Describing Data: Displaying and Exploring Data

Chapter 4

Learning Objectives

- Develop and interpret a dot plot.
- Develop and interpret a stem-and-leaf display.
- Compute and understand quartiles.
- Construct and interpret box plots.
- Draw and interpret a scatter diagram.
- Construct and interpret a contingency table.

Dot Plot

- A **dot plot** groups the data as little as possible and the identity of an individual observation is not lost.
- To develop a dot plot, each observation is simply displayed as a dot along a horizontal number line indicating the possible values of the data.
- If there are identical observations or the observations are too close to be shown individually, the dots are "piled" on top of each other.

Dot Plot

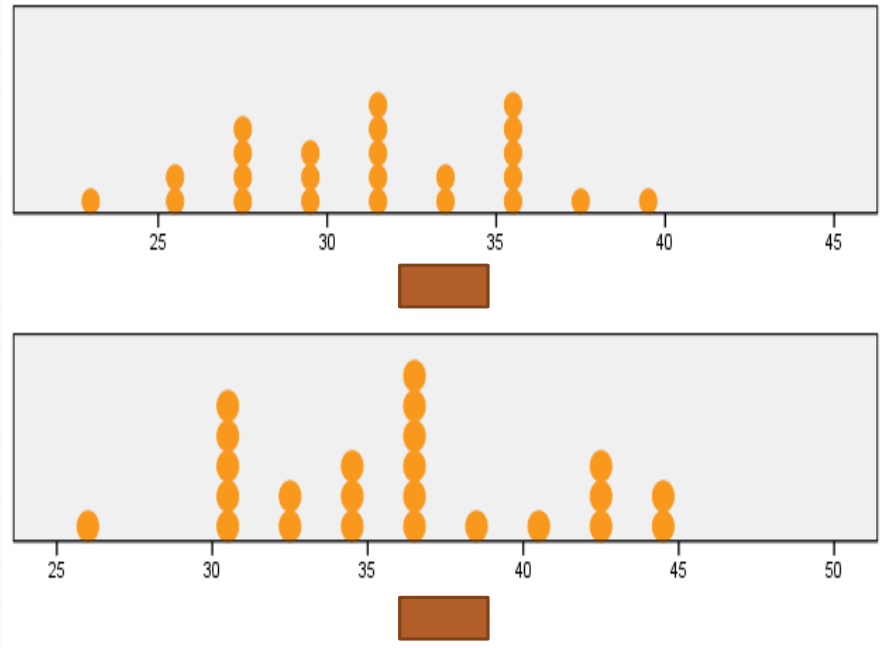
Example 1: Recall "Whitner Autoplex" from chapter 2, Develop a dot plot for the selling prices.

Dot Plot

Example:
Reported below are the number of vehicles sold in the last 24 months at Smith Ford Mercury Jeep, Inc., in Kane, Pennsylvania, and Brophy Honda Volkswagen in Greenville, Ohio. Construct dot plots and report summary statistics for the two small-town Auto USA lots.

Smith Ford Mercury Jeep, Inc.									
23	27	30	27	32	31	32	32	35	33
28	39	32	29	35	36	33	25	35	37
26	28	36	30						

Brophy Honda Volkswagen									
31	44	30	36	37	34	43	38	37	35
36	34	31	32	40	36	31	44	26	30
37	43	42	33						



Stem-and-Leaf

- In Chapter 2, we showed how to organize data into a frequency distribution. The major advantage to organizing the data into a frequency distribution is that we get a quick visual picture of the shape of the distribution.

- One technique that is used to display quantitative information in a condensed form is the **stem-and-leaf display.**

Stem-and-Leaf

- **Stem-and-leaf display** is a statistical technique to present a set of data. Each numerical value is divided into two parts. The leading digit(s) becomes the stem and the trailing digit the leaf. The stems are located along the vertical axis, and the leaf values are stacked against each other along the horizontal axis.

- Advantage of the stem-and-leaf display over a frequency distribution - the identity of each observation is not lost.

Stem-and-Leaf

- Suppose we have seven observations 96, 94, 93, 94, 95, 96, and 97.
- The stem value is the leading digit or digits, in this case 9. The leaves are the trailing digits. The stem is placed to the left of a vertical line and the leaf values to the right. The values in the 90 up to 100 class would appear as

- Then, we sort the values within each stem from smallest to largest. Thus, the second row of the stem-and-leaf display would appear as follows:

Stem-and-Leaf

- Example: Listed in Table 4–1 is the number of 30-second radio advertising spots purchased by each of the 45 members of the Greater Buffalo Automobile Dealers Association last year. Organize the data into a stem-and-leaf display. Around what values do the number of advertising spots tend to cluster? What is the fewest number of spots purchased by a dealer? The largest number purchased?

TABLE 4–1 Number of Advertising Spots Purchased by Members of the Greater Buffalo Automobile Dealers Association

96	93	88	117	127	95	113	96	108	94	148	156
139	142	94	107	125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139	134	119	97	89
118	136	125	143	120	103	113	124	138			

Stem-and-Leaf

Stem	Leaf
8	8 9
9	6 3 5 6 4 4 7
10	8 7 3 4 6 3
11	7 3 2 7 2 1 9 8 3
12	7 5 7 0 5 5 0 4
13	9 5 2 9 4 6 8
14	8 2 3
15	6 5 5

The Quartiles

- Quartiles are the three points that divide a set of observations into four equal parts.
- The first quartile is the value below which 25% of the observations occur and usually labeled as Q_1 . (the 25th percentile)
- The second quartile Q_2 is the "Median." (the 50th percentile)
- The third quartile is the value below which 75% of the observations occur and usually labeled as Q_3 . (the 75th percentile)

The Quartiles

LOCATION OF A PERCENTILE

$$L_p = (n + 1)\frac{P}{100} \qquad [4\text{–}1]$$

$$Q_1 = L_{25}$$
$$Q_2 = L_{50}$$
$$Q_3 = L_{75}$$

The Quartiles

Example: Listed below are the commissions earned last month by a sample of 15 brokers at Salomon Smith Barney's Oakland, California, office. Salomon Smith Barney is an investment company with offices located throughout the United States.

$2,038	$1,758	$1,721	$1,637
$2,097	$2,047	$2,205	$1,787
$2,287	$1,940	$2,311	$2,054
$2,406	$1,471	$1,460	

Locate the median, the first quartile, and the third quartile for the commissions earned.

The Quartiles

Step 1: Organize the data from lowest to largest value

$1,460	$1,471	$1,637	$1,721
$1,758	$1,787	$1,940	$2,038
$2,047	$2,054	$2,097	$2,205
$2,287	$2,311	$2,406	

The Quartiles

Step 2: Compute the first and third quartiles. Locate L_{25} and L_{75} using:

$$L_{25} = (15+1)\frac{25}{100} = 4 \qquad L_{75} = (15+1)\frac{75}{100} = 12$$

Therefore, the first and third quartiles are the 4th and 12th observation in the array, respectively

$L_{25} = \$1{,}721$

$L_{75} = \$2{,}205$

The Box Plot

- A box plot is a graphical display, based on the quartiles, that help us picture a set of data. To construct a box plot, we need only five statistics: the minimum value, Q_1 (the first quartile), the median, Q_3 (the third quartile) and the maximum value.

The Box Plot

Alexander's Pizza offers free delivery of its pizza within 15 miles. Alex, the owner, wants some information on the time it takes for delivery. How long does a typical delivery take? Within what range of times will most deliveries be completed? For a sample of 20 deliveries, he determined the following information:

Minimum value = 13 minutes

Q_1 = 15 minutes

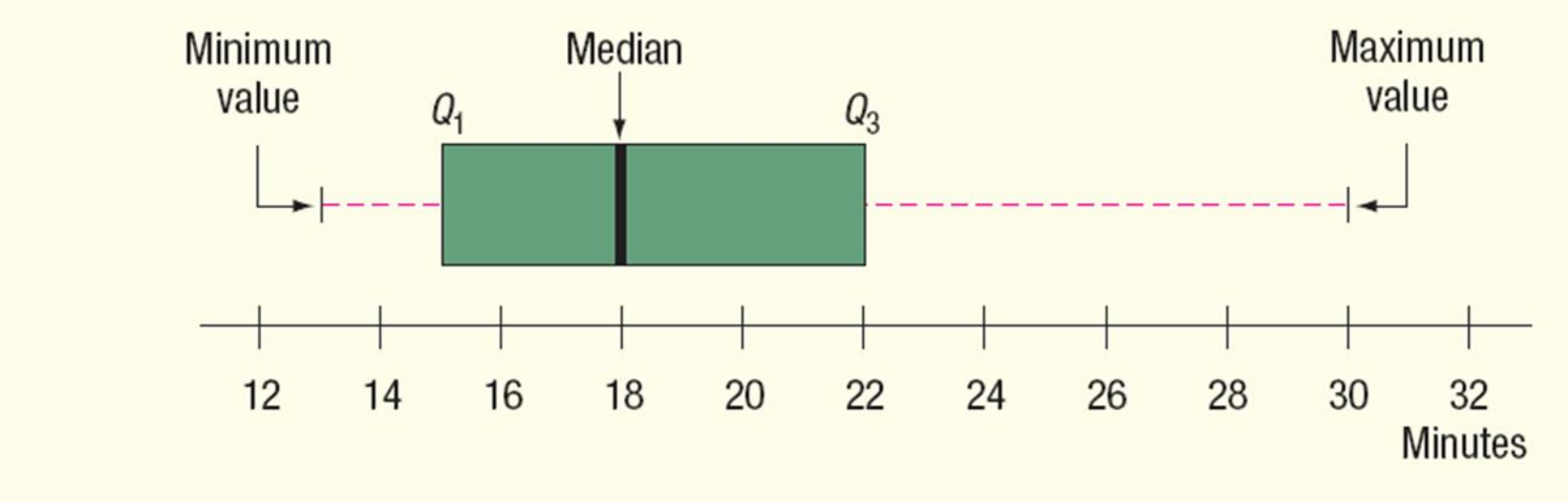
Median = 18 minutes

Q_3 = 22 minutes

Maximum value = 30 minutes

Develop a box plot for the delivery times. What conclusions can you make about the delivery times?

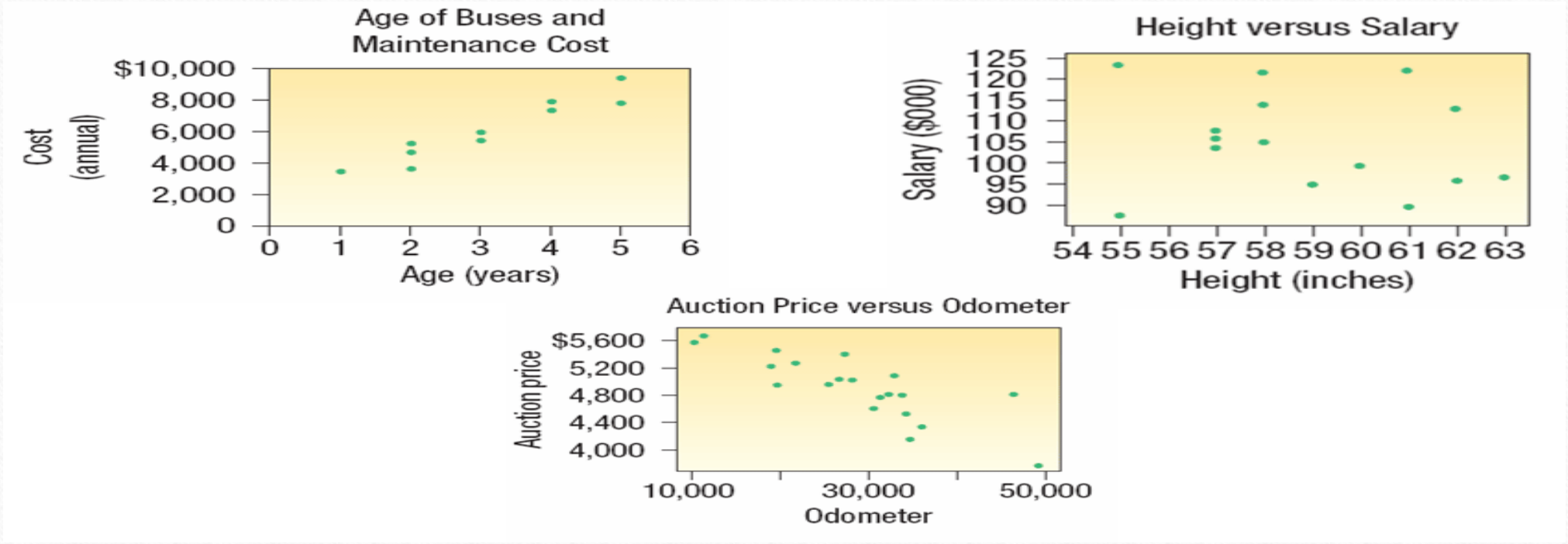
The Box Plot


Minimum value
Median
Maximum value
Q_1
Q_3
12
14
16
18
20
22
24
26
28
30
32
Minutes

The Scatter Diagram

- A **scatter diagram** is graphical technique to show the relationship between variables.
- To draw a scatter diagram we need two variables. We scale one variable along the horizontal axis (*X*-axis) of a graph and the other variable along the vertical axis (*Y*-axis).

The Scatter Diagram



The Contingency table

- A scatter diagram requires that both of the variables be at least interval scale.
- What if we wish to study the relationship between two variables when one or both are nominal or ordinal scale? In this case we tally the results in a contingency table.
- A **contingency table** is a table used to classify observations according to two identifiable characteristics.

The Contingency table

- Example: A manufacturer of preassembled windows produced 50 windows yesterday. This morning the quality assurance inspector reviewed each window for all quality aspects. Each was classified as acceptable or unacceptable and by the shift on which it was produced. Thus we reported two variables on a single item. The two variables are shift and quality. The results are reported in the following table.

	Shift			
	Day	Afternoon	Night	Total
Defective	3	2	1	6
Acceptable	17	13	14	44
Total	20	15	15	50