

Appendix B Concepts in Statistics

B.1 Representing Data

▶ What you should learn

- Use line plots to order and analyze data.
- Use stem-and-leaf plots to organize and compare data.
- Use histograms to represent frequency distributions.

▶ Why you should learn it

Line plots, stem-and-leaf plots, and histograms are quick methods for determining which elements in a set of data occur with the greatest frequency. For instance, in Exercise 8 on page B4, you are asked to construct a frequency distribution and a histogram of employees' monthly salary contributions to a retirement plan.

Line Plots

Statistics is the branch of mathematics that studies techniques for collecting, organizing, and interpreting data. In this section, you will study several ways to organize data. The first is a **line plot**, which uses a portion of a real number line to order numbers. Line plots are especially useful for ordering small sets of numbers (about 50 or less) by hand.

Example 1 ▶ Constructing a Line Plot



Use a line plot to organize the following test scores. Which score occurs with the greatest frequency?

93, 70, 76, 67, 86, 93, 82, 78, 83, 86, 64, 78, 76, 66, 83,
83, 96, 74, 69, 76, 64, 74, 79, 76, 88, 76, 81, 82, 74, 70

Solution

Begin by scanning the data to find the smallest and largest numbers. For this data set, the smallest number is 64 and the largest is 96. Next, draw a portion of a real number line that includes the interval $[64, 96]$. To create the line plot, start with the first number, 93, and enter an \times above 93 on the number line. Continue recording \times 's for each number in the list until you obtain the line plot shown in Figure B.1. From the line plot, you can see that 76 occurs with the greatest frequency.

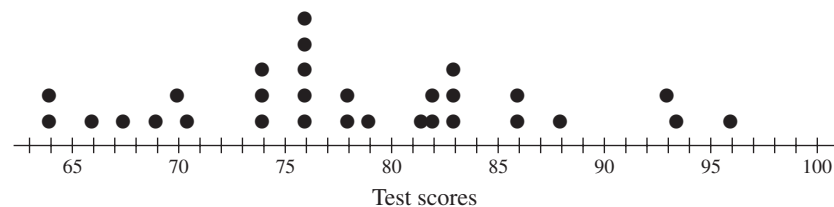


FIGURE B.1



CHECKPOINT

Now try Exercise 3.

Test Scores
93, 70, 76, 58, 86, 93, 82, 78,
83, 86, 64, 78, 76, 66, 83, 83,
96, 74, 69, 76, 64, 74, 79, 76,
88, 76, 81, 82, 74, 70

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

Stem-and-Leaf Plots

Another type of plot that can be used to organize sets of numbers by hand is a **stem-and-leaf plot**. A set of test scores and the corresponding stem-and-leaf plot are shown at the left.

Note that the *leaves* represent the units digits of the numbers and the *stems* represent the tens digits. Stem-and-leaf plots can also be used to compare two sets of data, as shown in the following example.

Example 2 ▶ Comparing Two Sets of Data



AK	6.3	MT	13.6
AL	13.2	NC	12.1
AR	13.9	ND	14.8
AZ	12.8	NE	13.4
CA	10.6	NH	12.0
CO	9.7	NJ	13.0
CT	13.5	NM	12.0
DC	12.0	NV	11.2
DE	13.1	NY	13.0
FL	17.0	OH	13.3
GA	9.5	OK	13.1
HI	13.5	OR	12.7
IA	14.7	PA	15.4
ID	11.4	RI	14.0
IL	11.9	SC	12.3
IN	12.3	SD	14.3
KS	13.0	TN	12.4
KY	12.4	TX	9.8
LA	11.7	UT	8.6
MA	13.3	VA	11.3
MD	11.3	VT	12.9
ME	14.4	WA	11.3
MI	12.3	WI	13.0
MN	12.0	WV	15.3
MO	13.3	WY	12.0
MS	12.1		

Use a stem-and-leaf plot to compare the test scores given on the previous page with the following test scores. Which set of test scores is better?

90, 81, 70, 62, 64, 73, 81, 92, 73, 81, 92, 93, 83, 75, 76,
83, 94, 96, 86, 77, 77, 86, 96, 86, 77, 86, 87, 87, 79, 88

Solution

Begin by ordering the second set of scores.

62, 64, 70, 73, 73, 75, 76, 77, 77, 77, 79, 81, 81, 81, 83,
83, 86, 86, 86, 86, 87, 87, 88, 90, 92, 92, 93, 94, 96, 96

Now that the data have been ordered, you can construct a *double* stem-and-leaf plot by letting the leaves to the right of the stems represent the units digits for the first group of test scores and letting the leaves to the left of the stems represent the units digits for the second group of test scores.

Leaves (2nd Group)	Stems	Leaves (1st Group)
	5	8
4 2	6	4 4 6 9
9 7 7 7 6 5 3 3 0	7	0 0 4 4 4 6 6 6 6 8 8 9
8 7 7 6 6 6 6 3 3 1 1 1	8	1 2 2 3 3 3 6 6 8
6 6 4 3 2 2 0	9	3 3 6

By comparing the two sets of leaves, you can see that the second group of test scores is better than the first group.

CHECKPOINT Now try Exercise 6.

Example 3 ▶ Using a Stem-and-Leaf Plot



Stems	Leaves
6.	3
7.	
8.	6
9.	5 7 8
10.	6
11.	2 3 3 3 4 7 9
12.	0 0 0 0 1 1 3 3 3 4 4 7 8 9
13.	0 0 0 0 1 1 2 3 3 3 4 5 5 6 9
14.	0 3 4 7 8
15.	3 4
16.	
17.	0

The table at the left above shows the percent of the population of each state and the District of Columbia that was at least 65 years old in 2003. Use a stem-and-leaf plot to organize the data. (Source: U.S. Census Bureau)

Solution

Begin by ordering the numbers, as shown below.

6.3, 8.6, 9.5, 9.7, 9.8, 10.6, 11.2, 11.3, 11.3, 11.3, 11.4, 11.7, 11.9, 12.0,
12.0, 12.0, 12.0, 12.0, 12.1, 12.1, 12.3, 12.3, 12.3, 12.4, 12.4, 12.7,
12.8, 12.9, 13.0, 13.0, 13.0, 13.0, 13.1, 13.1, 13.2, 13.3, 13.3, 13.3,
13.4, 13.5, 13.5, 13.6, 13.9, 14.0, 14.3, 14.4, 14.7, 14.8, 15.3, 15.4, 17.0

Next construct the stem-and-leaf plot using the leaves to represent the digits to the right of the decimal points, as shown at the left. From the stem-and-leaf plot, you can see that Alaska has the lowest percent and Florida has the highest percent.

CHECKPOINT Now try Exercise 5.

Histograms and Frequency Distributions

With data such as that given in Example 3, it is useful to group the numbers into intervals and plot the frequency of the data in each interval. For instance, the **frequency distribution** and **histogram** shown in Figure B.2 represent the data given in Example 3.

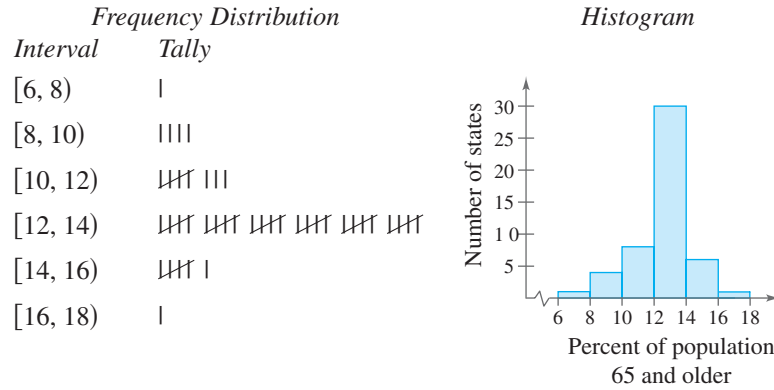


FIGURE B.2

A histogram has a portion of a real number line as its horizontal axis. A histogram is similar to a bar graph, except that the rectangles (bars) in a bar graph can be either horizontal or vertical and the labels of the bars are not necessarily numbers.

Another difference between a bar graph and a histogram is that the bars in a bar graph are usually separated by spaces, whereas the bars in a histogram are not separated by spaces.

Interval	Tally
100–109	
110–119	
120–129	
130–139	
140–149	
150–159	
160–169	
170–179	
180–189	
190–199	

Example 4 ▶ Constructing a Histogram



A company has 48 sales representatives who sold the following numbers of units during the first quarter of 2005. Construct a frequency distribution for this data set.

107	162	184	170	177	102	145	141
105	193	167	149	195	127	193	191
150	153	164	167	171	163	141	129
109	171	150	138	100	164	147	153
171	163	118	142	107	144	100	132
153	107	124	162	192	134	187	177

Solution

To begin constructing a frequency distribution, you must first decide on the number of intervals. There are several ways to group this data. However, because the smallest number is 100 and the largest is 195, it seems that 10 intervals would be appropriate. The first interval would be 100–109, the second interval would be 110–119, and so on. By tallying the data into the 10 intervals, you obtain the distribution shown at the left above. A histogram for the distribution is shown in Figure B.3.

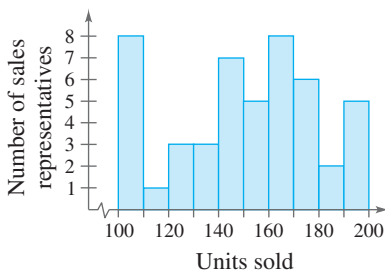


FIGURE B.3



CHECKPOINT

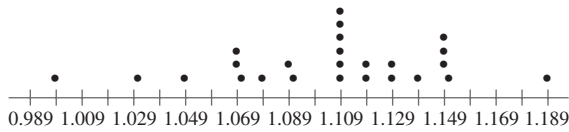
Now try Exercise 7.

B.1 Exercises

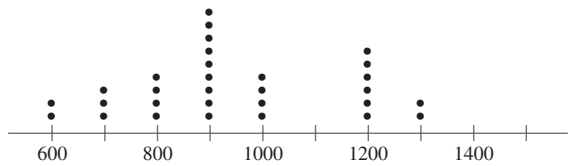
VOCABULARY CHECK: Fill in the blanks.

- _____ is the branch of mathematics that studies techniques for collecting, organizing, and interpreting data.
- A _____ plot and a _____ plot are useful for ordering small sets of numbers by hand.
- A _____ has a portion of a real number line as its horizontal axis, and the bars are not separated by spaces.
- You use a _____ to create a histogram.

1. **Gasoline Prices** The line plot shows a sample of prices of unleaded regular gasoline from 25 different cities.



- What price occurred with the greatest frequency?
 - What is the range of prices?
2. **Agriculture** The line plot shows the weights (to the nearest hundred pounds) of 30 head of cattle sold by a rancher.



- What weight occurred with the greatest frequency?
- What is the range of weights?

Quiz and Exam Scores In Exercises 3–6, use the following scores from a math class of 30 students. The scores are for two 25-point quizzes and two 100-point exams.

Quiz #1 20, 15, 14, 20, 16, 19, 10, 21, 24, 15, 15, 14, 15, 21, 19, 15, 20, 18, 18, 22, 18, 16, 18, 19, 21, 19, 16, 20, 14, 12

Quiz #2 22, 22, 23, 22, 21, 24, 22, 19, 21, 23, 23, 25, 24, 22, 22, 23, 23, 23, 23, 22, 24, 23, 22, 24, 21, 24, 16, 21, 16, 14

Exam #1 77, 100, 77, 70, 83, 89, 87, 85, 81, 84, 81, 78, 89, 78, 88, 85, 90, 92, 75, 81, 85, 100, 98, 81, 78, 75, 85, 89, 82, 75

Exam #2 76, 78, 73, 59, 70, 81, 71, 66, 66, 73, 68, 67, 63, 67, 77, 84, 87, 71, 78, 78, 90, 80, 77, 70, 80, 64, 74, 68, 68, 68

3. Construct a line plot for each quiz. For each quiz, which score occurred with the greatest frequency?

- Construct a line plot for each exam. For each exam, which score occurred with the greatest frequency?
- Construct a stem-and-leaf plot for Exam #1.
- Construct a double stem-and-leaf plot to compare the scores for Exam #1 and Exam #2. Which set of scores is higher?
- Agriculture** The table shows the total numbers of farms (in thousands) in the 50 states in 2002. Use a frequency distribution and a histogram to organize the data. (Source: U.S. Dept. of Agriculture)

AK	1	AL	45	AR	47	AZ	7	CA	80
CO	31	CT	4	DE	2	FL	44	GA	49
HI	5	IA	91	ID	25	IL	73	IN	60
KS	64	KY	87	LA	27	MA	6	MD	12
ME	7	MI	53	MN	81	MO	107	MS	42
MT	28	NC	54	ND	31	NE	49	NH	3
NJ	10	NM	15	NV	3	NY	37	OH	78
OK	83	OR	40	PA	58	RI	1	SC	25
SD	32	TN	88	TX	229	UT	15	VA	48
VT	7	WA	36	WI	77	WV	21	WY	9

- Retirement Contributions** The employees of a company must contribute 7% of their monthly salaries to a company-sponsored retirement plan. The contributed amounts (in dollars) for the company's 35 employees are shown below. Use a frequency distribution and a histogram to organize the data.
100, 200, 130, 136, 161, 156, 209, 126, 135, 98, 114, 117, 168, 133, 140, 124, 172, 127, 143, 157, 124, 152, 104, 126, 155, 92, 194, 115, 120, 136, 148, 112, 116, 146, 96
- Meteorology** The data shows the seasonal snowfall (in inches) at Lincoln, Nebraska for the years 1975 through 2004 (the amounts are listed in order by year). How would you organize the data? Explain your reasoning. (Source: University of Nebraska–Lincoln)
21.2, 21.8, 31.0, 34.4, 23.3, 13.0, 32.3, 38.0, 47.5, 21.5, 18.9, 15.7, 13.0, 19.1, 18.7, 25.8, 23.8, 32.1, 21.3, 21.8, 30.7, 29.0, 44.6, 24.4, 11.9, 37.9, 29.5, 31.7, 35.9, 16.3