could go in the next-higher class is called the **upper cutpoint** of the class. Note that the lower cutpoint of a class is the same as its lower limit and that the upper cutpoint of a class is the same as the lower limit of the next higher class.

The method of grouping quantitative data by using cutpoints is called **cutpoint grouping**. This method is particularly useful when the data are continuous and are expressed with decimals.

EXAMPLE 2.14 Cutpoint Grouping

Weights of 18- to 24-Year-Old Males The U.S. National Center for Health Statistics publishes data on weights and heights by age and sex in the document *Vital and Health Statistics*. The weights shown in Table 2.8, given to the nearest tenth of a pound, were obtained from a sample of 18- to 24-year-old males. Use cutpoint grouping to organize these data into frequency and relative-frequency distributions. Use a class width of 20 and a first cutpoint of 120.

TABLE 2.8
Weights, in pounds, of 37 males aged 18–24 years

129.2	185.3	218.1	182.5	142.8
155.2	170.0	151.3	187.5	145.6
167.3	161.0	178.7	165.0	172.5
191.1	150.7	187.0	173.7	178.2
161.7	170.1	165.8	214.6	136.7
278.8	175.6	188.7	132.1	158.5
146.4	209.1	175.4	182.0	173.6
149.9	158.6			

Solution

Weight (lb)	Frequency	Relative frequency
120–under 140	3	0.081
140–under 160	9	0.243
160–under 180	14	0.378
180–under 200	7	0.189
200–under 220	3	0.081
220–under 240	0	0.000
240–under 260	0	0.000
260–under 280	1	0.027
	37	0.999

Organizing Quantitative Data

Terms Used in Cutpoint Grouping

Lower class cutpoint: The smallest value that could go in a class.

Upper class cutpoint: The smallest value that could go in the next-higher class (equivalent to the lower cutpoint of the next-higher class).

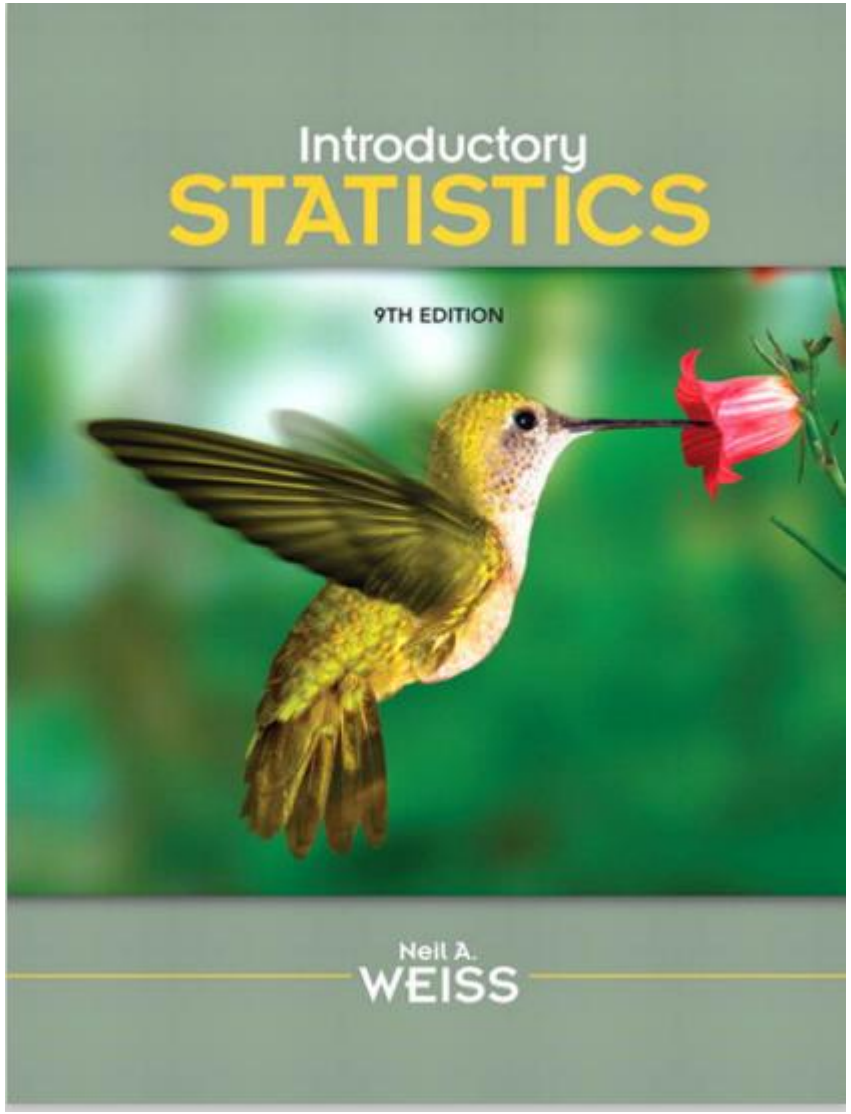
Class width: The difference between the cutpoints of a class.

Class midpoint: The average of the two cutpoints of a class.

Exercise # 2.3

Do questions 2.52-2.63

Page # 65-66



Graphs (Presentation of Data)

Bar Charts

Another graphical display for qualitative data is the *bar chart*. Frequencies, relative frequencies, or percents can be used to label a bar chart. Although we primarily use relative frequencies, some of our applications employ frequencies or percents.

Bar Chart

A **bar chart** displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value. The bars should be positioned so that they do not touch each other.

Note: Bar chart can be drawn for quantitate data as well. (Usually x-axis represents Class interval where y-axis represents frequency of data)

To Construct a Bar Chart

Step 1 **Obtain a relative-frequency distribution of the data by applying Procedure 2.2.**

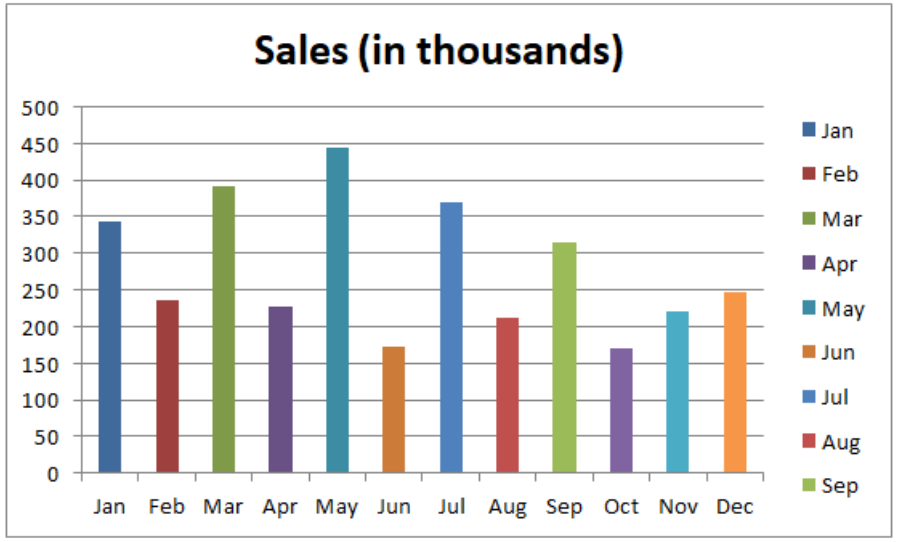
Step 2 **Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.**

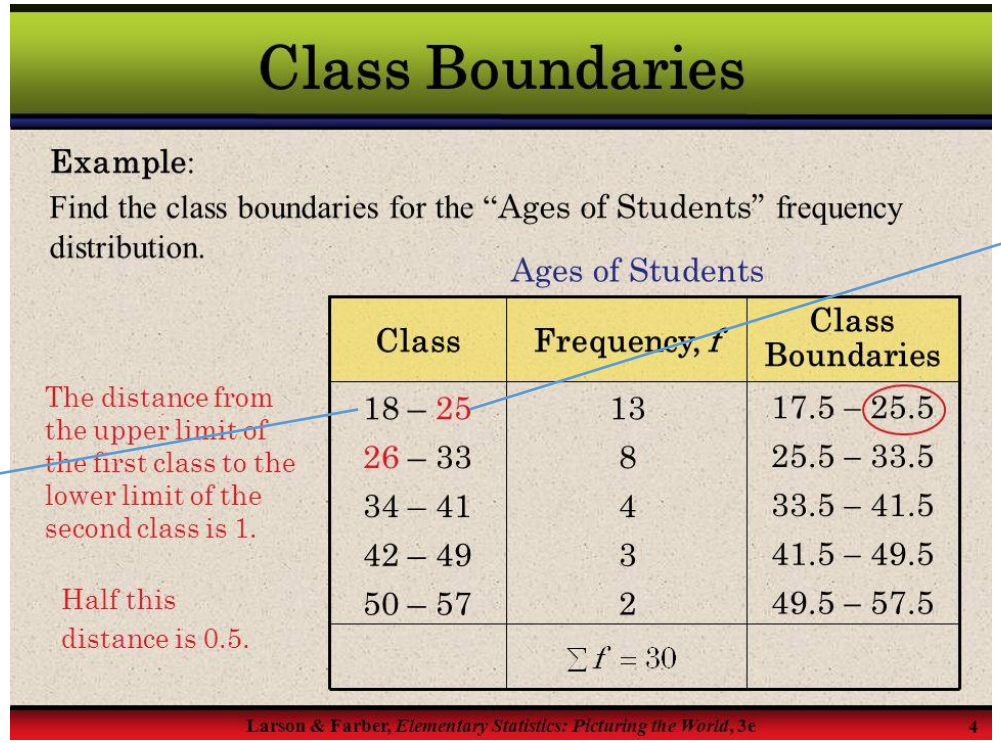
Step 3 **For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.**

Step 4 **Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative frequency."**

Example

Month	Sales (in thousands)
Jan	344
Feb	237
Mar	391
Apr	227
May	445
Jun	173
Jul	370
Aug	212
Sep	315
Oct	171
Nov	220
Dec	246





25+c.f

18-c.f

$$\boldsymbol{Correction\ factor(c.f)} = \frac{lower\ class\ limit\ of\ next\ class - upper\ class\ limit\ of\ previous\ class}{2}$$

Histogram

A **histogram** displays the classes of the quantitative data on a horizontal axis and the frequencies (relative frequencies, percents) of those classes on a vertical axis. The frequency (relative frequency, percent) of each class is represented by a vertical bar whose height is equal to the frequency (relative frequency, percent) of that class. The bars should be positioned so that they touch each other.

- For single-value grouping, we use the distinct values of the observations to label the bars, with each such value centered under its bar.
- For limit grouping or cutpoint grouping, we use the lower class limits (or, equivalently, lower class cutpoints) to label the bars. *Note:* Some statisticians and technologies use class marks or class midpoints centered under the bars.

To Construct a Histogram

Step 1 **Obtain a frequency (relative-frequency, percent) distribution of the data.**

Step 2 **Draw a horizontal axis on which to place the bars and a vertical axis on which to display the frequencies (relative frequencies, percents).**

Step 3 **For each class, construct a vertical bar whose height equals the frequency (relative frequency, percent) of that class.**

Step 4 **Label the bars with the classes, as explained in Definition 2.9, the horizontal axis with the name of the variable, and the vertical axis with "Frequency" ("Relative frequency," "Percent").**

EXAMPLE 2.15 Histograms

TVs, Days to Maturity, and Weights Construct frequency histograms and relative-frequency histograms for the data on number of televisions per household (Example 2.12), days to maturity for short-term investments (Example 2.13), and weights of 18- to 24-year-old males (Example 2.14).

Solution We previously grouped the three data sets using single-value grouping, limit grouping, and cutpoint grouping, respectively, as shown in Tables 2.5, 2.7, and 2.9. We repeat those tables here in Table 2.10.

TABLE 2.10 Frequency and relative-frequency distributions for the data on (a) number of televisions per household, (b) days to maturity for short-term investments, and (c) weights of 18- to 24-year-old males

Number of TVs	Frequency	Relative frequency
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04

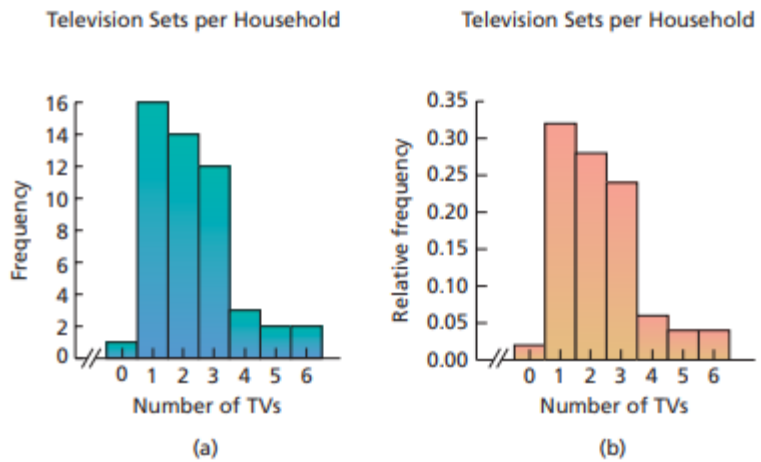
(a) Single-value grouping

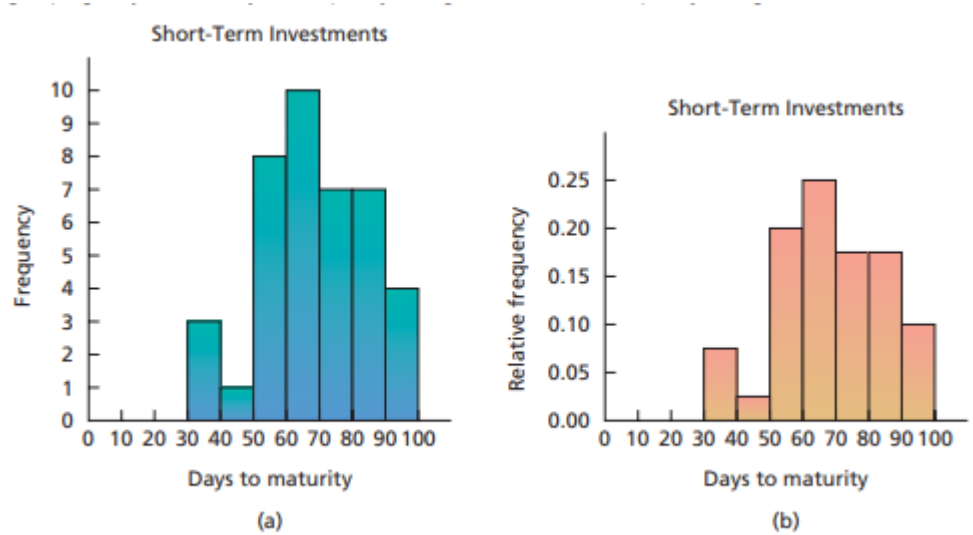
Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

(b) Limit grouping

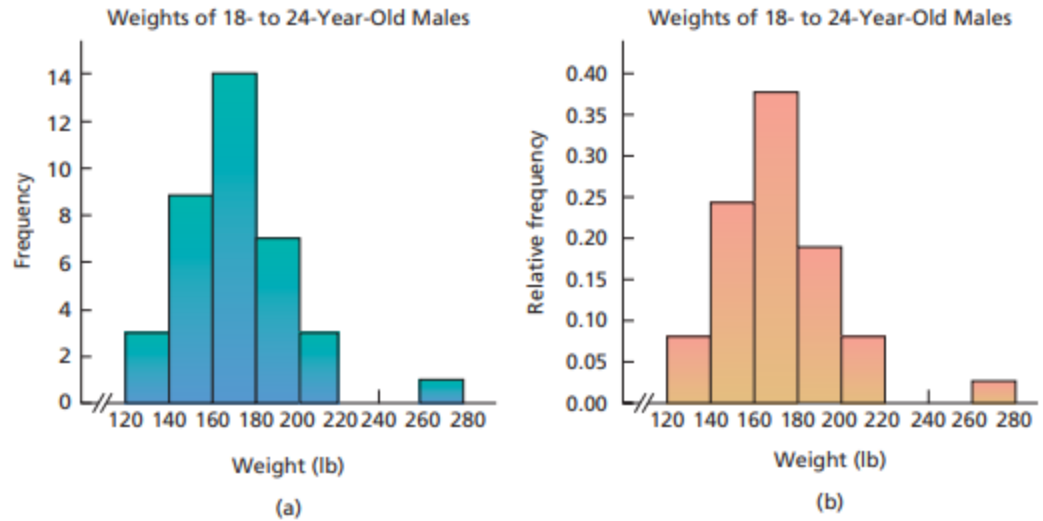
Weight (lb)	Frequency	Relative frequency
120–under 140	3	0.081
140–under 160	9	0.243
160–under 180	14	0.378
180–under 200	7	0.189
200–under 220	3	0.081
220–under 240	0	0.000
240–under 260	0	0.000
260–under 280	1	0.027

(c) Cutpoint grouping





Cutpoint grouping. Weight of 18- to 24-year-old males: (a) frequency histogram; (b) relative-frequency histogram

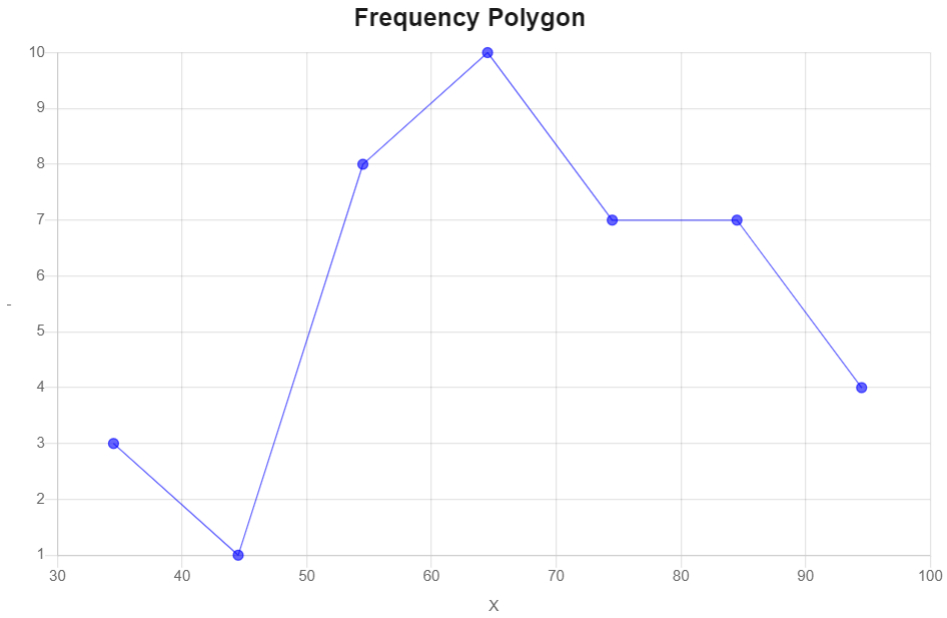


Frequency polygon: (Drawn b/w x and f)

Example

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

x	Frequency f
24.5	0
34.5	3
44.5	1
54.5	8
64.5	10
74.5	7
84.5	7
94.5	4
104.5	0

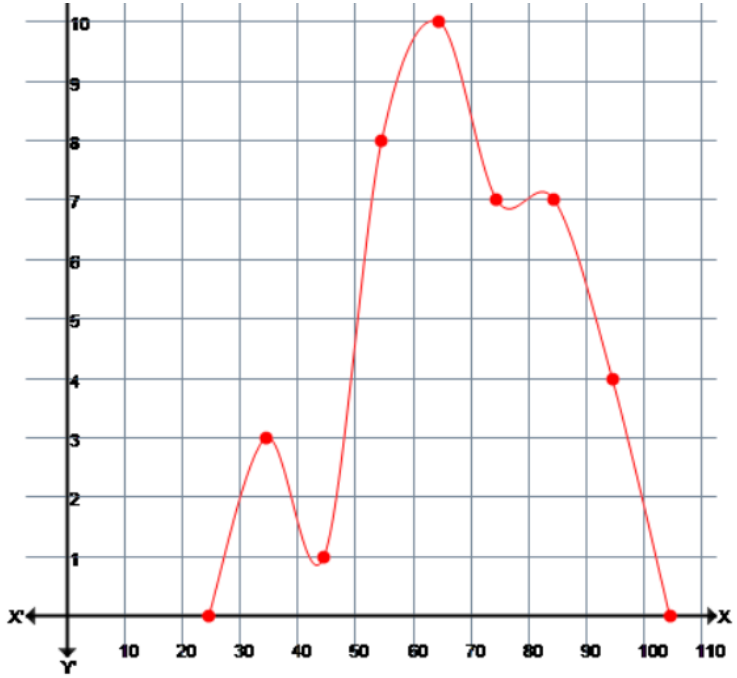


Frequency curve:
(Drawn b/w x and f)

Example

Days to maturity	Frequency	Relative frequency
30–39	3	0.075
40–49	1	0.025
50–59	8	0.200
60–69	10	0.250
70–79	7	0.175
80–89	7	0.175
90–99	4	0.100

x	Frequency f
24.5	0
34.5	3
44.5	1
54.5	8
64.5	10
74.5	7
84.5	7
94.5	4
104.5	0



Dotplots

Another type of graphical display for quantitative data is the **dotplot.** Dotplots are particularly useful for showing the relative positions of the data in a data set or for comparing two or more data sets. Procedure 2.6 presents a method for constructing a dotplot.

PROCEDURE 2.6 To Construct a Dotplot

Step 1 **Draw a horizontal axis that displays the possible values of the quantitative data.**

Step 2 **Record each observation by placing a dot over the appropriate value on the horizontal axis.**

Step 3 **Label the horizontal axis with the name of the variable.**

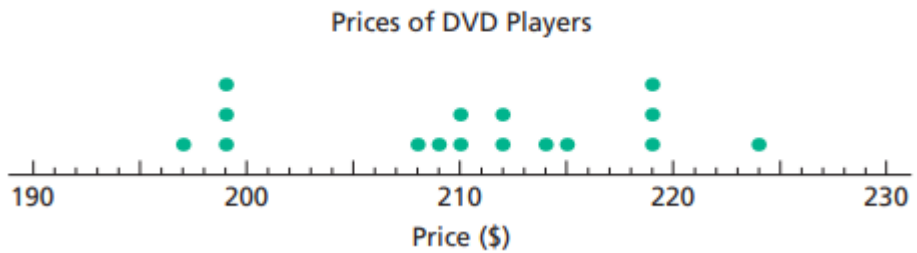
Example

Prices of DVD Players One of Professor Weiss's sons wanted to add a new DVD player to his home theater system. He used the Internet to shop and went to pricewatch.com. There he found 16 quotes on different brands and styles of DVD players. Table 2.11 lists the prices, in dollars. Construct a dotplot for these data.

TABLE 2.11

Prices, in dollars, of 16 DVD players

210	219	214	197
224	219	199	199
208	209	215	199
212	212	219	210



Stem-and-Leaf Diagrams

Statisticians continue to invent ways to display data. One method, developed in the 1960s by the late Professor John Tukey of Princeton University, is called a **stem-and-leaf diagram,** or **stemplot.** This ingenious diagram is often easier to construct than either a frequency distribution or a histogram and generally displays more information.

With a stem-and-leaf diagram, we think of each observation as a **stem**—consisting of all but the rightmost digit—and a **leaf,** the rightmost digit. In general, stems may use as many digits as required, but each leaf must contain only one digit.

Procedure 2.7 presents a step-by-step method for constructing a stem-and-leaf diagram.

PROCEDURE 2.7 To Construct a Stem-and-Leaf Diagram

Step 1 **Think of each observation as a stem—consisting of all but the rightmost digit—and a leaf, the rightmost digit.**

Step 2 **Write the stems from smallest to largest in a vertical column to the left of a vertical rule.**

Step 3 **Write each leaf to the right of the vertical rule in the row that contains the appropriate stem.**

Step 4 **Arrange the leaves in each row in ascending order.**

Example

Days to Maturity for Short-Term Investments Table 2.6 displays the number of days to maturity for 40 short-term investments. The data are from *BARRON'S* magazine. Use limit grouping, with grouping by 10s, to organize these data into frequency and relative-frequency distributions.

TABLE 2.6
Days to maturity for 40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Stems	Leaves
3	6 8 9
4	7
5	0 1 1 3 5 5 6 7
6	0 2 3 4 4 5 6 7 8 9
7	0 0 0 1 5 8 9
8	0 1 3 5 6 7 9
9	5 8 9 9

Stem-leaf diagram

EXAMPLE 2.18 Stem-and-Leaf Diagrams

Cholesterol Levels According to the *National Health and Nutrition Examination Survey*, published by the Centers for Disease Control, the average cholesterol level for children between 4 and 19 years of age is 165 mg/dL. A pediatrician tested the cholesterol levels of several young patients and was alarmed to find that many had levels higher than 200 mg/dL. Table 2.13 presents the readings of 20 patients with high levels. Construct a stem-and-leaf diagram for these data by using

a. one line per stem. **b.** two lines per stem.

TABLE 2.13
Cholesterol levels for 20 high-level patients

210	209	212	208
217	207	210	203
208	210	210	199
215	221	213	218
202	218	200	214

19	9
20	0 2 3 7 8 8 9
21	0 0 0 0 2 3 4 5 7 8 8
22	1

(a)

19	
19	9
20	0 2 3
20	7 8 8 9
21	0 0 0 0 2 3 4
21	5 7 8 8
22	1
22	

(b)

Pie Charts

Another method for organizing and summarizing data is to draw a picture of some kind. The old saying "a picture is worth a thousand words" has particular relevance in statistics—a graph or chart of a data set often provides the simplest and most efficient display.

Two common methods for graphically displaying qualitative data are *pie charts* and *bar charts*. We begin with pie charts.

Pie-chart

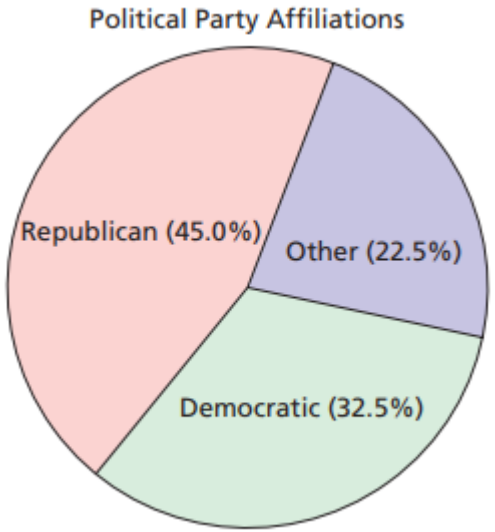
DEFINITION 2.5

Pie Chart

A **pie chart** is a disk divided into wedge-shaped pieces proportional to the relative frequencies of the qualitative data.

Example

Party	Relative frequency		Angle
Democratic	0.325	← *13/40*	$0.325 * 360 = 117$
Republican	0.450	← *18/40*	$0.450 * 360 = 162$
Other	0.225	← *9/40*	$0.225 * 360 = 81$
	1.000		



Note: Pie chart can be drawn for quantitate data as well. Each sector of circle represents x and class intervals in single value grouping and limit grouping

Exercise # 2.2
Do questions 2.18-2.29
Page # 48-50

Exercise # 2.3
Do questions 2.64-2.71
Page # 66-67