

Chapter

Descriptive Statistics

Chapter Outline

- 1 Frequency Distributions and Their Graphs
- 2 More Graphs and Displays
- 3 Measures of Central Tendency
- 4 Measures of Variation
- 5 Measures of Position

Section 1

Frequency Distributions and Their Graphs

Section 1 Objectives

- Construct frequency distributions
- Construct frequency histograms, frequency polygons, relative frequency histograms, and ogives

Frequency Distribution

Frequency Distribution

- A table that shows **classes** or **intervals** of data with a count of the number of entries in each class.
- The **frequency, f** , of a class is the number of data entries in the class.

Class width
 $6 - 1 = 5$ ←

Class	Frequency, f
1 – 5	5
6 – 10	8
11 – 15	6
16 – 20	8
21 – 25	5
26 – 30	4

Lower class
limits

Upper class
limits

Constructing a Frequency Distribution

1. Decide on the number of classes.
 - Usually between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width.
 - Determine the range of the data.
 - Divide the range by the number of classes.
 - *Round up to the next convenient number.*

Constructing a Frequency Distribution

3. Find the class limits.
 - You can use the minimum data entry as the lower limit of the first class.
 - Find the remaining lower limits (add the class width to the lower limit of the preceding class).
 - Find the upper limit of the first class. Remember that classes cannot overlap.
 - Find the remaining upper class limits.

Constructing a Frequency Distribution

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

Example: Constructing a Frequency Distribution

The following sample data set lists the number of minutes 50 Internet subscribers spent on the Internet during their most recent session. Construct a frequency distribution that has seven classes.

50 40 41 17 11 7 22 44 28 21 19 23 37 51 54 42 86
41 78 56 72 56 17 7 69 30 80 56 29 33 46 31 39 20
18 29 34 59 73 77 36 39 30 62 54 67 39 31 53 44

Solution: Constructing a Frequency Distribution

50 40 41 17 11 7 22 44 28 21 19 23 37 51 54 42 86
41 78 56 72 56 17 7 69 30 80 56 29 33 46 31 39 20
18 29 34 59 73 77 36 39 30 62 54 67 39 31 53 44

1. Number of classes = 7 (given)
2. Find the class width

$$\frac{\text{max} - \text{min}}{\text{\#classes}} = \frac{86 - 7}{7} \approx 11.29$$

Round up to 12

Solution: Constructing a Frequency Distribution

3. Use 7 (minimum value) as first lower limit. Add the class width of 12 to get the lower limit of the next class.

$$7 + 12 = 19$$

Find the remaining lower limits.

Class width = 12

Lower limit	Upper limit
7	
19	
31	
43	
55	
67	
79	

Solution: Constructing a Frequency Distribution

The upper limit of the first class is 18 (one less than the lower limit of the second class).

Add the class width of 12 to get the upper limit of the next class.

$$18 + 12 = 30$$

Find the remaining upper limits.

Lower limit	Upper limit
7	18
19	30
31	42
43	54
55	66
67	78
79	90

Class width = 12

Solution: Constructing a Frequency Distribution

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

Class	Tally	Frequency, f
7 – 18	I	6
19 – 30		10
31 – 42	III	13
43 – 54	III	8
55 – 66		5
67 – 78	I	6
79 – 90		2

$$\Sigma f = 50$$

Determining the Midpoint

Midpoint of a class

$$\frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

Class	Midpoint	Frequency, f
7 – 18	$\frac{7+18}{2} = 12.5$	6
19 – 30	$\frac{19+30}{2} = 24.5$	10
31 – 42	$\frac{31+42}{2} = 36.5$	13

Class width = 12

Determining the Relative Frequency

Relative Frequency of a class

- Portion or percentage of the data that falls in a particular class.

- relative frequency = $\frac{\text{class frequency}}{\text{Sample size}} = \frac{f}{n}$

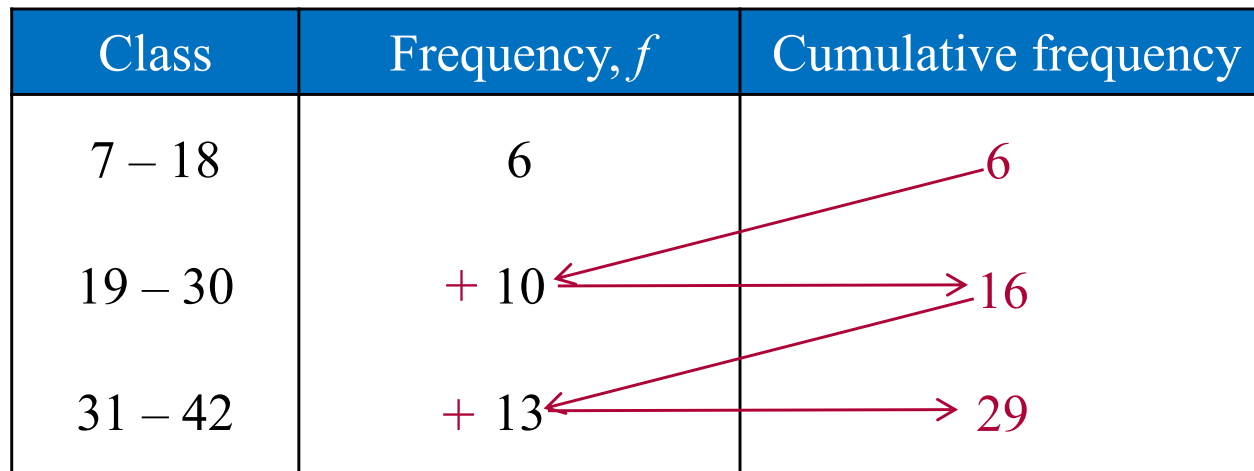
Class	Frequency, f	Relative Frequency
7 – 18	6	$\frac{6}{50} = 0.12$
19 – 30	10	$\frac{10}{50} = 0.20$
31 – 42	13	$\frac{13}{50} = 0.26$

Determining the Cumulative Frequency

Cumulative frequency of a class

- The sum of the frequency for that class and all previous classes.

Class	Frequency, f	Cumulative frequency
7 – 18	6	6
19 – 30	+ 10	16
31 – 42	+ 13	29



Expanded Frequency Distribution

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
7 – 18	6	12.5	0.12	6
19 – 30	10	24.5	0.20	16
31 – 42	13	36.5	0.26	29
43 – 54	8	48.5	0.16	37
55 – 66	5	60.5	0.10	42
67 – 78	6	72.5	0.12	48
79 – 90	2	84.5	0.04	50

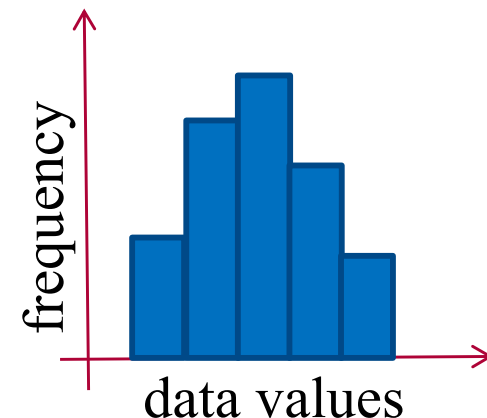
$$\Sigma f = 50$$

$$\Sigma \frac{f}{n} = 1$$

Graphs of Frequency Distributions

Frequency Histogram

- A bar graph that represents the frequency distribution.
- The horizontal scale is quantitative and measures the data values.
- The vertical scale measures the frequencies of the classes.
- Consecutive bars must touch.



Class Boundaries

Class boundaries

- The numbers that separate classes without forming gaps between them.
- The distance from the upper limit of the first class to the lower limit of the second class is $19 - 18 = 1$.
- Half this distance is 0.5.

Class	Class Boundaries	Frequency, f
7 – 18	6.5 – 18.5	6
19 – 30		10
31 – 42		13

- First class lower boundary = $7 - 0.5 = 6.5$
- First class upper boundary = $18 + 0.5 = 18.5$

Class Boundaries

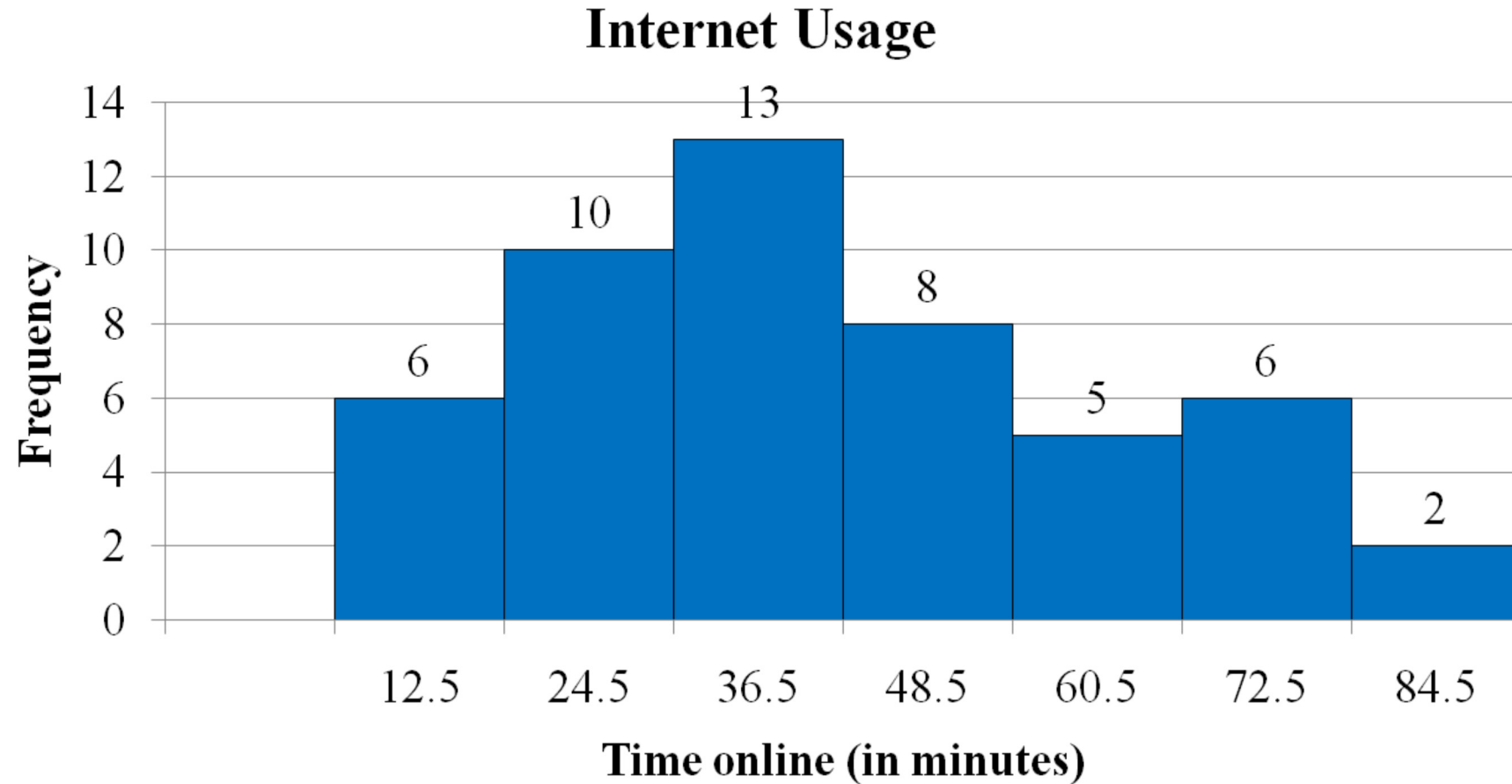
Class	Class boundaries	Frequency, f
7 – 18	6.5 – 18.5	6
19 – 30	18.5 – 30.5	10
31 – 42	30.5 – 42.5	13
43 – 54	42.5 – 54.5	8
55 – 66	54.5 – 66.5	5
67 – 78	66.5 – 78.5	6
79 – 90	78.5 – 90.5	2

Example: Frequency Histogram

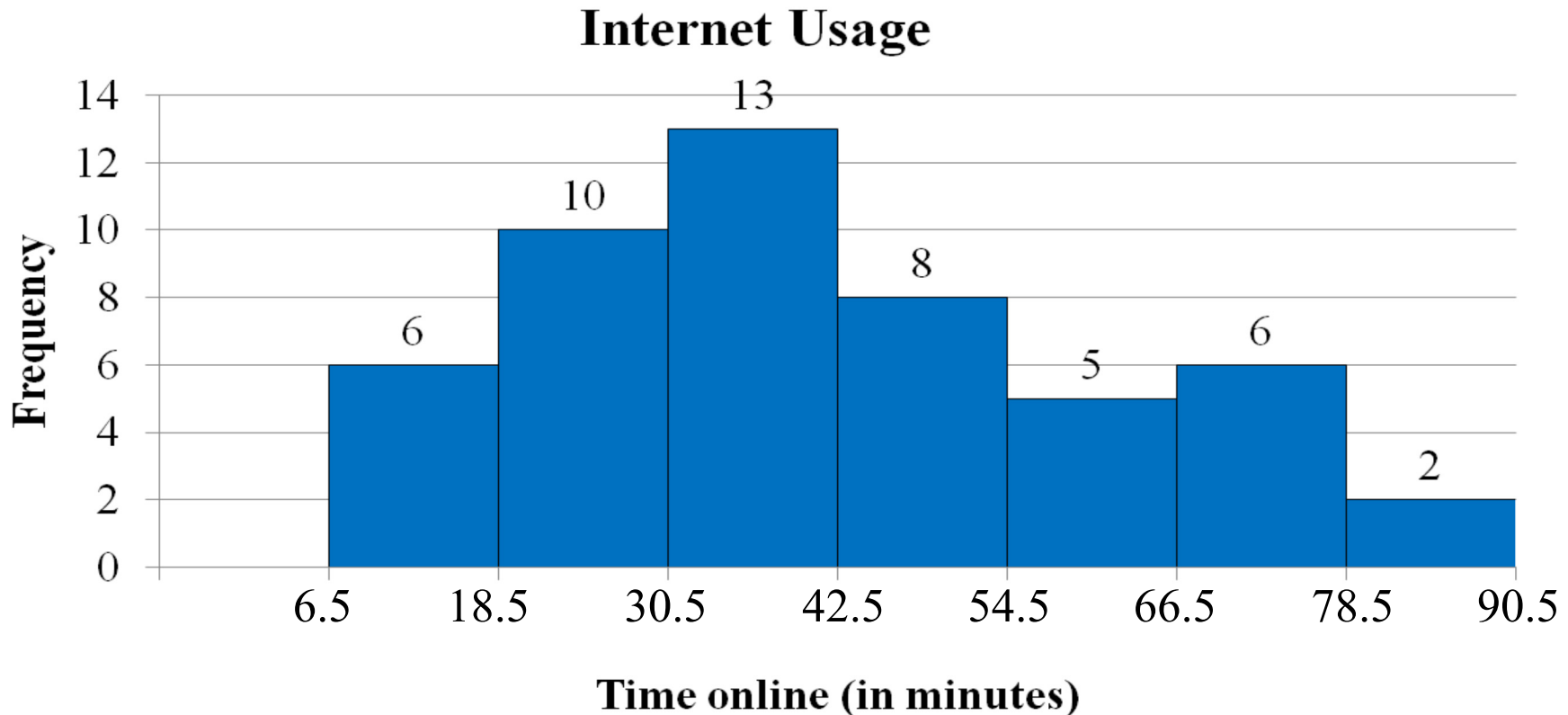
Construct a frequency histogram for the Internet usage frequency distribution.

Class	Class boundaries	Midpoint	Frequency, f
7 – 18	6.5 – 18.5	12.5	6
19 – 30	18.5 – 30.5	24.5	10
31 – 42	30.5 – 42.5	36.5	13
43 – 54	42.5 – 54.5	48.5	8
55 – 66	54.5 – 66.5	60.5	5
67 – 78	66.5 – 78.5	72.5	6
79 – 90	78.5 – 90.5	84.5	2

Solution: Frequency Histogram (using Midpoints)



Solution: Frequency Histogram (using class boundaries)

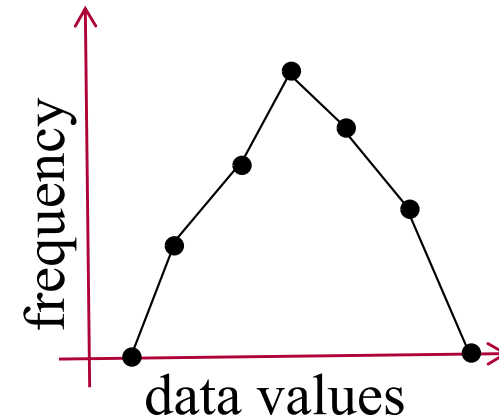


You can see that more than half of the subscribers spent between 19 and 54 minutes on the Internet during their most recent session.

Graphs of Frequency Distributions

Frequency Polygon

- A line graph that emphasizes the continuous change in frequencies.



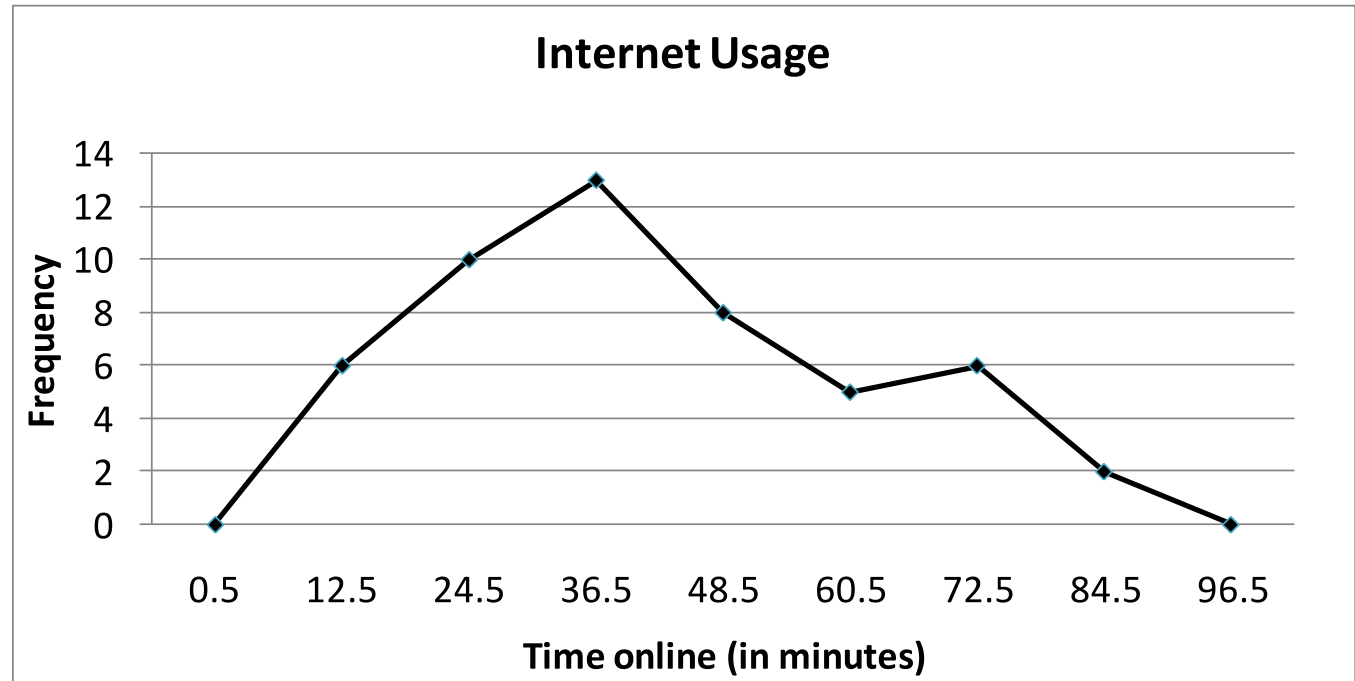
Example: Frequency Polygon

Construct a frequency polygon for the Internet usage frequency distribution.

Class	Midpoint	Frequency, f
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2

Solution: Frequency Polygon

The graph should begin and end on the horizontal axis, so extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.

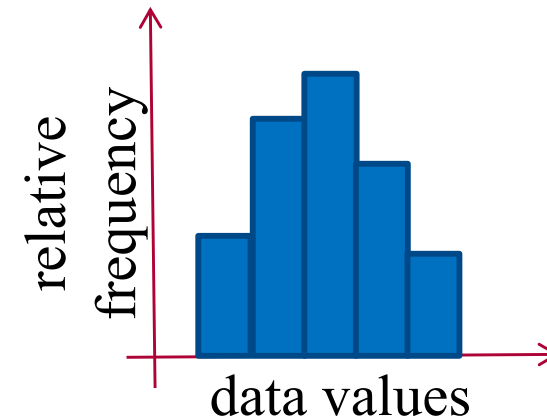


You can see that the frequency of subscribers increases up to 36.5 minutes and then decreases.

Graphs of Frequency Distributions

Relative Frequency Histogram

- Has the same shape and the same horizontal scale as the corresponding frequency histogram.
- The vertical scale measures the **relative frequencies**, not frequencies.

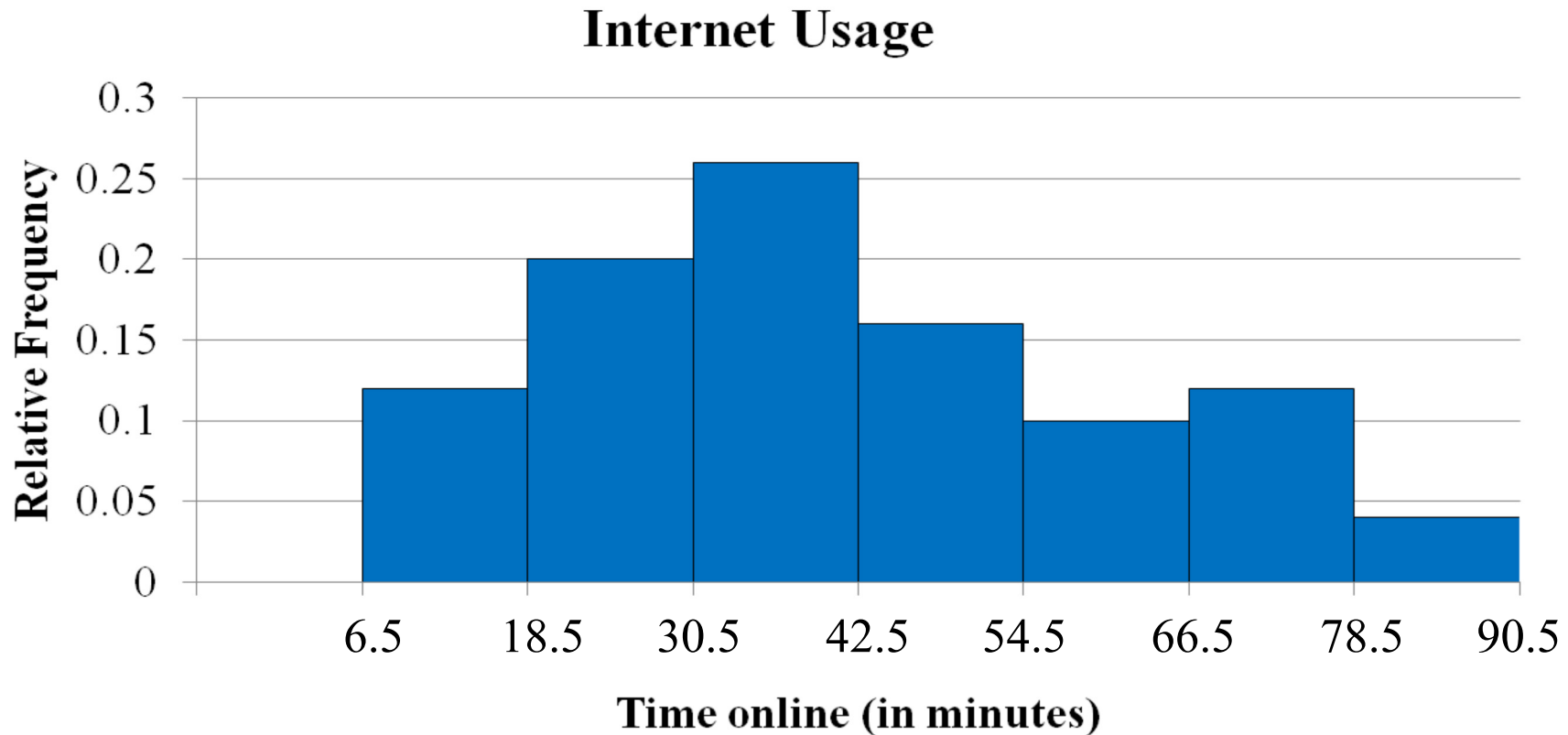


Example: Relative Frequency Histogram

Construct a relative frequency histogram for the Internet usage frequency distribution.

Class	Class boundaries	Frequency, f	Relative frequency
7 – 18	6.5 – 18.5	6	0.12
19 – 30	18.5 – 30.5	10	0.20
31 – 42	30.5 – 42.5	13	0.26
43 – 54	42.5 – 54.5	8	0.16
55 – 66	54.5 – 66.5	5	0.10
67 – 78	66.5 – 78.5	6	0.12
79 – 90	78.5 – 90.5	2	0.04

Solution: Relative Frequency Histogram

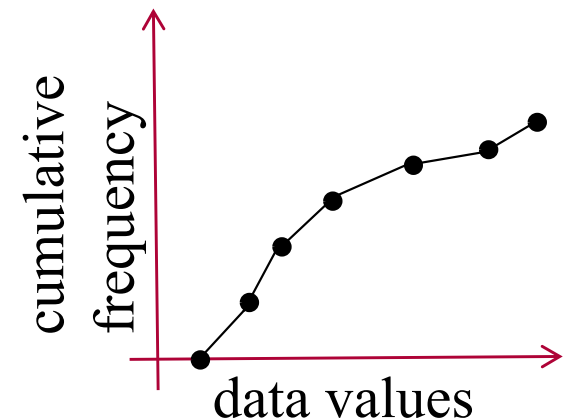


From this graph you can see that 20% of Internet subscribers spent between 18.5 minutes and 30.5 minutes online.

Graphs of Frequency Distributions

Cumulative Frequency Graph or Ogive

- A line graph that displays the cumulative frequency of each class at its upper class boundary.
- The upper boundaries are marked on the horizontal axis.
- The cumulative frequencies are marked on the vertical axis.



Constructing an Ogive

1. Construct a frequency distribution that includes cumulative frequencies as one of the columns.
2. Specify the horizontal and vertical scales.
 - The horizontal scale consists of the upper class boundaries.
 - The vertical scale measures cumulative frequencies.
3. Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.

Constructing an Ogive

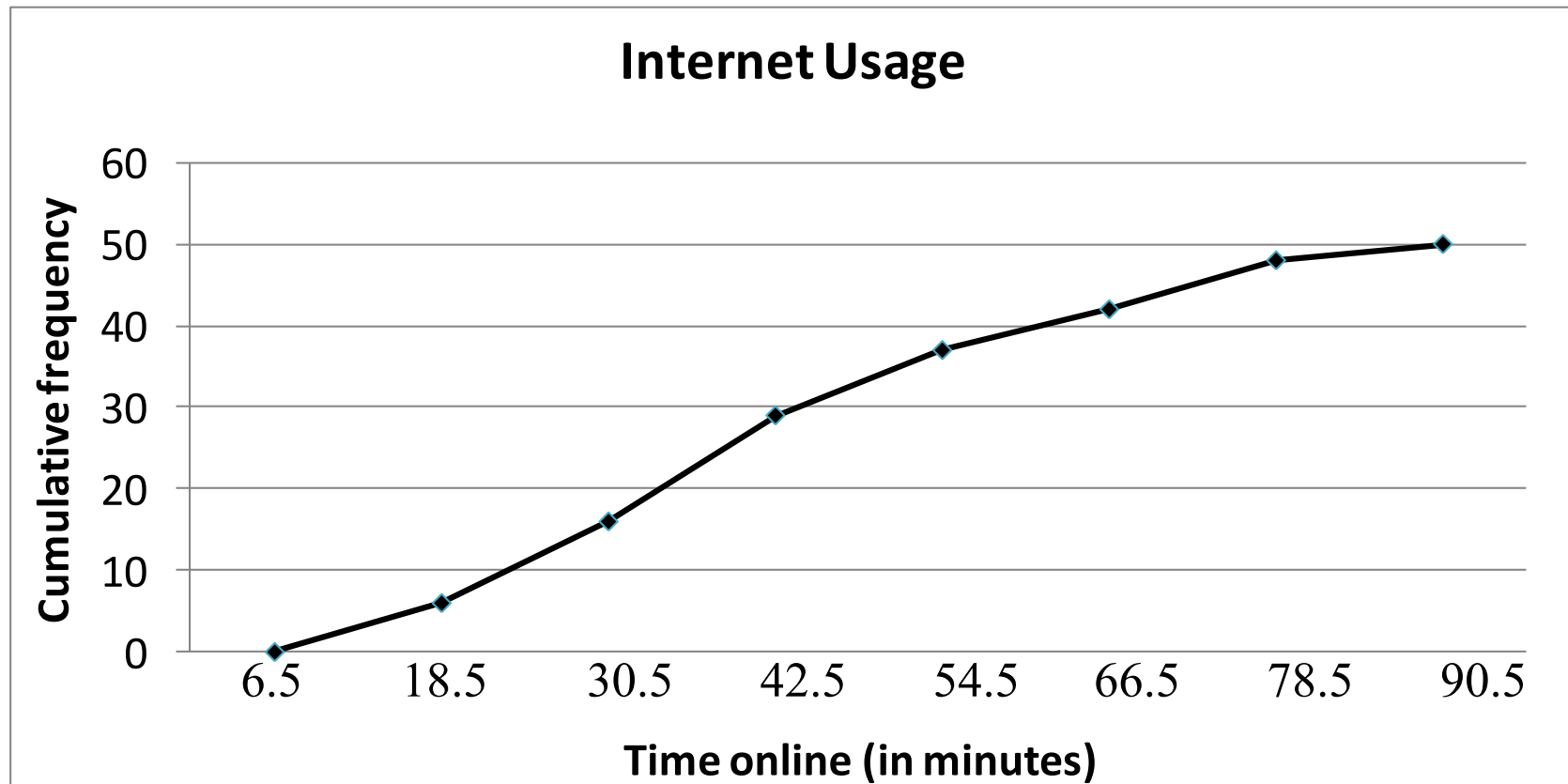
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).

Example: Ogive

Construct an ogive for the Internet usage frequency distribution.

Class	Class boundaries	Frequency, f	Cumulative frequency
7 – 18	6.5 – 18.5	6	6
19 – 30	18.5 – 30.5	10	16
31 – 42	30.5 – 42.5	13	29
43 – 54	42.5 – 54.5	8	37
55 – 66	54.5 – 66.5	5	42
67 – 78	66.5 – 78.5	6	48
79 – 90	78.5 – 90.5	2	50

Solution: Ogive



From the ogive, you can see that about 40 subscribers spent 60 minutes or less online during their last session. The greatest increase in usage occurs between 30.5 minutes and 42.5 minutes.

Section 1 Summary

- Constructed frequency distributions
- Constructed frequency histograms, frequency polygons, relative frequency histograms and ogives

Section 2

More Graphs and Displays

Section 2 Objectives

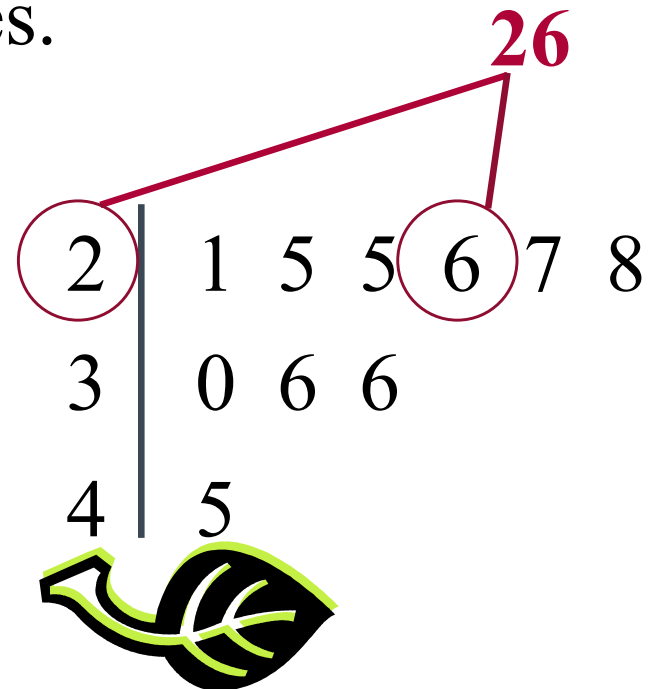
- Graph quantitative data using stem-and-leaf plots and dot plots
- Graph qualitative data using pie charts and Pareto charts
- Graph paired data sets using scatter plots and time series charts

Graphing Quantitative Data Sets

Stem-and-leaf plot

- Each number is separated into a **stem** and a **leaf**.
- Similar to a histogram.
- Still contains original data values.

Data: 21, 25, 25, **26**, 27, 28,
30, 36, 36, 45



Example: Constructing a Stem-and-Leaf Plot

The following are the numbers of text messages sent last month by the cellular phone users on one floor of a college dormitory. Display the data in a stem-and-leaf plot.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

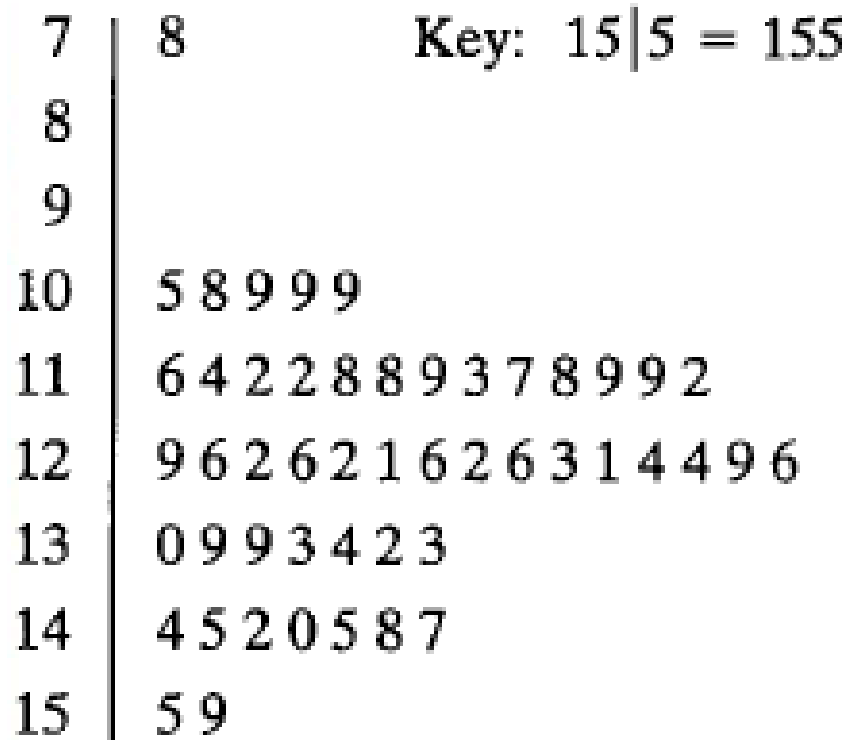
Solution: Constructing a Stem-and-Leaf Plot

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

- The data entries go from a low of 78 to a high of 159.
- Use the rightmost digit as the leaf.
 - For instance,
 $78 = 7 | 8$ and $159 = 15 | 9$
- List the stems, 7 to 15, to the left of a vertical line.
- For each data entry, list a leaf to the right of its stem.

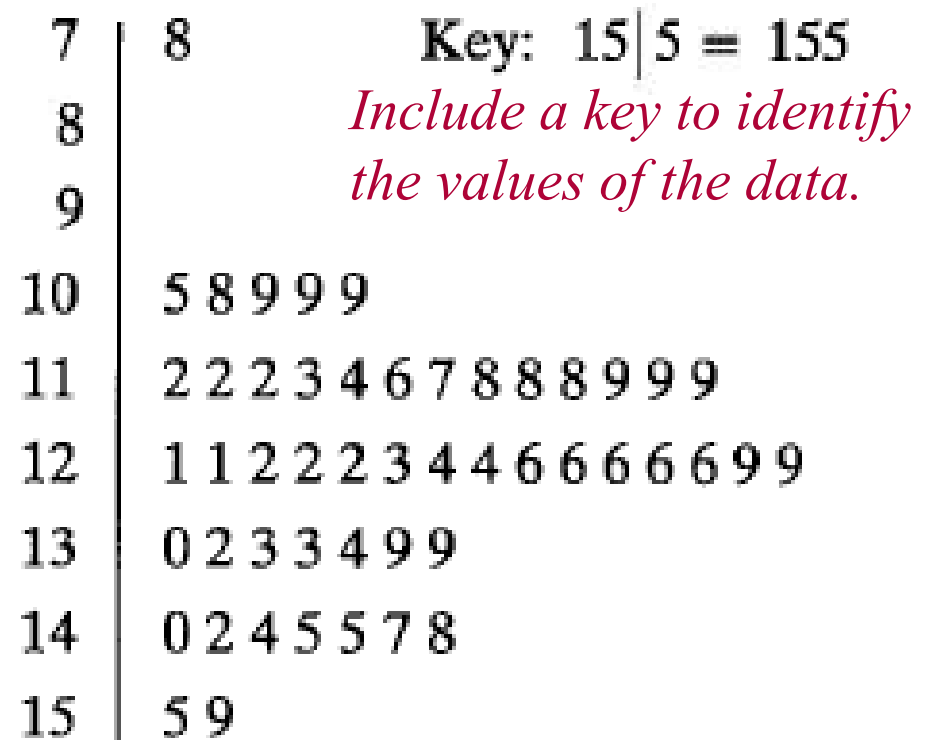
Solution: Constructing a Stem-and-Leaf Plot

Number of Text Messages Sent



Unordered Stem-and-Leaf Plot

Number of Text Messages Sent



Ordered Stem-and-Leaf Plot

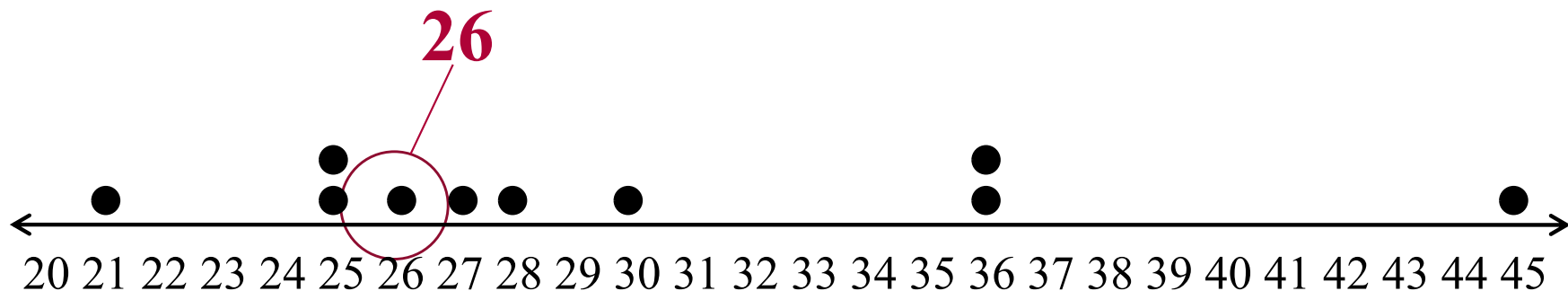
From the display, you can conclude that more than 50% of the cellular phone users sent between 110 and 130 text messages.

Graphing Quantitative Data Sets

Dot plot

- Each data entry is plotted, using a point, above a horizontal axis

Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



Example: Constructing a Dot Plot

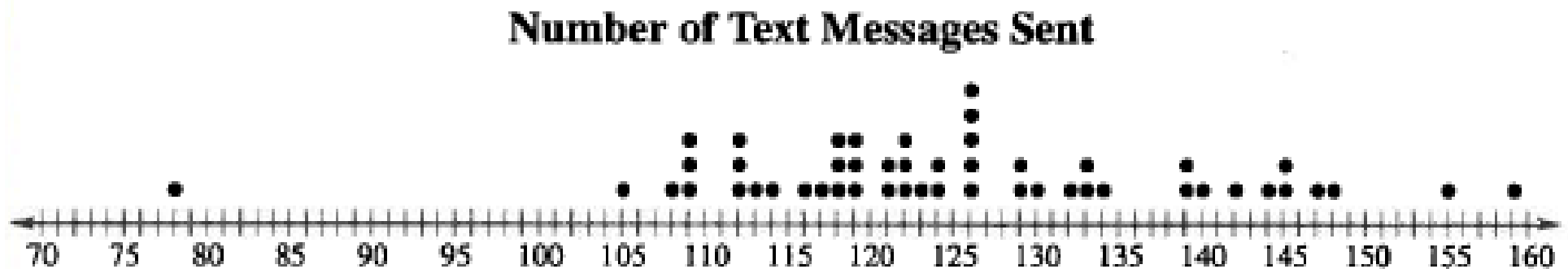
Use a dot plot to organize the text messaging data.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

- So that each data entry is included in the dot plot, the horizontal axis should include numbers between 70 and 160.
- To represent a data entry, plot a point above the entry's position on the axis.
- If an entry is repeated, plot another point above the previous point.

Solution: Constructing a Dot Plot

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

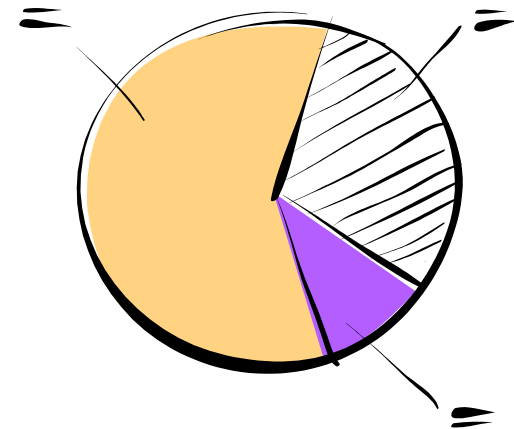


From the dot plot, you can see that most values cluster between 105 and 148 and the value that occurs the most is 126. You can also see that 78 is an unusual data value.

Graphing Qualitative Data Sets

Pie Chart

- A circle is divided into sectors that represent categories.
- The area of each sector is proportional to the frequency of each category.



Example: Constructing a Pie Chart

The numbers of motor vehicle occupants killed in crashes in 2005 are shown in the table. Use a pie chart to organize the data.

Vehicle type	Killed
Cars	18,440
Trucks	13,778
Motorcycles	4,553
Other	823

Solution: Constructing a Pie Chart

- Find the relative frequency (percent) of each category.

Vehicle type	Frequency, f	Relative frequency
Cars	18,440	$\frac{18440}{37594} \approx 0.49$
Trucks	13,778	$\frac{13778}{37594} \approx 0.37$
Motorcycles	4,553	$\frac{4553}{37594} \approx 0.12$
Other	823	$\frac{823}{37594} \approx 0.02$

37,594

Solution: Constructing a Pie Chart

- Construct the pie chart using the central angle that corresponds to each category.
 - To find the central angle, multiply 360° by the category's relative frequency.
 - For example, the central angle for cars is

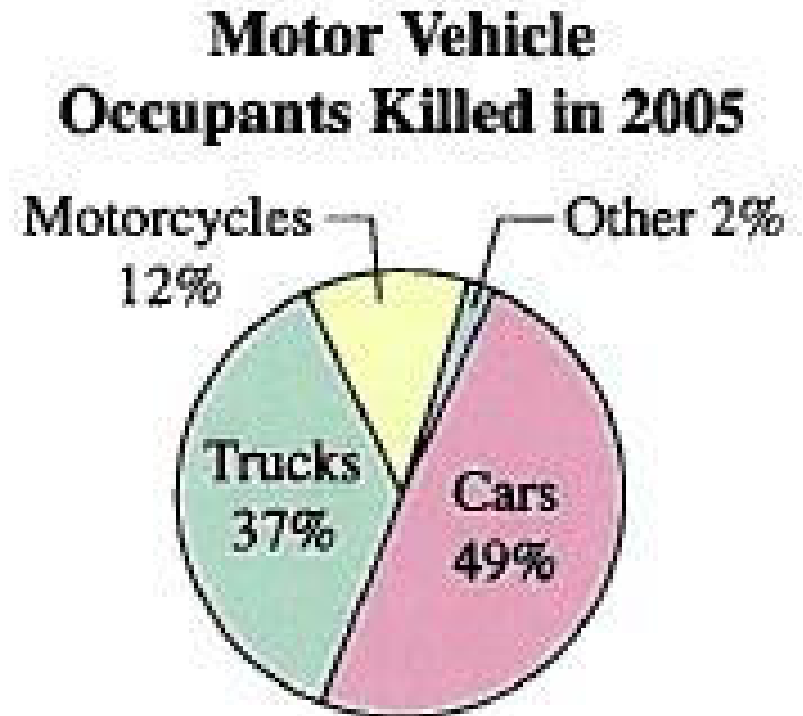
$$360(0.49) \approx 176^\circ$$

Solution: Constructing a Pie Chart

Vehicle type	Frequency, f	Relative frequency	Central angle
Cars	18,440	0.49	$360^\circ(0.49) \approx 176^\circ$
Trucks	13,778	0.37	$360^\circ(0.37) \approx 133^\circ$
Motorcycles	4,553	0.12	$360^\circ(0.12) \approx 43^\circ$
Other	823	0.02	$360^\circ(0.02) \approx 7^\circ$

Solution: Constructing a Pie Chart

Vehicle type	Relative frequency	Central angle
Cars	0.49	176°
Trucks	0.37	133°
Motorcycles	0.12	43°
Other	0.02	7°

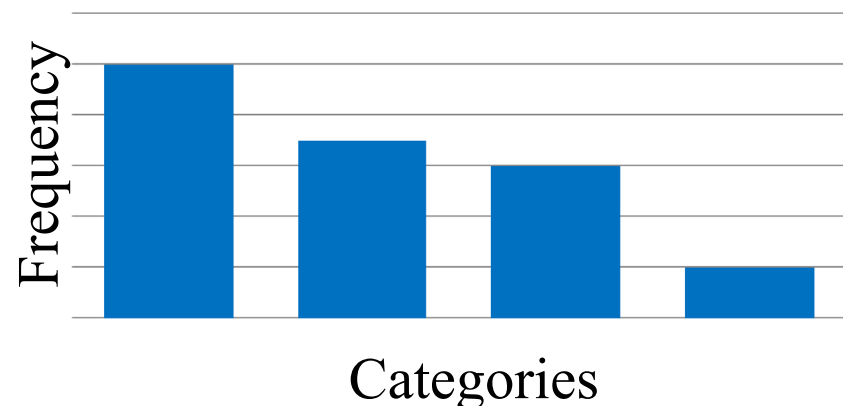


From the pie chart, you can see that most fatalities in motor vehicle crashes were those involving the occupants of cars.

Graphing Qualitative Data Sets

Pareto Chart

- A vertical bar graph in which the height of each bar represents frequency or relative frequency.
- The bars are positioned in order of decreasing height, with the tallest bar positioned at the left.

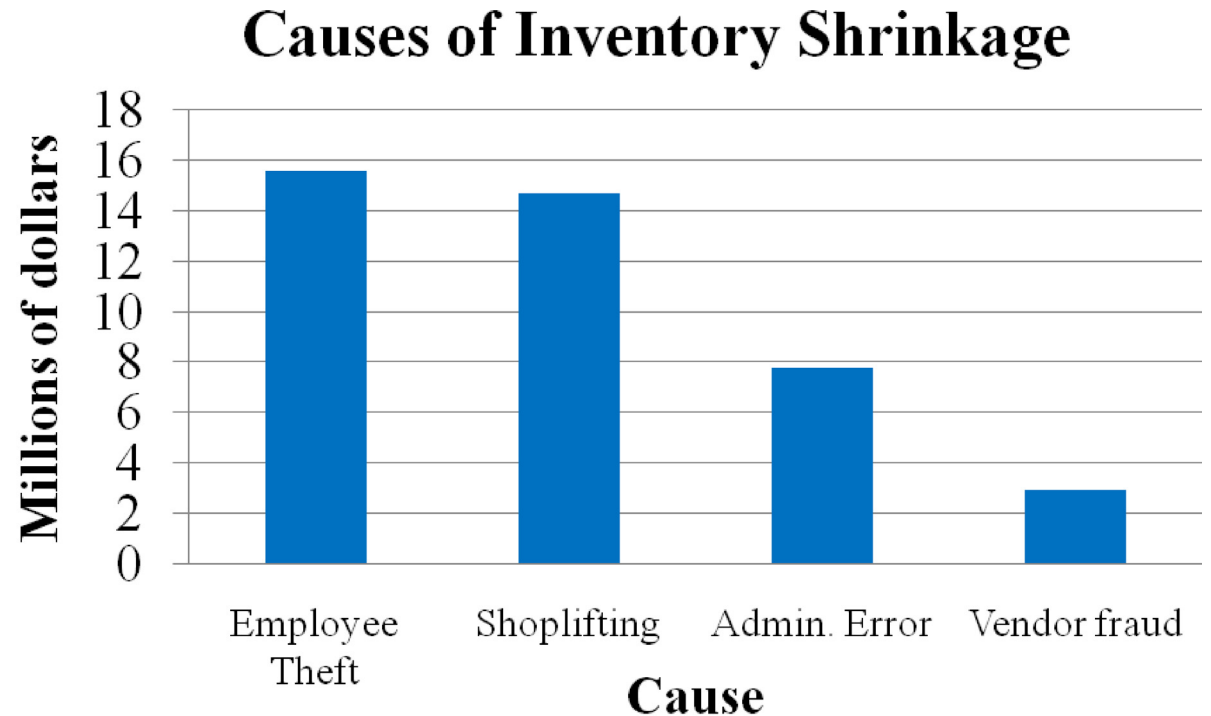


Example: Constructing a Pareto Chart

In a recent year, the retail industry lost \$41.0 million in inventory shrinkage. Inventory shrinkage is the loss of inventory through breakage, pilferage, shoplifting, and so on. The causes of the inventory shrinkage are administrative error (\$7.8 million), employee theft (\$15.6 million), shoplifting (\$14.7 million), and vendor fraud (\$2.9 million). Use a Pareto chart to organize this data.

Solution: Constructing a Pareto Chart

Cause	\$ (million)
Admin. error	7.8
Employee theft	15.6
Shoplifting	14.7
Vendor fraud	2.9

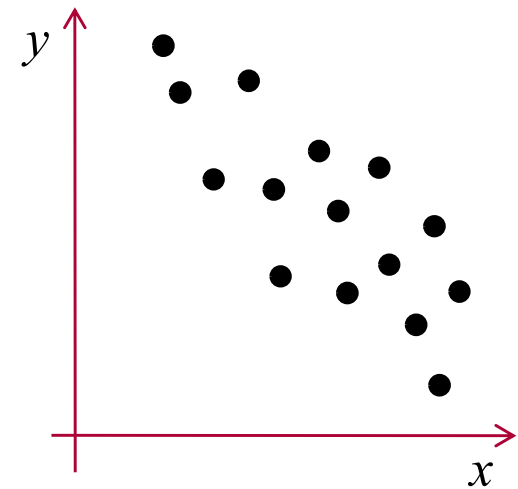


From the graph, it is easy to see that the causes of inventory shrinkage that should be addressed first are employee theft and shoplifting.

Graphing Paired Data Sets

Paired Data Sets

- Each entry in one data set corresponds to one entry in a second data set.
- Graph using a **scatter plot**.
 - The ordered pairs are graphed as points in a coordinate plane.
 - Used to show the relationship between two quantitative variables.



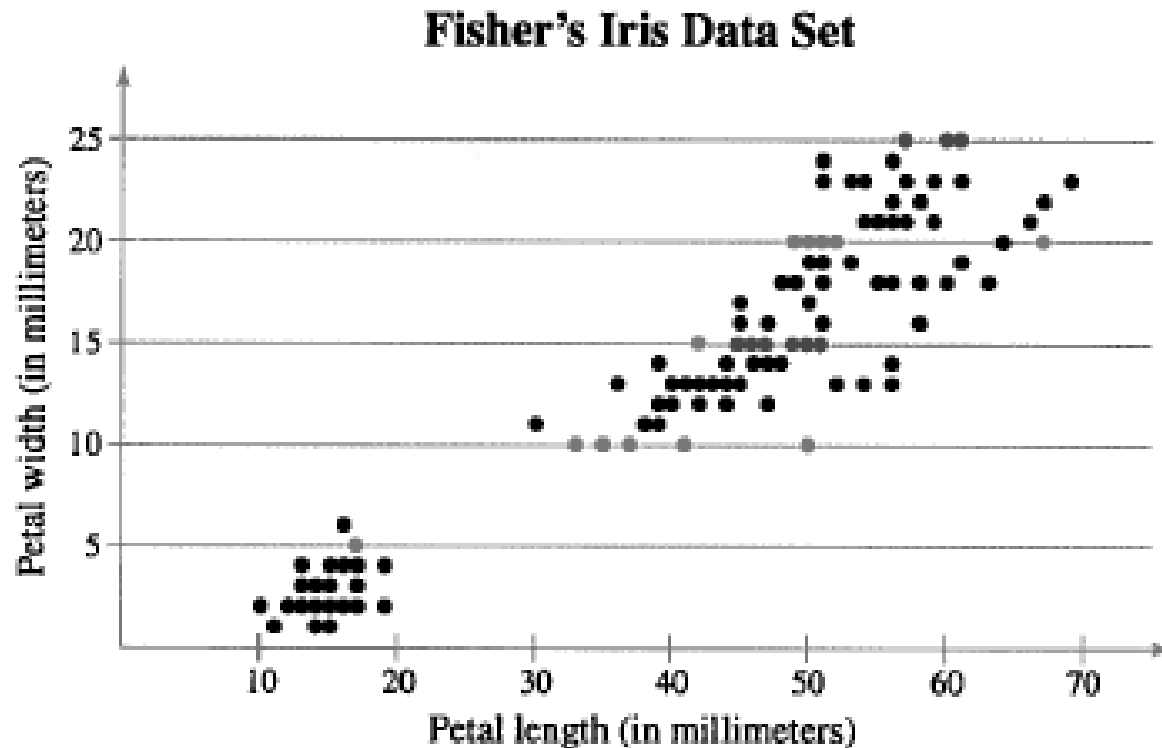
Example: Interpreting a Scatter Plot

The British statistician Ronald Fisher introduced a famous data set called Fisher's Iris data set. This data set describes various physical characteristics, such as petal length and petal width (in millimeters), for three species of iris. The petal lengths form the first data set and the petal widths form the second data set.



Example: Interpreting a Scatter Plot

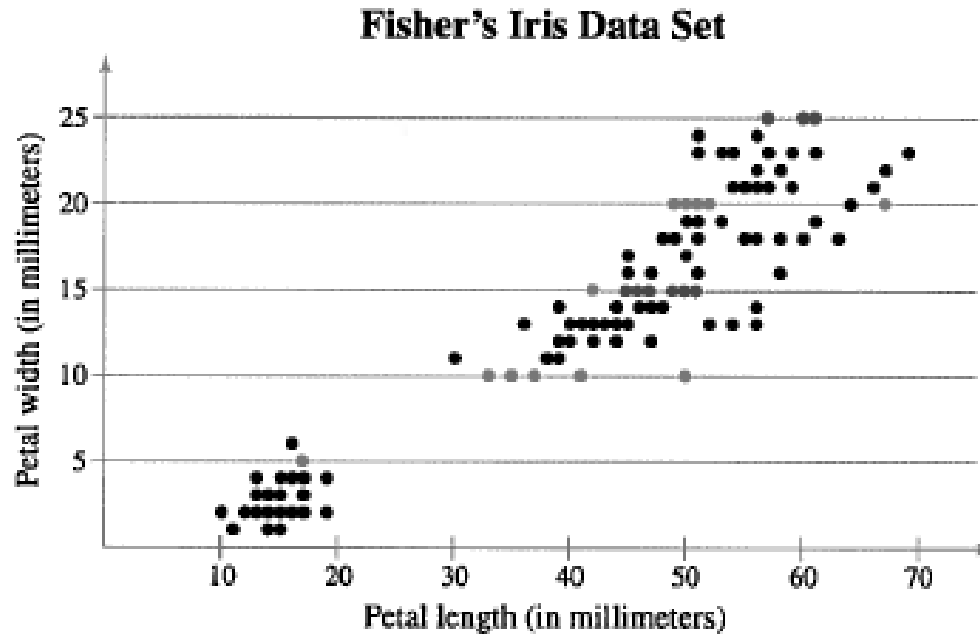
As the petal length increases, what tends to happen to the petal width?



Each point in the scatter plot represents the petal length and petal width of one flower.



Solution: Interpreting a Scatter Plot



Interpretation

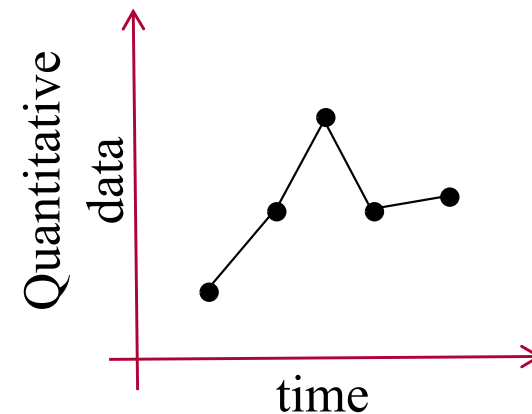
From the scatter plot, you can see that as the petal length increases, the petal width also tends to increase.



Graphing Paired Data Sets

Time Series

- Data set is composed of quantitative entries taken at regular intervals over a period of time.
 - e.g., The amount of precipitation measured each day for one month.
- Use a **time series chart** to graph.



Example: Constructing a Time Series Chart

The table lists the number of cellular telephone subscribers (in millions) for the years 1995 through 2005. Construct a time series chart for the number of cellular subscribers.

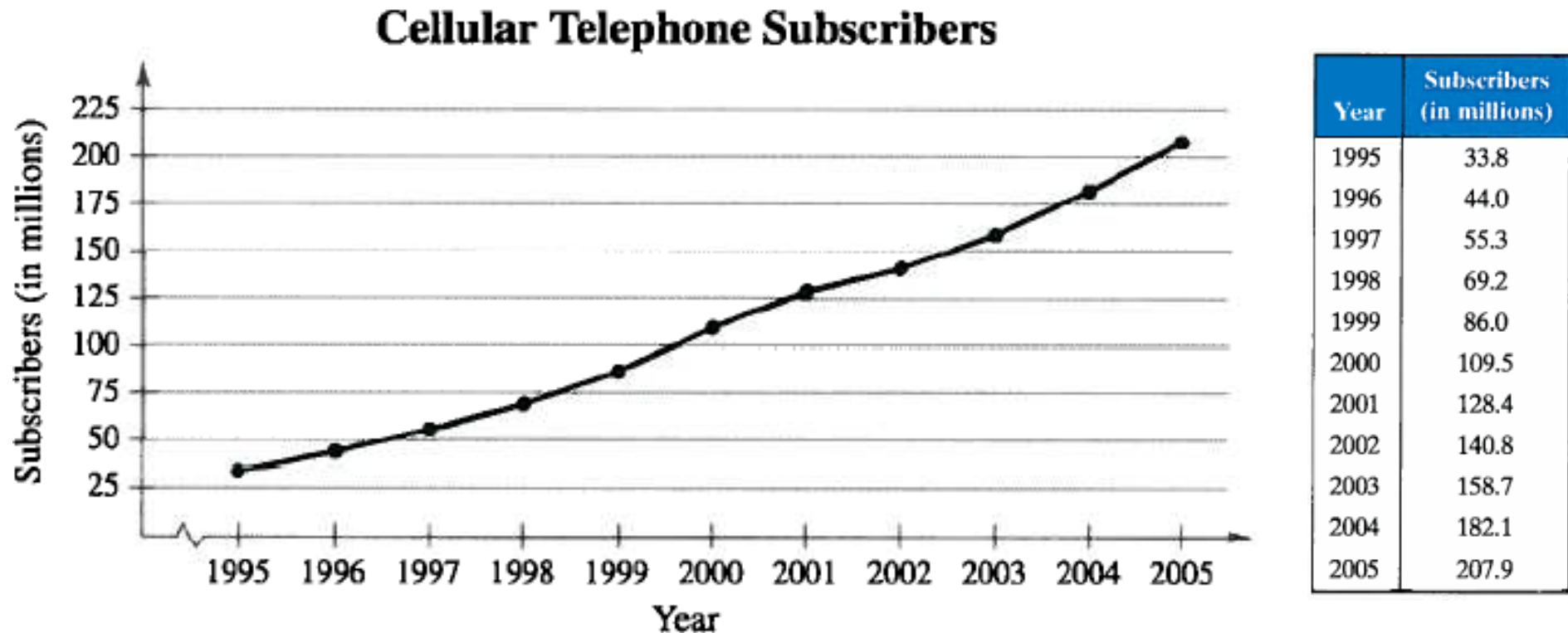
Year	Subscribers (in millions)
1995	33.8
1996	44.0
1997	55.3
1998	69.2
1999	86.0
2000	109.5
2001	128.4
2002	140.8
2003	158.7
2004	182.1
2005	207.9

Solution: Constructing a Time Series Chart

- Let the horizontal axis represent the years.
- Let the vertical axis represent the number of subscribers (in millions).
- Plot the paired data and connect them with line segments.

Year	Subscribers (in millions)
1995	33.8
1996	44.0
1997	55.3
1998	69.2
1999	86.0
2000	109.5
2001	128.4
2002	140.8
2003	158.7
2004	182.1
2005	207.9

Solution: Constructing a Time Series Chart



The graph shows that the number of subscribers has been increasing since 1995, with greater increases recently.

Section 2 Summary

- Graphed quantitative data using stem-and-leaf plots and dot plots
- Graphed qualitative data using pie charts and Pareto charts
- Graphed paired data sets using scatter plots and time series charts