CHAPTER 2

Frequency Distributions and Graphs

Summary

When you understand the material in this chapter, you will be able to:

1. organize a set of measurements into a frequency distribution;
2. choose and construct an appropriate graph; and
3. identify graphs by name and recognize the direction of skew, if any.

This chapter starts by presenting a set of unorganized scores and proceeds with organization techniques of descriptive statistics. The unorganized Satisfaction With Life Scale scores were compiled into a ***simple frequency distribution*** table, which was then further compiled into a ***grouped frequency distribution*** table. Constructing class intervals for grouped frequency distributions is explained in Appendix B in the textbook.

Graphs are a different (and sometimes superior) way to present frequency distributions and relationships among variables. Frequency distributions give information about one variable and can be presented as a frequency polygon, a histogram, or a bar graph. ***Frequency polygons,*** which show data points connected by lines, and ***histograms,*** which have bars raised to the appropriate frequency, are both used for *continuous quantitative* data. ***Bar graphs***, which have separated bars raised to the appropriate frequency, are used for *categorical* data and *discrete quantitative* data. A ***line graph*** is used to present the relationship between two variables.

We also use terms to describe the shapes of distributions we see in graphs. The ***normal distribution***—which will be very important in the chapters to come—follows a bell-shaped pattern. ***Rectangular distributions*** are flat on top, representing an equal frequency for all values on the *x*-axis. Both of these distributions are symmetrical. ***Skewed distributions*** are unbalanced, nonsymmetrical distributions, in which scores cluster at either the lower end of the distribution (positively skewed) or the higher end of the distribution (negatively skewed). Finally, ***bimodal distributions*** have two distinct peaks.

Because of the persuasive power of graphs, it is important to remember that a properly constructed graph allows the reader to quickly and easily make informed conclusions about data, but a poorly designed graph gives your audience a false impression of the scores. Thus, drawing a graph requires careful thought. You should make some rough drafts of a graph and then refine one that conveys the information best.

Multiple-Choice Questions

1. The graph that is used to present data on two variables rather than one is the _____.

   a. frequency polygon
   b. histogram
   c. bar graph
   d. line graph

2. Assume you give a survey to 25 college students about their attitudes toward having dress codes in high schools. Students rate their attitudes on a scale of 1 (*not at all in favor*) to 5 (*very much in favor*). When the data are initially collected, the raw scores could be organized into _____.

   a. a grouped frequency distribution
   b. a simple frequency distribution
   c. a line graph
   d. both a. and b

3. The best way to determine if a graph is a histogram or a bar graph is to look at _____.

a. the height of the bars
b. whether the bars are wide or narrow
c. the kind of variable on the *y*-axis
d. the kind of variable on the *x*-axis

4. To present a frequency distribution of nominal data you should use a _____.

a. frequency polygon
b. bar graph
c. histogram
d. line graph

5. Which of the following is *not* used to present a frequency distribution?

a. bar graph
b. histogram
c. frequency polygon
d. line graph

6. The two graphs explained in this chapter that are most similar in terms of the data they represent are _____ and _____.

a. bar graph; frequency polygon
b. histogram; frequency polygon
c. bar graph; line graph
d. histogram; line graph

7. Psychologists sometimes identify parenting styles as authoritarian, authoritative, and permissive. If 100 people were surveyed and their parenting styles determined, the data could be compiled into a frequency distribution. This frequency distribution would show values on a(n) _____ scale and would be best represented visually with a _____.

a. nominal; bar graph
b. ordinal; frequency polygon
c. interval; histogram
d. nominal; histogram

8. A distribution with two separated peaks is a _____ distribution; one with a bell-shape is a _____ distribution.

a. normal; rectangular
b. rectangular; bimodal
c. bimodal; normal
d. normal; bell-shaped

9. Skewness refers to _____.

a. the shape of the curve
b. the number of items in the curve
c. the height of the curve
d. none of the above

10. Suppose a frequency distribution with a range from 0 to 100 was positively skewed. The greatest frequency of scores would be expected around _____.

a. 25
b. 50
c. 75
d. any of the above are possible for such a distribution

11. When the shape of a distribution has two peaks, it is called _____.

a. normal
b. skewed
c. rectangular
d. bimodal

12. The horizontal axis of a graph for a frequency distribution is called _____.

a. a line graph
b. the ordinate
c. the abscissa
d. a histogram

13. Assume you collect data from psychology majors who indicate their favorite class in psychology. Students choose from five classes. What kind of graph should you use to display the data?

a. histogram
b. bar graph
c. line graph
d. cannot be determined from the information given

14. Grouped frequency distributions and simple frequency distributions differ in _____.

a. the type of data displayed
b. the range of scores covered by the distribution
c. the conclusions that can be drawn about skewness
d. all of the above

15. Graphs are beneficial because _____.

a. they allow you to easily see the shape of a distribution
b. they communicate numbers, words, and pictures simultaneously
c. they can be used to make a persuasive argument
d. all of the above

Short-Answer Questions

1. In a few sentences, distinguish among the frequency polygon, the histogram, the bar graph, and the line graph.

2. For each situation described below, decide which of the four types of graphs covered in this chapter is most appropriate to represent the data and why. Briefly describe what the graph might look like.

a. a poll of 100 houses to determine what TV show the household is watching
b. a poll of 100 houses to determine household income
c. hours of contact with students per week and salaries for professors
d. numbers reported by each member of a large sociology class when asked, "How many close friends do you

have?"

e. number of correct answers on a psychology exam
f. income level and IQ scores of 50 people
g. attendance figures at a political debate, a ballet, a dramatic production, and a rock concert, all of which were held on the same night
h. average annual tuition at 200 liberal arts colleges from 2009 to 2018
i. number of people entering a museum each hour it is open
j. number of people entering each of four pizza restaurants during one hour

3. For each description below, identify the mistake in the researcher's approach.

a. I will use a line graph to represent how many units were sold (*y*-axis) for each of five car companies (*x*-axis) during the 2017 calendar year.
b. I will use a histogram to plot how the cost of laundry detergent (*y*-axis) has changed over time (*y*-axis).
c. I will use a frequency polygon to represent the relationship between pressure applied to skin (*x*-axis) and experiences of pain (*y*-axis).
d. I will use a histogram to represent how many of seven different types of birds I spotted during a 1-hour observation. I will put bird type on the *x*-axis and number observed on the *y*-axis.

Problems

1. For each distribution of scores, compile a simple frequency distribution and identify the shape of the distribution or the direction of skew.

a. 9, 10, 9, 9, 11, 9, 10, 9, 9, 10
b. 5, 4, 2, 1, 6, 5, 3, 2, 4, 2, 5, 5, 2
c. 9, 11, 12, 10, 12, 11, 8, 12, 10, 11, 12
d. 8, 7, 7, 9, 5, 7, 7, 8, 6, 6, 7, 7

2. A small group of people in a college town set out to promote greater use of bicycles. One member of the group was eager to measure progress. Before any promotion activity, she assessed the frequency of five different modes of transportation by asking people how they got to work or school that morning. Responses were scored as auto = 0, bicycle = 1, motorcycle = 2, walked = 3, bus = 4. Arrange the scores into an appropriate frequency distribution and then graph it. Write a sentence or two explaining what your analysis shows.

4	1	0	0	3	0	2	0	4	0
3	1	0	4	0	0	0	2	0	0
0	0	2	0	0	4	0	0	2	0
1	0	4	0	3	0	0	4	0	2

3. Below is a grouped frequency distribution of age data collected by the U.S. Census Bureau in 2017 (https://tinyurl.com/peokrom, Retrieved June 23, 2018) for adults aged 18 to 62 (rounded to the nearest million). Your task is to create both a histogram and a frequency polygon of these data and then write a sentence describing the shape of the distribution.

Age (class interval)	*Midpoint X*	*f* (in millions)
58-62	60	21
53-57	55	22
48-52	50	21
43-47	45	20
38-42	40	20
33-37	35	21
28-32	30	22
23-27	25	23
18-22	20	21
		$N = 191$

4. The data below represent the average temperature (rounded to the nearest degree) in New York City's Central Park (in degrees Fahrenheit) during each month for the years 2007 and 2017 (https://www.weather.gov/media/okx/Climate/CentralPark/monthlyannualtemp.pdf, Retrieved June 23, 2018). Draw a line graph for these two data sets and write a short interpretation that compares the two graphs.

2007		2017	
X	*Y*	*X*	*Y*
January	38	January	38
February	36	February	42
March	43	March	40
April	55	April	57
May	60	May	61
June	74	June	72
July	78	July	77
August	74	August	74
September	69	September	71
October	55	October	64
November	46	November	47
December	38	December	33

ANSWERS

Multiple-Choice Questions

1. d
2. b
   Explanation: Because there are only five possible values for the *X* variable (1-5), a grouped frequency distribution is not recommended. A general rule is to use grouped frequency distributions when the range of scores is greater than 20.
3. d
   Explanation: Histograms are used to represent *continuous quantitative* data, whereas bar graphs are used to represent *categorical* and *discrete* data. Also, bar graphs have a space between separate bars and histograms have bars that touch.
4. b
5. d
   Explanation: Frequency distributions summarize data for *one* variable; line graphs represent relationships between *two* variables.
6. b
   Explanation: Both histograms and frequency polygons are used to present data for *continuous quantitative* variables. Bar graphs are used for *categorical* and *discrete quantitative* variables. Line graphs are used to show the relationship between *two* variables.
7. a
8. c
9. a
10. a
    Explanation: For a positively skewed distribution, the majority of scores cluster at the lower end of the scale, with smaller frequencies at higher values.
11. d
12. c
13. b
    Explanation: Because the data would be *categorical* in nature (nominal scale), a bar graph is most appropriate.
14. b
    Explanation: If a distribution covers a wide range of scores (general rule: more than 20 values on the *X* variable), a grouped frequency distribution is preferable.
15. d

Short-Answer Questions

1. A frequency polygon uses dots connected by lines to represent frequencies in one or more sets of continuous quantitative data. A histogram is also used to represent continuous quantitative data, but histograms use bar heights instead of dots to represent frequencies. A histogram is typically used to represent one set of data, but frequency polygons often show two or more sets of data. Bar graphs are visually similar to histograms, but represent *categorical* and *discrete quantitative* data, instead of continuous quantitative data. Because of this difference, bars do not touch one another in bar graphs. Finally, line graphs are used to show the relationship *between two variables*. Unlike frequency polygons, line graphs only connect to the *x*-axis if data exist at those values.
2.
   a. Bar graph, because TV show is a categorical variable. TV show would be presented on the abscissa (perhaps alphabetically or by popularity), with frequency on the ordinate.
   b. Histogram or frequency polygon, because income is a continuous quantitative variable. Income would be presented on the abscissa, with frequency on the ordinate. Because of the wide range of values the *X* variable (income) can have, class intervals would be helpful. If using a frequency polygon, the polygon will be closed at both ends.
   c. Line graph, because the description contains two variables (hours of student contact and salary) whose relationship we want to show. In this case, either variable could be presented on the abscissa or the ordinate.

d. Bar graph, because number of friends is a *discrete* quantitative variable. Number of friends would be presented on the abscissa, with frequency on the ordinate.

e. Bar graph, because number of correct answers is a *discrete* quantitative variable. Number of correct answers would be presented on the abscissa, with frequency on the ordinate.

f. Line graph, because the description contains two variables (income and IQ score) whose relationship we want to show. In this case, either variable could be presented on the abscissa or the ordinate.

g. Bar graph, because type of event is a categorical variable. Type of event would be presented on the abscissa (perhaps alphabetically or by popularity), with number in attendance on the ordinate.

h. Line graph, because the description contains two variables (tuition and time) whose relationship we want to show. It would make the most sense to put year on the abscissa and tuition on the ordinate.

i. Histogram or frequency polygon, because hour of the day is a continuous quantitative variable. Hour of the day would be presented on the abscissa, with number of entries on the ordinate. If using a frequency polygon, the polygon will be closed at both ends.

j. Bar graph, because restaurant is a categorical variable. Each restaurant would be placed on the abscissa (perhaps alphabetically or by popularity), with number entering on the ordinate.

3.

a. A line graph should only be used to represent the relationship between two variables. Here, the researcher wants to represent frequencies (units sold) of a nominal variable (car company), so a bar graph is most appropriate.

b. A line graph should be used when you want to represent the relationship between two variables (cost and time).

c. A line graph should be used when you want to represent the relationship between two variables (pressure and pain).

d. Because the researcher wants to represent frequencies of a nominal variable (bird type), a bar graph is most appropriate.

Problems

1.

a.

X	*f*
11	1
10	3
9	6
	$N = 10$

This distribution is positively skewed, because most scores “pile up” at the lower end of the distribution.

b.

X	*f*
6	1
5	4
4	2
3	1
2	4
1	1
	N = 13

This distribution is bimodal, with peaks at both $X = 2$ and $X = 5$.

c.

X	*f*
12	4
11	3
10	2
9	1
8	1
	N = 11

This distribution is negatively skewed, because most scores "pile up" at the higher end of the distribution.

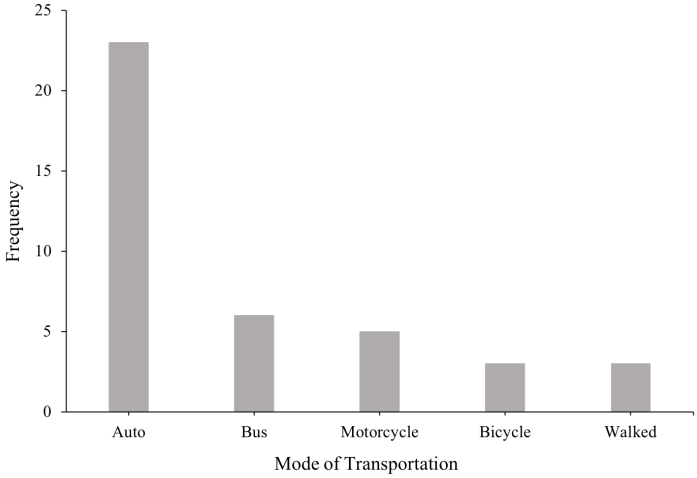
d.

X	*f*
9	1
8	2
7	6
6	2
5	1
	N = 12

This distribution is symmetrical and bell-shaped.

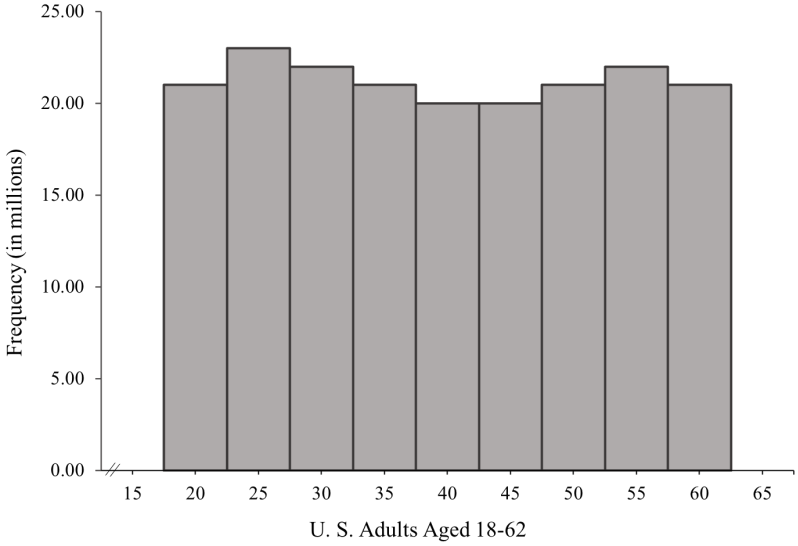
2.

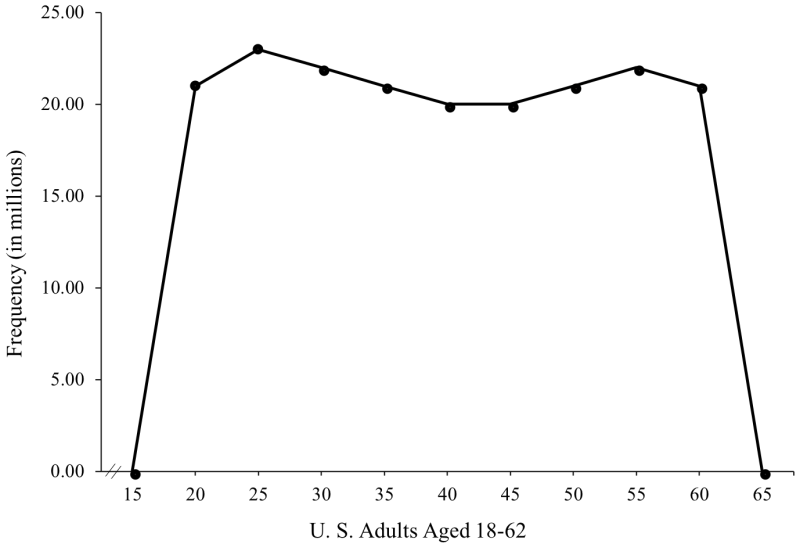
Transportation Mode X	***f***
4	**6**
3	**3**
2	**5**
1	**3**
0	**23**
	***N* = 40**



Presently, most people drive to work or school and very few people walk or ride a bicycle.

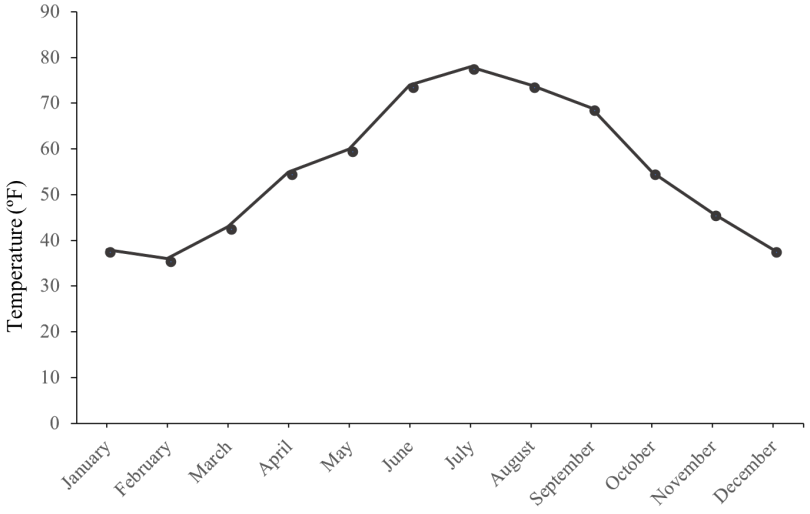
3.





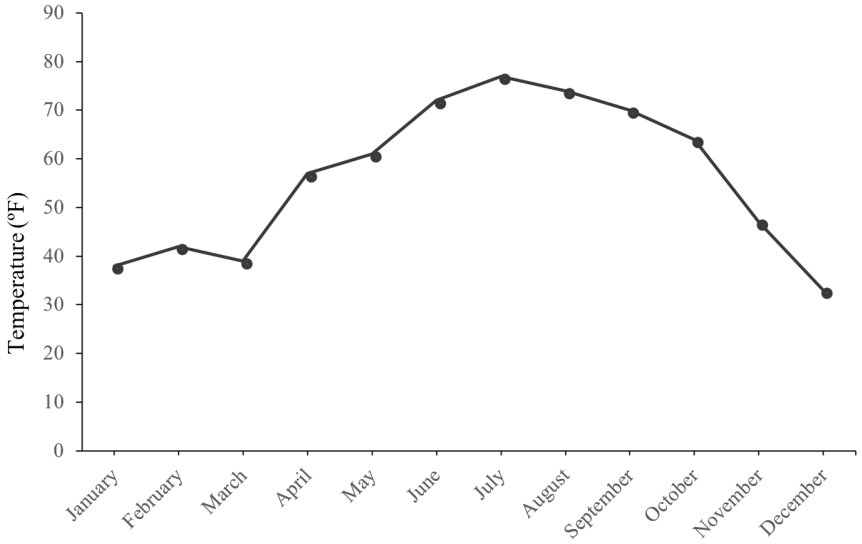
Although the overall shape of the distribution looks somewhat rectangular, remember that the data are presented in *millions* of people; thus, the shape of this distribution is probably best described as bimodal, with peaks around ages 25 and 55.

4.



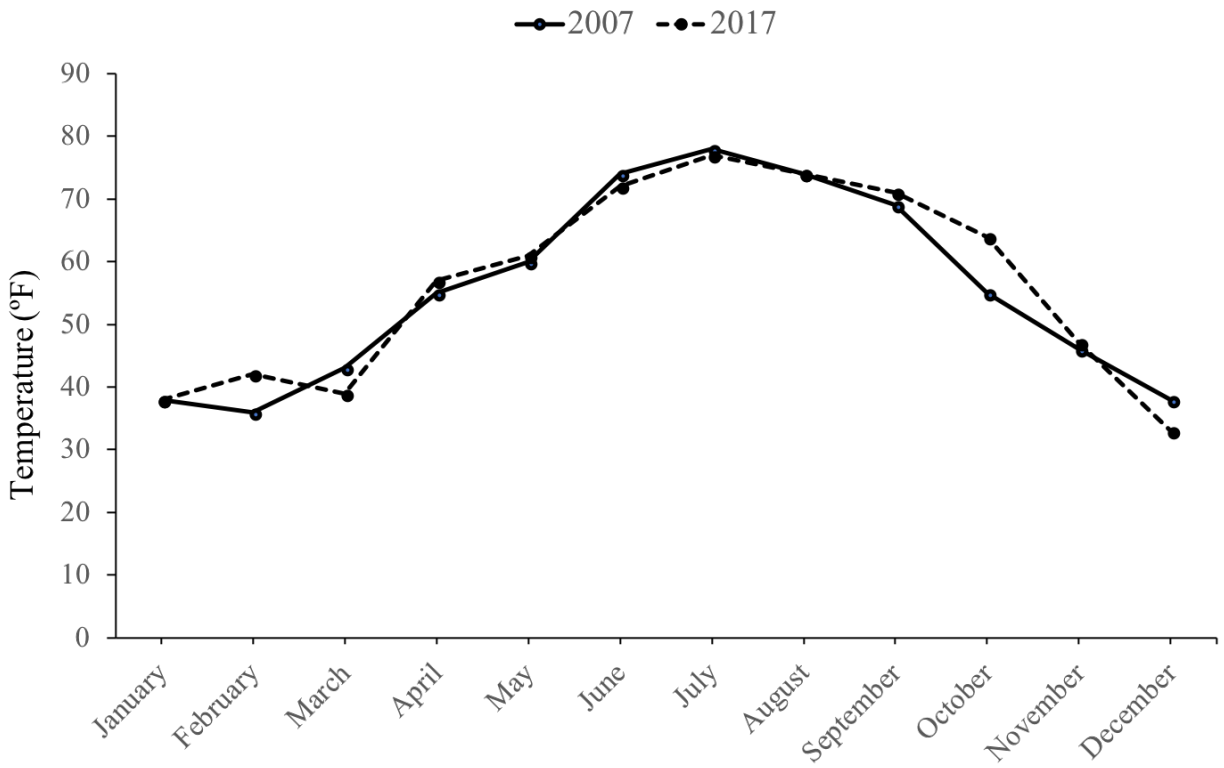
Average Temperature in Central Park for 2007

5.



Average Temperature in Central Park for 2017

These two graphs show similar trends in temperature in Central Park across 2007 and 2017, with the summer months being the hottest. However, October of 2017 was markedly warmer than October 2007. As a note, a nice feature of line graphs is that two distributions of data on the same graph make comparisons easier. If you put both years' data on the same plot, it looks like the graph that follows.



Average Temperature in Central Park in 2007 and 2017