Organizing and Graphing Quantitative Data

To organize quantitative data, we first group the observations into classes (also known as categories or bins) and then treat the classes as the distinct values of qualitative data. Consequently, once we group the quantitative data into classes, we can construct frequency and relative-frequency distributions of the data in exactly the same way as we did for qualitative data. Several methods can be used to group quantitative data into classes. Here we discuss three of the most common methods: single-value grouping, limit grouping, and cut-point grouping.

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 singlevalue classes, and this method of grouping quantitative data is called single-value grouping.

Example:

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.

TABLE 2.4
Number of TV sets in each of 50 randomly selected households

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 The (single-value) classes are the distinct values of the data in Table 2.4, which are the numbers 0, 1, 2, 3, 4, 5, and 6. See the first column of Table 2.5.

Tallying the data in Table 2.4, we get the frequencies shown in the second column of Table 2.5. Dividing each such frequency by the total number of observations, 50, we get the relative frequencies in the third column of Table 2.5.

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

TABLE 2.5

Frequency and relative-frequency distributions, using single-value grouping, for the number-of-TVs data in Table 2.4

Thus, the first and second columns of Table 2.5 provide a frequency distribution of the data in Table 2.4, and the first and third columns provide a relative-frequency distribution.

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

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.

important guidelines for grouping:

1. *The number of classes should be small enough to provide an effective summary but large enough to display the relevant characteristics of the data.*
2. *Each observation must belong to one, and only one, class.*
3. *Whenever feasible, all classes should have the same width.*

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

Solution Because we are grouping by 10s and the shortest maturity period is 36 days, our first class is 30–39, that is, for maturity periods from 30 days up to, and including, 39 days. The longest maturity period is 99 days, so grouping by 10s results in the seven classes given in the first column of Table 2.7 on the next page.

Next we tally the data in Table 2.6 into the classes. For instance, the first investment in Table 2.6 has a 70-day maturity period, calling for a tally mark on the line for the class 70–79 in Table 2.7. The results of the tallying procedure are shown in the second column of Table 2.7.

TABLE 2.7

Frequency and relative-frequency distributions, using limit grouping, for the days-to-maturity data in Table 2.6

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

Counting the tallies for each class, we get the frequencies in the third column of Table 2.7. Dividing each such frequency by the total number of observations, 40, we get the relative frequencies in the fourth column of Table 2.7.

Thus, the first and third columns of Table 2.7 provide a frequency distribution of the data in Table 2.6, and the first and fourth columns provide a relative-frequency distribution.

Cut point Grouping:

A third way to group quantitative data is to use class cut points. 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 cut point of the class, and the smallest value that could go in the next higher class is called the upper cut point of the class. Note that the lower cut point of a class is the same as its lower limit and that the upper cut point of a class is the same as the lower limit of the next higher class. The method of grouping quantitative data by using cut points is called cut point grouping. This method is particularly useful when the data are continuous and are expressed with decimals.

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.

Example:

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 Because we are to use a first cutpoint of 120 and a class width of 20, our first class is 120–under 140, as shown in the first column of Table 2.9. This class is for weights of 120 lb up to, but not including, weights of 140 lb. The largest weight in Table 2.8 is 278.8 lb, so the last class in Table 2.9 is 260–under 280.

Tallying the data in Table 2.8 gives us the frequencies in the second column of Table 2.9. Dividing each such frequency by the total number of observations, 37, we get the relative frequencies (rounded to three decimal places) in the third column of Table 2.9.

TABLE 2.9

Frequency and relative-frequency distributions, using cutpoint grouping, for the weight data in Table 2.8

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

Thus, the first and second columns of Table 2.9 provide a frequency distribution of the data in Table 2.8, and the first and third columns provide a relative-frequency distribution.

Cumulative Frequency Distributions

Cumulative Frequency Distribution A *cumulative frequency distribution* gives the total number of values that fall below the upper boundary of each class.

Graphing Grouped Data

Histograms:

Another method for organizing and summarizing data is to draw a picture of some kind. Three common methods for graphically displaying quantitative data are histograms, dot plots, and stem-and-leaf diagrams. We begin with histograms

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.

As expected, a histogram that uses frequencies on the vertical axis is called a frequency histogram. Similarly, a histogram that uses relative frequencies or percentage on the vertical axis is called a relative-frequency histogram or percent histogram, respectively

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:

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

Referring to Tables 2.10(a), 2.10(b), and 2.10(c), we applied Procedure 2.5 to construct the histograms in Figs. 2.4, 2.5, and 2.6, respectively, on the next page.

You should observe the following facts about the histograms in Figs. 2.4, 2.5, and 2.6:

- In each figure, the frequency histogram and relative-frequency histogram have the same shape, and the same would be true for the percent histogram. This result holds because frequencies, relative-frequencies, and percents are proportional.
- Because the histograms in Fig. 2.4 are based on single-value grouping, the distinct values (numbers of TVs) label the bars, with each such value centered under its bar.
- Because the histograms in Figs. 2.5 and 2.6 are based on limit grouping and cutpoint grouping, respectively, the lower class limits (or, equivalently, lower class cutpoints) label the bars.
- We did not show percent histograms in Figs. 2.4, 2.5, and 2.6. However, each percent histogram would look exactly like the corresponding relative-frequency histogram, except that the relative frequencies would be changed to percents (obtained by multiplying each relative frequency by 100) and "Percent," instead of "Relative frequency," would be used to label the vertical axis.

FIGURE 2.4 Single-value grouping. Number of TVs per household: (a) frequency histogram; (b) relative-frequency histogram

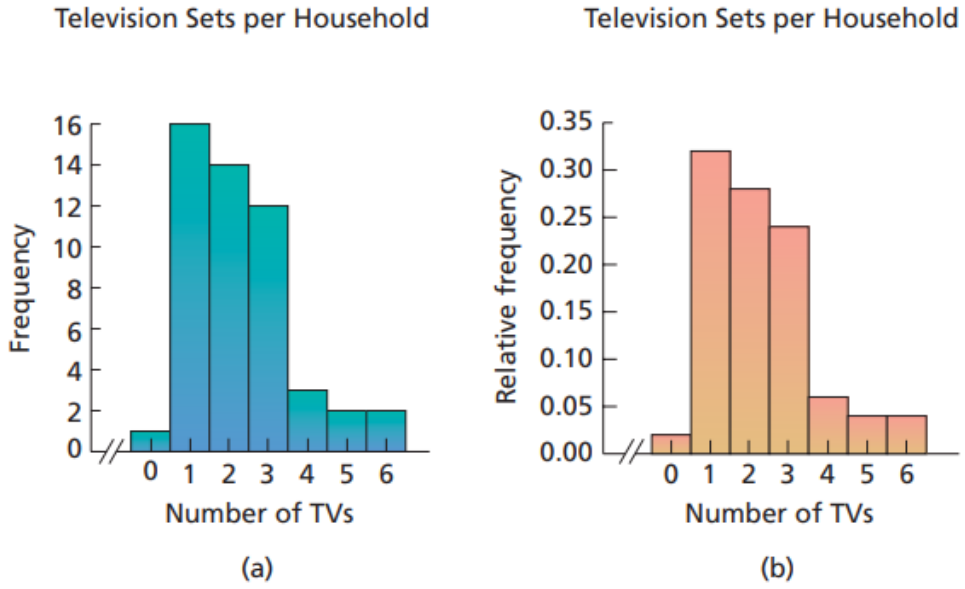


FIGURE 2.5 Limit grouping. Days to maturity: (a) frequency histogram; (b) relative-frequency histogram

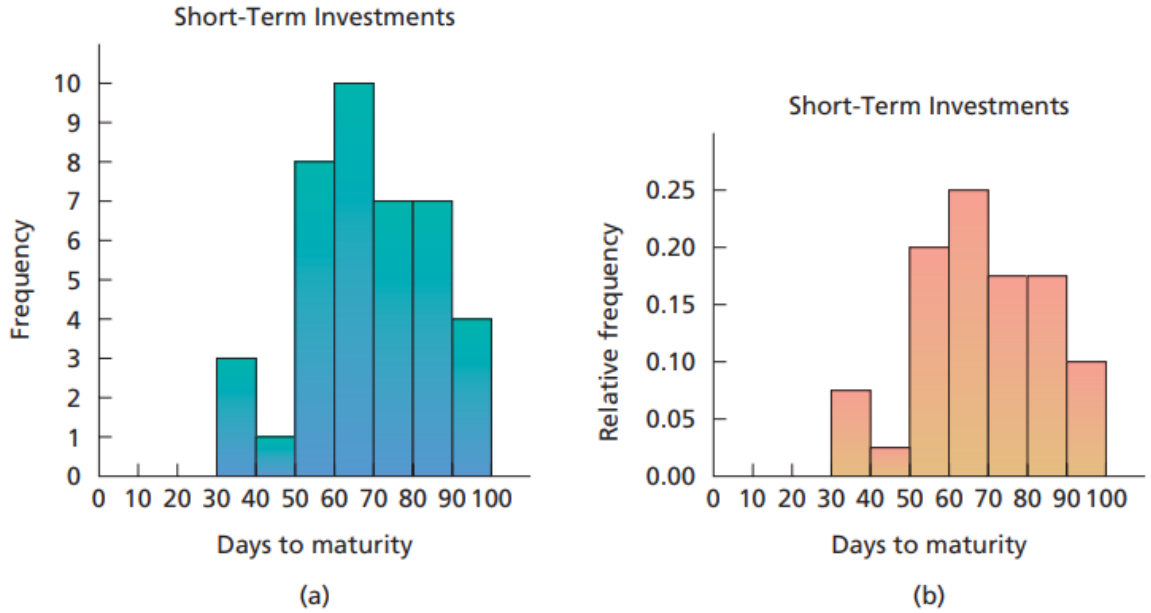
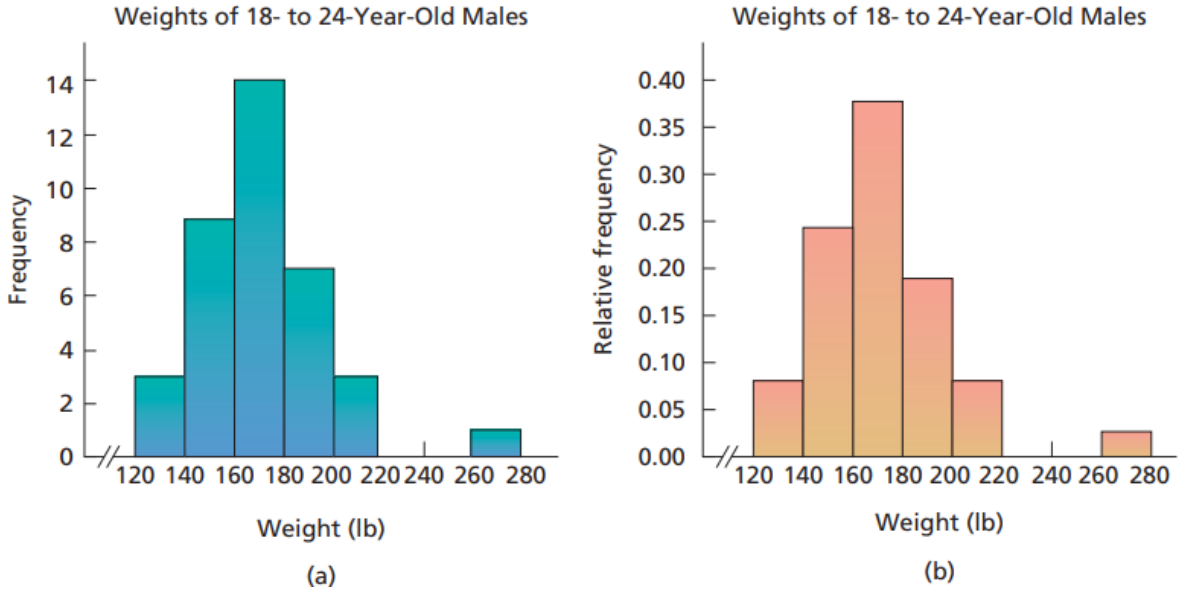


FIGURE 2.6 Cutpoint grouping. Weight of 18- to 24-year-old males: (a) frequency histogram; (b) relative-frequency histogram



- The symbol // is used on the horizontal axes in Figs. 2.4 and 2.6. This symbol indicates that the zero point on that axis is not in its usual position at the intersection of the horizontal and vertical axes. Whenever any such modification is made, whether on a horizontal axis or a vertical axis, the symbol // or some similar symbol should be used to indicate that fact.

Dot plots:

Another type of graphical display for quantitative data is the dot plot. Dot plots are particularly useful for showing the relative positions of the data in a data set or for comparing two or more data sets

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

Solution We apply Procedure 2.6.

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

See the horizontal axis in Fig. 2.7 at the top of the next page.

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

The first price is \$210, which calls for a dot over the "210" on the horizontal axis in Fig. 2.7. Continuing in this manner, we get all the dots shown in Fig. 2.7.

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

The variable here is "Price," with which we label the horizontal axis in Fig. 2.7.

FIGURE 2.7
Dotplot for DVD-player prices in Table 2.11