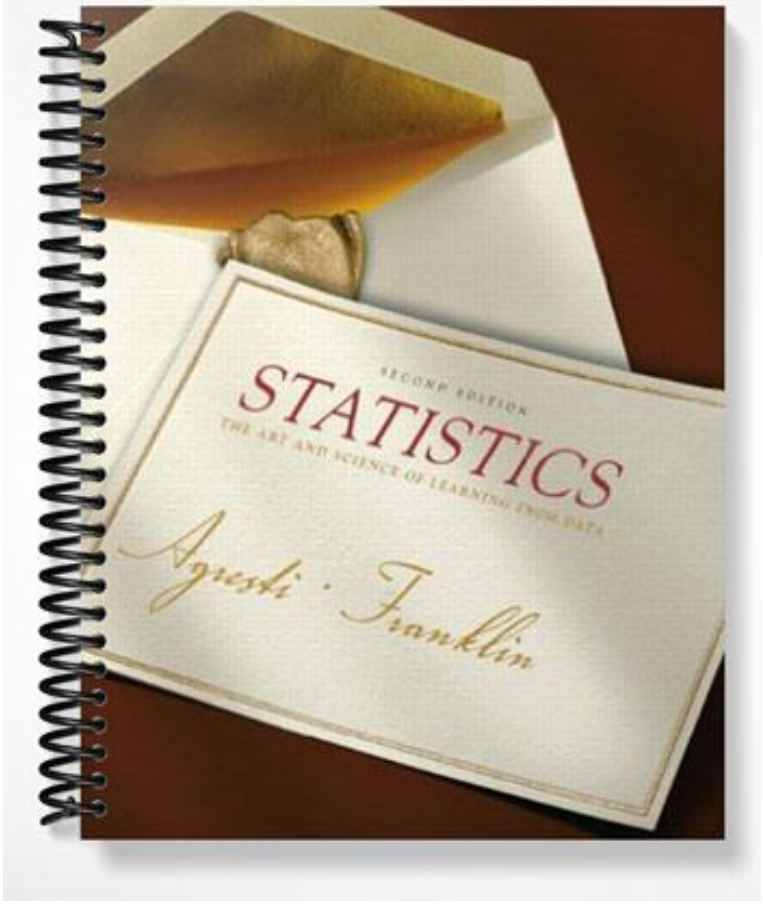
TEST BANK


SECOND EDITION
STATISTICS
THE ART AND SCIENCE OF LEARNING FROM DATA
Agresti · Franklin

Ch. 2 Exploring Data with Graphs and Numerical Summaries

2.1 What Are the Types of Data?

1 Categorical or Quantitative Variables

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Classify as categorical or qualitative data.

1) A survey of automobiles parked in the student and staff lots at a large college recorded the make and model of the automobiles. The variable "make" is:

2) The amount of time spent watching television or playing video games is considered a significant factor on predicting childhood obesity. 290 parents of school–aged children were asked to estimate the number of hours per week that their child spent watching television or playing video games. This is an example of what type of variable?

3) Your statistics teacher has gathered information on each of the students in your class in order to illustrate the difference between categorical and quantitative variables. For each student, she has recorded their major, gender, age and height. The variable "major" is an example of what type of variable?

4) Your statistics teacher has gathered information on each of the students in your class in order to illustrate the difference between categorical and quantitative variables. For each student, she has recorded their major, gender, age and height. The variable "age" is an example of what type of variable?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

5) A variable is called _____ if each observation belongs to one of a set of categories.

6) A variable is called _____ if observations on it take numerical values that represent different magnitudes of the variable.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Select the most appropriate answer.

7) The characteristics observed to address the questions posed in a study are called

2 Discrete or Continuous Variables

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Classify the variable as either discrete or continuous.

1) The time it takes an athlete to run 100 meters.

2) The number of calls received between 8 a.m. and 5 p.m. by a technical support professional.

3) The following table shows the heights of the five tallest mountains in North America.

Mountain	Height (ft)	Rank
McKinley	20,320	1
Logan	19,850	2
Citlaltepec	18,700	3
St. Elias	18,008	4
Popocatepetl	17,930	5

The ranks given in the third column represent what type of data?

4) The following table shows the heights of the five tallest mountains in North America.

Mountain	Height (ft)	Rank
McKinley	20,320	1
Logan	19,850	2
Citlaltepec	18,700	3
St. Elias	18,008	4
Popocatepetl	17,930	5

The heights given in the second column represent what type of data?

5) Your statistics teacher has gathered information on each of the students in your class in order to illustrate the difference between discrete and continuous variables. For each student, she has recorded their height, number of credit hours completed and the time it took for them to complete their last exam. The variable "height" is

6) Your statistics teacher has gathered information on each of the students in your class in order to illustrate the difference between discrete and continuous variables. For each student, she has recorded their height, number of credit hours completed and the time it took for them to complete their last exam. The variable "number of credit hours completed" is

Select the most appropriate answer.

7) Which of the following is a continuous variable?

8) Which of the following is a discrete variable?

3 Interpret Frequency Table

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Answer true or false.

1) The frequency for a particular category is the proportion of observations that fall in the category.

2) A frequency table is a listing of possible values for a variable, together with their frequencies and/or relative frequencies.

The heights (in inches) of 30 adult males are listed below. A frequency distribution show the frequency and relative frequency using five classes.

70 72 71 70 69 73 69 68 70 71
67 71 70 74 69 68 71 71 71 72
69 71 68 67 73 74 70 71 69 68

Height (in inches)	Frequency	Relative Frequency
67.0–68.4	6	0.20
68.5–69.9	5	0.167
70.0–71.4	13	0.433
71.5–72.9	2	0.067
73.0–74.4	4	0.133

3) Identify the variable.

4) Is the variable "height" continuous or discrete?

5) A height of 69 inches belongs to the class having what frequency?

6) What percentage of the 30 adult males had heights between 73 and 74.4 inches?

7) What proportion of the 30 adult males had heights less than 70 inches?

8) Which category of heights represents the mode?

Provide an appropriate response.

9) A safety engineer wishes to use the following data to show the number of deaths in a year from the collision of passenger cars with trucks on a particular highway.

Year	Number of Deaths
1995	12
1996	17
1997	22
1998	21
1999	16
2000	13
2001	11
2002	12

What is the mode of the number of deaths?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

10) A stock broker has been following different stocks over the last month and has recorded whether the various stock values are up, unchanged, or down at the end of the month. The results were

Stock performance	up	same	down
Count	21	7	12

a. What is the variable of interest?
b. Is the variable categorical or quantitative?
c. Which response is the mode?
d. Add proportions to this frequency table.

11) A local school district wants to know the number of children under the age of five living in the district in order to predict future enrollment. Households were randomly sampled in the district, and the head of household was asked to disclose the number of children under the age of five living in the household. The results were

Number of children under five	0	1	2	3	4
Count	15	18	12	12	3

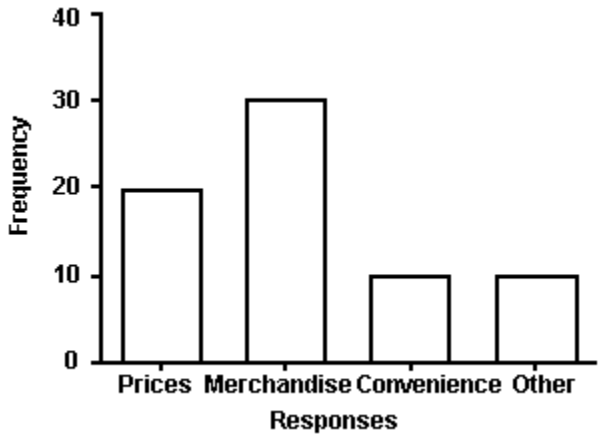
a. What is the variable of interest?
b. Is the variable categorical or quantitative?
c. Which response is the mode?
d. Add proportions to this frequency table.

2.2 How Can We Describe Data Using Graphical Summaries?

1 Construct/Interpret Bar Chart

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

A sporting goods retailer conducted a customer survey to determine its customers primary reason for shopping at their store. The results are shown in the graph below.



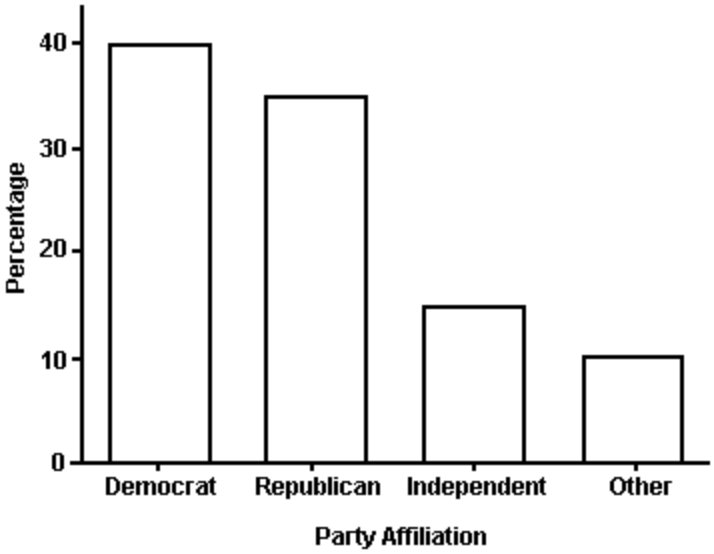
1) What proportion of the customers responded that the merchandise was the reason they shopped at the store?

2) What response represents the mode?

3) Is the variable "reason for shopping at our store" categorical or quantitative?

4) What percentage of the customers gave "prices" or "convenience" as their answer?

The bar graph below shows the political party affiliation of 1000 registered U.S. voters.



5) What percentage of the 1000 registered U.S. voters belongs to one of the two traditional parties (Democratic and Republican)?

6) About how many of the registered U.S. voters stated "Independent" as their political party affiliation?

7) Which response represents the mode?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

8) During the Fall Semester 2004 at the University of Georgia, the enrollment can be summarized as follows.

Enrollment	Count
Undergraduate	24,814
Graduate/Professional	8386
Independent Study	205

Source: http://www.uga.edu/profile/facts.html

a. Construct a bar graph for these data.
b. Would a dot plot or a stem–and–leaf plot make sense for these data? Explain.

2 Construct/Interpret Pie Chart

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

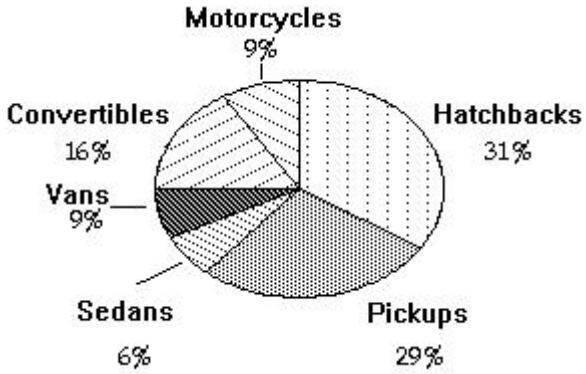
1) Parking at a large university has become a major issue. University administrators would like to determine the average time it takes a student to find a parking spot in a university lot. Students who are willing to participate in the study were asked to record the time between entering campus and pulling into a parking spot. Which of the following would not be appropriate for displaying the parking time data?

2) Each year advertisers spend billions of dollars purchasing commercial time on network sports television. A recent article listed the top 10 leading spenders (in millions of dollars) over a 6 month period:

Company A	$72.0	Company F	$26.9
Company B	63.1	Company G	25.0
Company C	54.7	Company H	23.9
Company D	54.3	Company I	23.0
Company E	29.0	Company J	20.0

Which of the following graphs would not be appropriate for displaying this data?

3) Results from a survey of 7116 vehicle types on the campus of State College are summarized in the following pie chart.

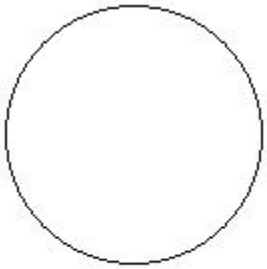


How many of the vehicles were sedans? Give your answer to the nearest whole number.

Construct a pie chart illustrating the given data set.

4) After reviewing a movie, 900 people rated the movie as excellent, good, or fair. The following data give the rating distribution.

Excellent	Good	Fair
180	450	270



SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

5) A sample of recent car buyers was asked to identify what they considered to be the most useful source of information about the cars they purchased. The results follow.

Source	Count
Consumer guide	172
Dealership	93
Word of mouth	40
Internet	26

Source: Automotive Retailing Today, The Gallup Organization.

a. Construct a pie chart for these data.
b. In creating a bar graph of these data, would it be more useful to list the sources of consumer information in the same order in which they appear in the table above or in the form of a Pareto chart?

6) The article "Physicians Accessing the Internet: The PAI project" contained the following information on Internet usage given by 331 randomly selected doctors who were asked to indicate the category that best described how often they used the Internet.

Internet Usage Pattern	Count
Never	31
Rarely (about 3 times per year)	15
Occasionally (about once a month)	52
Often (about once a week)	109
Daily	117

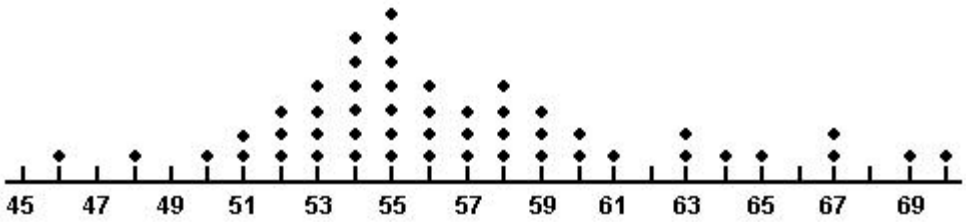
Source: Journal of the American Medical Association (1999): 633–634.

a. Construct a pie chart for these data.
b. In creating a bar graph of these data, would it be more useful to list the patterns as given in the table above or in the order of a Pareto chart?

3 Construct/Interpret Dot Plot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

A sample of fifty motorists was taken on a Federal highway where the speed limit was 60 miles per hour. A dot plot of their speeds is shown below.



1) What proportion of the motorists were speeding?

2) What is the mode for speed?

3) Would a pie chart be appropriate for displaying this data?

4) What is the variable of interest?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

5) The Highway Patrol, using radar, clocked the speeds (in mph) of 30 passing motorists at a checkpoint. The results are listed below. Construct a dot diagram for the data.

44	38	41	50	36	36	43	42	49	48
35	40	37	41	43	50	45	45	39	38
50	41	47	36	35	40	42	43	48	33

6) The following data represent the number of grams of fat in various breakfast foods at McDonald's.

Egg McMuffin®	12
Sausage McMuffin®	22
Sausage McMuffin® with Egg	27
English Muffin	3
Bacon, Egg & Cheese Biscuit (Regular Size Biscuit)	25
Bacon, Egg & Cheese Biscuit (Large Size Biscuit)	30
Sausage Biscuit with Egg (Regular Size Biscuit)	32
Sausage Biscuit with Egg (Large Size Biscuit)	37
Sausage Biscuit (Regular Size Biscuit)	27
Sausage Biscuit (Large Size Biscuit)	31
Biscuit (Regular Size)	11
Biscuit (Large Size)	16
Bacon, Egg & Cheese McGriddles®	21
Sausage, Egg & Cheese McGriddles®	32
Sausage McGriddles®	2
	2

Source: McDonald's Corporation

Construct a dot plot for these data.

7) The article "Tobacco and Alcohol Use in G–Rated Children's Animated Films" investigated exposure to tobacco and alcohol use in all G–rated animated films released between 1937 and 1997 by five major film studios. Data on the total tobacco exposure time (in seconds) for films with tobacco use produced by Walt Disney, Inc., follow.

223	176	548	37	158	51	299	37	11
165	74	9	2	6	23	206	9	

Source: Journal of the American Medical Association (1999): 1131–1136.

Construct a dot plot for these data. Comment on the shape of the distribution.

8) In order to reduce pollutants from motor vehicle exhaust emissions, three–way catalytic converters have been installed in new vehicles. However, these converters increase the level of ammonia in the air. A study was published on the ammonia levels near the exit ramp of a San Francisco highway tunnel. The data below represent daily ammonia concentrations (parts per million) on eight randomly selected days during afternoon drive–time in the summer of 1999.

1.53	1.50	1.37	1.51	1.55	1.42	1.41	1.48

Source: Environmental Science & Technology (Sept. 1, 2000)

Construct a dot plot for these data.

4 Construct/Interpret Stem–and–Leaf Plot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

A survey was conducted to determine how people rated the quality of programming available on television. Respondents were asked to rate the overall quality from 0 (no quality at all) to 100 (extremely good quality). The stem–and–leaf display of the data is shown below.

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

1) What percentage of the respondents rated overall television quality as very good (regarded as ratings of 80 and above)?

2) What is the mode rating?

3) The variable "quality" is

4) Identify the minimum quality rating.

5) Identify the maximum quality rating.

Find the original data from the stem–and–leaf plot.

6)

Stem	Leaves
8	5 8
9	1 8
10	5 5

The following data show the number of laps run by each participant in a timed running race:

46 65 55 43 51 48 57 30 43 49 32 56

7) If the stems are 3, 4, 5 and 6, how many leaves are on the "4 stem"?

8) If the stems are 3, 4, 5 and 6,what are the values of the leaves are on the "4 stem"?

9) Is the variable "number of laps run" discrete or continuous?

10) What is the mode for number of laps run?

Provide an appropriate response.

11) Twenty–four workers were surveyed and asked how long it takes them to travel to work each day. The data below are given in minutes.

20 35 42 52 65 20 60 49 24 37 23 24 22 20 41 25 28 27 50 47 58 30 32 48

Which of the following shows the data in a stem–and–leaf plot?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

12) The scores for a statistics test are as follows:

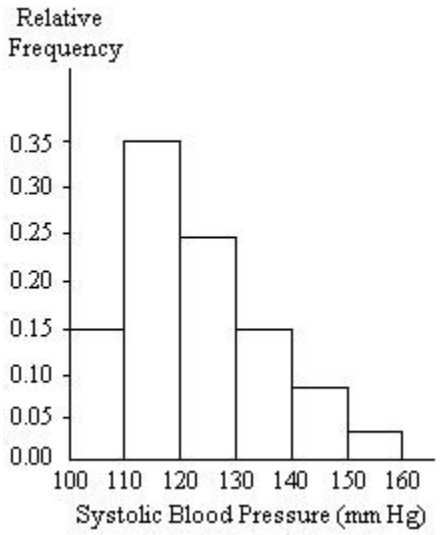
87 76 94 77 95 96 88 85 66 89
79 98 54 90 83 88 82 55 14 69

Create a stem–and–leaf display for the data. The stem should consist of the tens digit and range from 1 to 9. The leaves should be drawn aside the appropriate stem based on the data values.

5 Interpret Histogram

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

A nurse measured the blood pressure of each person who visited her clinic. Following is a relative –frequency histogram for the systolic blood pressure readings for those people aged 25 to 40. Use the histogram to answer the question. The blood pressure readings were given to the nearest whole number.



1) Approximately what percentage of the people aged 25–40 had a systolic blood pressure reading of at least 110 but less than 120?

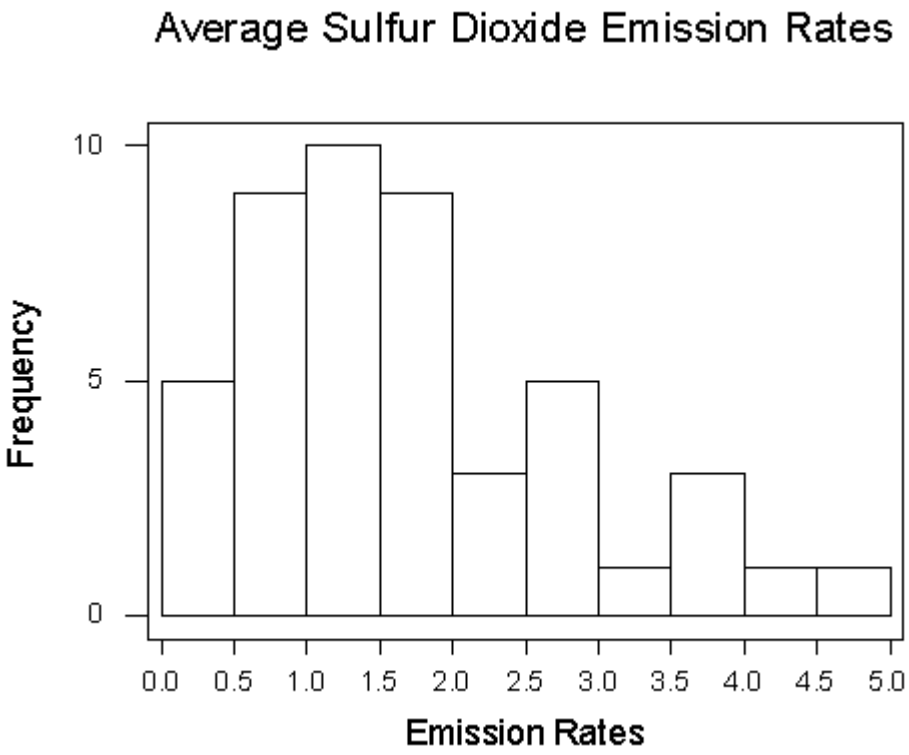
2) Approximately what percentage of the people aged 25–40 had a systolic blood pressure reading less than 120?

3) Given that 200 people were aged between 25 and 40, approximately how many had a systolic blood pressure reading less than 130?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

4) The paper "The Acid Rain Controversy: The Limits of Confidence" gave the following frequency histogram on average SO_2 (sulfur dioxide) emission rates from utility and industrial boilers (lb/million Btu) for 47 states (data for Idaho, Alaska, and Hawaii were not given).

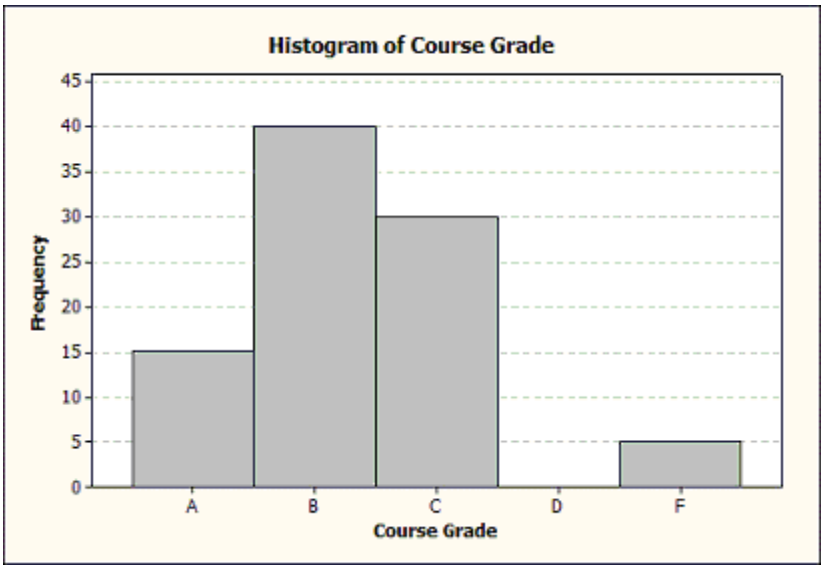


Source: American Statistician (1983): 385–394.

a. Identify the intervals of emission rates used for the plot.
b. Describe the shape of the distribution.
c. What information can you get from the dot plot or stem–and–leaf plot of these data that you cannot get from this plot?
d. This histogram shows frequencies. If you were to construct a histogram using the percentages for each interval, how (if at all) would the shape of this histogram change?

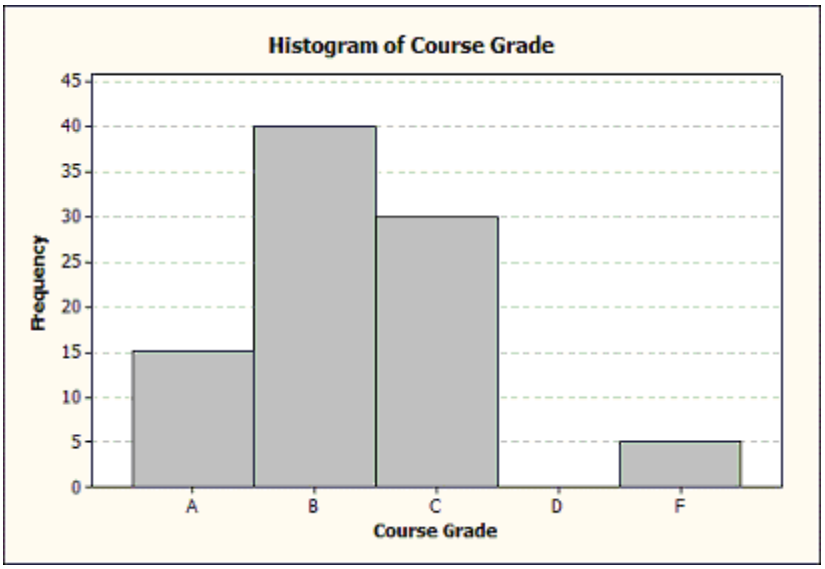
MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

5) The following is a partial histogram illustrating the final course grade distribution for an introductory level statistics class.



The data for a grade of "D" is missing. What is the correct frequency for the grade of "D"?

6) The following is a partial histogram illustrating the final course grade distribution for an introductory level statistics class.



The data for a grade of "D" is missing. Assuming 160 students were in the class, how many received a grade of "D"?

6 Construct/Interpret Time Plot

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

1) The data below represent the federal minimum wage from 1980 to 1989 in constant 1996 dollars. Constant dollars are dollars adjusted for inflation.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Min. Wage	5.90	5.78	5.45	5.28	5.06	4.88	4.80	4.63	4.44	4.24

Source: U.S. Employment Standards Administration

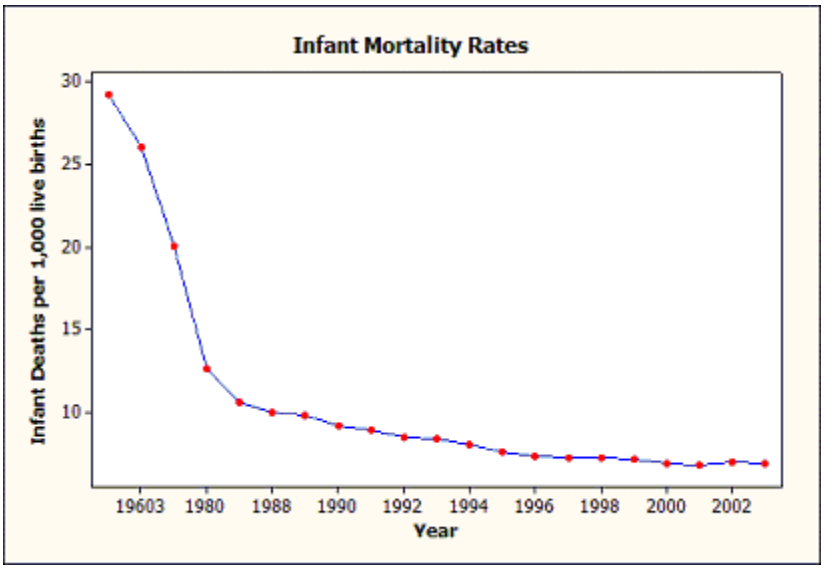
a. Construct a time plot for these data.
b. Is there a trend? If so, what kind?
c. Would a histogram more clearly describe the above dataset? Explain.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

2) The following is a time plot of U.S. wine exports (in millions of gallons) between the years 1986 and 2000. Is there a trend evident in the data?

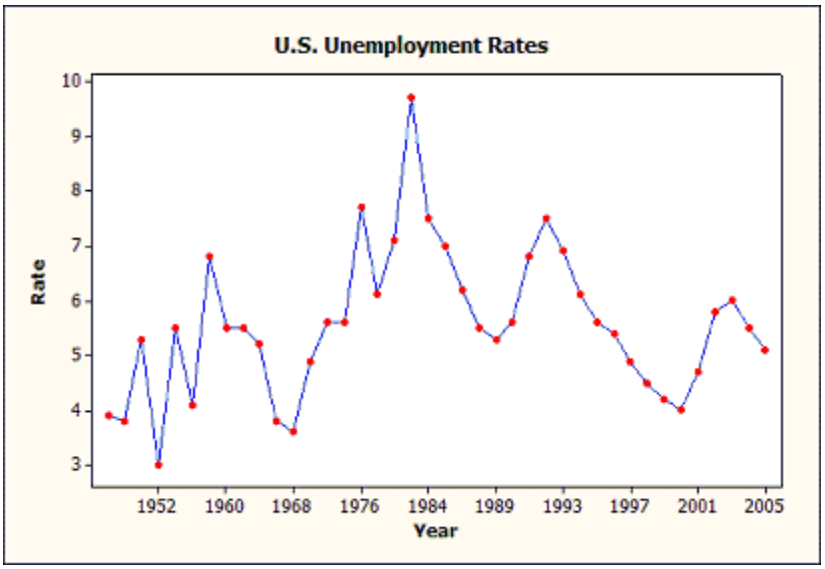
(source: U.S. department of Commerce, National Trade Data Bank)

3) The following is a time plot of infant mortality rates in the United States from the years 1950 to 2003. Is there an obvious trend in the data?



(source: Centers for Disease Control and Prevention, National Center for Health Statistics. From *Health, United States, 2005)*

4) The following plot illustrates a time series of U.S. unemployment rates in the civilian labor force between the years 1946 and 2005. Is a trend evident in the data set?



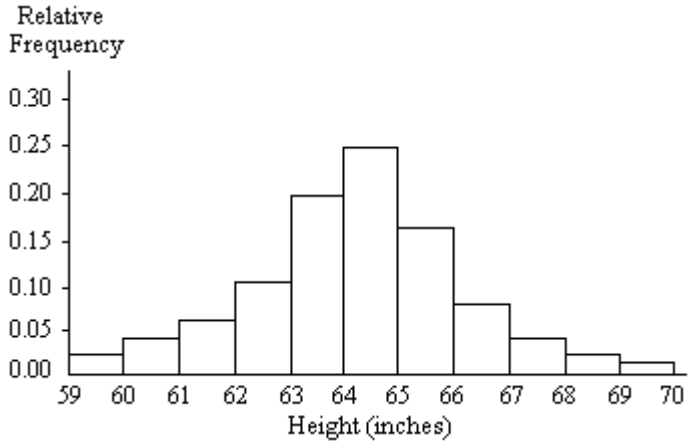
(Source: U.S. Department of Labor, Bureau of Labor Statistics)

7 Identify Shape of Distribution

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

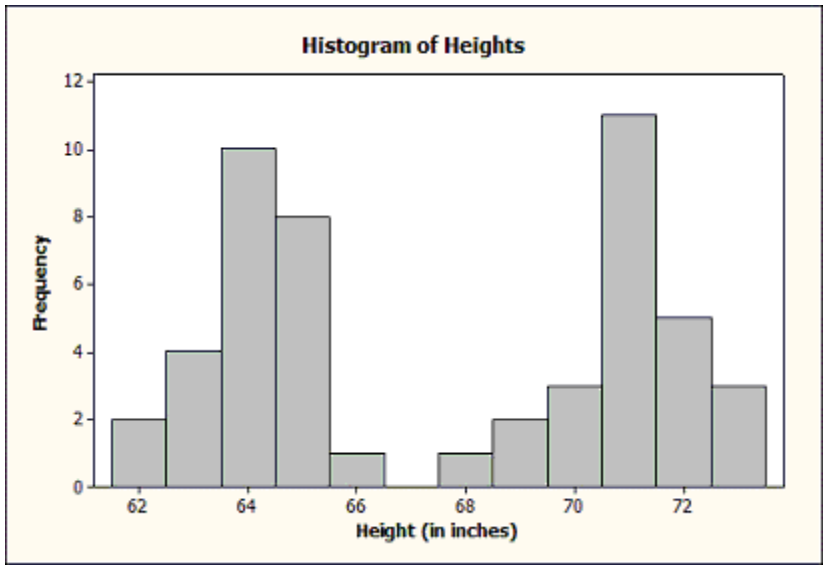
A graphical display of a data set is given. Identify the overall shape of the distribution.

1) A relative frequency histogram for the heights of a sample of adult women is shown below.



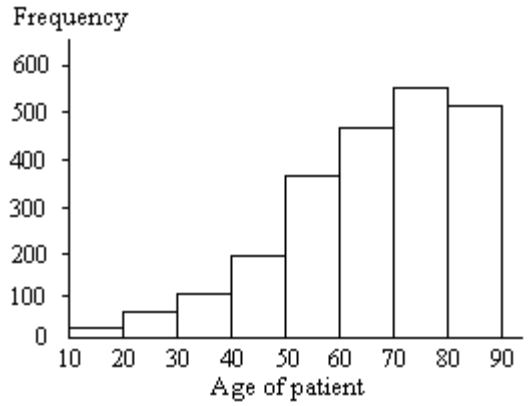
Which of the following best describes the shape of the distribution?

2) The following histogram depicts the heights of 50 women and 50 men.



Which of the following best describes the shape of the distribution?

3) The ages of a group of patients being treated at one hospital for osteoporosis are summarized in the frequency histogram below.



Which of the following best describes the shape of the distribution?

4) A stem–and–leaf diagram is given below for the ages of the patients at a hospital.

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

Which of the following best describes the shape of the distribution?

Select the most appropriate answer.

5) A distribution that shows an overall pattern with a single mound is called

6) A distribution that shows an overall pattern with two mounds is called

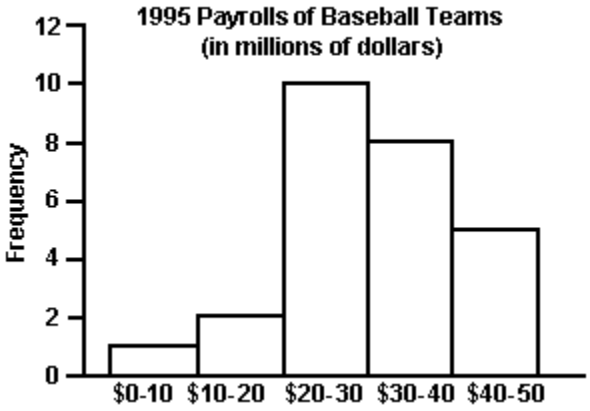
7) A distribution that has a left tail longer than the right tail is considered

8) A distribution that has the right tail longer than the left tail is considered

8 Graphical Summaries

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

The 1995 payroll amounts for all major–league baseball teams are shown below. Answer the following question concerning this graph.

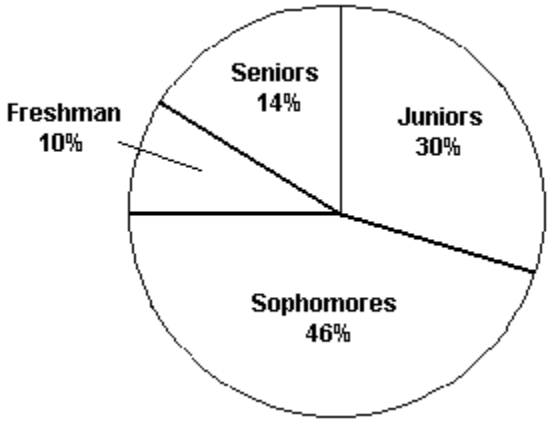


1) How many of the major–league payrolls exceeded $20 million in 1995? (Assume that no payroll was exactly $20 million in 1995.)

2) What percentage of the payrolls exceeded $30 million in 1995? (Assume that no payroll was exactly $30 million in 1995.)

Provide an appropriate response.

3) The professor of economics at a small Texas University wanted to determine what year in school students were taking his tough economics course. Shown below is a pie chart of the results.



What percentage of the class took the course prior to reaching their senior year?

Answer true or false.

4) Bar graphs and pie charts are graphical methods that are often used in summarizing quantitative data.

5) Dot plots and stem–and–leaf plots are often used to summarize small quantitative datasets.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

6) A _____ is a graph that uses bars to portray the frequencies or the relative frequencies of the possible outcomes for a quantitative variable.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Select the most appropriate answer.

7) Which of the following graphical methods cannot be used to summarize a quantitative dataset?

8) A set of data collected over time is called a

9) A common pattern observed over time is called a/an

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

10) Why is it beneficial to label each pie slice of a pie chart with its corresponding percent?

2.3 How Can We Describe the Center of Quantitative Data?

1 Measures of Center Analysis

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

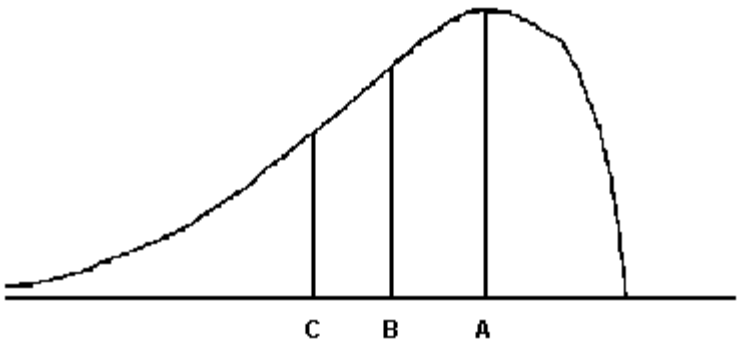
Provide an appropriate response.

1) At the U.S. Open Tennis Championship a statistician keeps track of every serve that a player hits during the tournament. The statistician reported that the mean serve speed of a particular player was 98 miles per hour. Suppose that the statistician indicated that the serve speed distribution was skewed to the left. Which of the following values is most likely the value of the median serve speed?

2) In 1990, U.S. consumers redeemed 6.12 billion manufacturers' coupons and saved themselves $2.86 billion. Calculate and interpret the mean savings of U.S. consumers per coupon.

3) The distribution of salaries of professional basketball players is skewed to the right. Which measure of central tendency would be the best measure to determine the location of the center of the distribution?

4) For the distribution shown below, identify the mean, median, and mode



5) The mean is less than the median

Answer true or false.

6) A numerical summary of the observations is called resistant if extreme observations have little, if any, influence on its value.

7) If a distribution is very highly skewed, the mean is usually preferred over the mean because it better represents what is typical.

8) In skewed distributions, we expect the values of the mean, median, and mode to be approximately equal, since they are all measures of center.

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

9) The _____ is the balance point of the data values; while, the _____ is the midpoint of the ordered data values.

10) Extreme observations in the dataset are called _____.

2 Find Mean

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) Brandon kept track of the number of hours he spent exercising each week for four months. The results are shown below. Find the mean number of hours Brandon spent exercising per week. Round your answer to two decimal places.

7.50	8.20	7.10	7.90	8.00	7.50
7.80	7.10	7.30	7.50	7.90	8.90
7.10	8.20	8.20	8.20	8.00	7.80

2) The normal monthly precipitation (in inches) for September is listed for 20 different U.S. cities. Find the mean of the data.

3.5	1.6	2.4	3.7	4.1
3.9	1.0	3.6	4.2	3.4
3.7	2.2	1.5	4.2	3.4
2.7	0.4	3.7	2.0	3.6

3) The age at inauguration for the last 15 U.S. presidents is given below. Find the mean age.

G.W. Bush	54
Clinton	46
G.H.W. Bush	64
Reagan	69
Carter	52
Ford	61
Nixon	56
L.B. Johnson	55
Kennedy	43
Eisenhower	62
Truman	60
F.D. Roosevelt	51
Hoover	54
Coolidge	51
Harding	55

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

4) In order to reduce pollutants from motor vehicle exhaust emissions, three–way catalytic converters have been installed in new vehicles. However, these converters increase the level of ammonia in the air. A study was published on the ammonia levels near the exit ramp of a San Francisco highway tunnel. The data below represent daily ammonia concentrations (parts per million) on eight randomly selected days during afternoon drive–time in the summer of 1999.

1.53	1.50	1.37	1.51	1.55	1.42	1.41	1.48

Source: Environmental Science & Technology (Sept. 1, 2000)

Find the mean.

3 Find Median

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Find the median for the given sample data.

1) Health care issues are receiving much attention in both academic and political arenas. A sociologist recently conducted a survey of senior citizens whose net worth is too high to qualify for Medicaid but who have no private health insurance. The ages of 25 uninsured senior citizens were as follows:

67 72 65 75 85 73
60 88 64 89 68 91
75 61 80 62 67 80
69 72 59 86 74 63 81

Find the median of the observations.

2) A store manager kept track of the number of newspapers sold each week over a seven–week period. The results are shown below.
95, 38, 221, 122, 258, 237, 233
Find the median number of newspapers sold.

Provide an appropriate response.

3) The age at inauguration for the last 15 U.S. presidents is given below. Find the median age.

G.W. Bush	54
Clinton	46
G.H.W. Bush	64
Reagan	69
Carter	52
Ford	61
Nixon	56
L.B. Johnson	55
Kennedy	43
Eisenhower	62
Truman	60
F.D. Roosevelt	51
Hoover	54
Coolidge	51
Harding	55

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

4) In order to reduce pollutants from motor vehicle exhaust emissions, three–way catalytic converters have been installed in new vehicles. However, these converters increase the level of ammonia in the air. A study was published on the ammonia levels near the exit ramp of a San Francisco highway tunnel. The data below represent daily ammonia concentrations (parts per million) on eight randomly selected days during afternoon drive–time in the summer of 1999.

1.53	1.50	1.37	1.51	1.55	1.42	1.41	1.48

Source: Environmental Science & Technology (Sept. 1, 2000)

Find the median.

5) The following data on daily protein intake (in grams of protein per kilogram of body weight) for 20 competitive athletes was obtained from a plot in the article "A Comparison of Plasma Glutamine Concentration in Athletes from Different Sports."

1.4	2.2	2.7	1.5	2.3	1.7	2.3	1.5	1.8	2.8
1.8	1.9	2.0	2.3	1.5	1.9	1.7	1.8	1.6	3.0

Source: Medicine and Science in Sports and Exercise (1998): 1693–1697.

Find the mean and the median. Which measure of center seems more appropriate for this dataset? Explain.

2.4 How Can We Describe the Spread of Quantitative Data?

1 Range

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) Last year, batting averages in the National League averaged .257 with a high of .323 and a low of .250 (minimum 250 at bats). Based on this information, which measure of variation could be calculated?

2) For the stem–and–leaf plot below, find the range of the data set.

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

3) The heights (in inches) of 20 adult males are listed below. Find the range of the data.

70 72 71 70 69 73 69 68 70 71
67 71 70 74 69 68 71 71 71 72

4) The age at inauguration for the last 15 U.S. presidents is given below. Find the range of the ages.

G.W. Bush	54
Clinton	46
G.H.W. Bush	64
Reagan	69
Carter	52
Ford	61
Nixon	56
L.B. Johnson	55
Kennedy	43
Eisenhower	62
Truman	60
F.D. Roosevelt	51
Hoover	54
Coolidge	51
Harding	55

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

5) The _____ is the difference between the largest and the smallest data values.

2 Standard deviation/variance

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) The cost for one semester's books (in dollars) are given below for a sample of five college students. Calculate the sample standard deviation, s of the book costs.

340, 170, 145, 420, 120

2) The heights (in inches) of 10 adult males are listed below. Find the standard deviation, s.

70 72 71 70 69 73 69 68 70 71

3) According to The College Board, the mean score on the SAT writing section was 497 for the 2006 graduating class. Noting that this test is scored on a scale of 200 to 800, what is the most plausible value for the standard deviation of the scores?

4) At the end of 2007, the proportion of adults aged 15–49 who were living with HIV/AIDS was 0.5% in Latin America, 1.0% in the Caribbean, 0.9% in Eastern Europe and Central Asia and 0.6% in North America. Suppose we include the proportion for Sub–Saharan Africa (5.0%) to this data set and calculate the standard deviation. Would you expect it to be significantly larger, smaller or remain about the same as the standard deviation of the proportions WITHOUT the observation from Sub–Saharan Africa? (www.avert.org)

Answer true or false.

5) The sum of the deviations, the differences between the observations and the sample mean $\sum(x - \bar{x})$, is always equal to zero.

Select the most appropriate answer.

6) Which of the following numerical summary measures cannot be negative?

3 Empirical Rule

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) A competency test has scores with a mean of 69 and a standard deviation of 4. A histogram of the data shows that the distribution is normal. Use the Empirical Rule to find the percentage of scores between 61 and 77.

2) SAT verbal scores are normally distributed with a mean of 433 and a standard deviation of 90. Use the Empirical Rule to determine what percent of the scores lie between 433 and 523.

3) Use the following summary information for a data set of 100 observations to determine whether the data set is likely to be bell–shaped, skewed to the right or skewed to the left.
Mean = 120, s=22, Minimum=37, Maximum=136

4) Use the following summary information for a data set of 100 observations to determine whether the data set is likely to be bell–shaped, skewed to the right or skewed to the left.
Mean = 120, s=22, Minimum=103, Maximum=170

5) According to the Empirical Rule, approximately 95% of the data values from a bell–shaped distribution fall within _____ standard deviations of the mean.

2.5 How Can Measures of Position Describe Spread?

1 Quartiles, IQR, Percentiles

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Determine the quartile, percentile or interquartile range as specified.

1) The test scores of 15 students are listed below. Find the first quartile, Q_1.

44 46 51 57 60
63 65 70 75 76
85 87 90 94 95

2) The test scores of 19 students are listed below. Find the interquartile range.

91 46 86 70 61
63 97 56 90 77
82 83 52 88 74
43 92 94 67

3) When Scholastic Achievement Scores (SAT's) are sent to test–takers, the percentiles associated with their scores are also given. Suppose a test–taker scored at the 75th percentile for their verbal grade and at the 37th percentile for their quantitative grade. Interpret these results.

4) The cholesterol levels (in milligrams per deciliter) of 30 adults are listed below. Find the interquartile range for the cholesterol level of the 30 adults.

154 156 165 165 170 171 172 180 184 185
189 189 190 192 195 198 198 200 200 200
205 205 211 215 220 220 225 238 255 265

Answer true or false.

5) The median is always the midpoint of Q1 and Q3.

Select the most appropriate answer.

6) One–fourth of the dataset lies

7) The median is equivalent to which quartile?

8) What percent of the data falls below Q1?

9) What percent of the data falls above Q2?

10) Which of the following numerical summary measures is not sensitive to outliers in a dataset?

2 Identify Outliers

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Identify potential outliers, if any, for the given data.

1) The test scores of 15 students are listed below.

36	40	48	65	67
69	70	73	75	76
79	82	87	90	99

2) The normal annual precipitation (in inches) is given below for 21 different U.S. cities.

32.4	29.4	34.6	65.3	22.1	31.8	16.6
28.2	36.2	59.4	24.3	47.2	45.6	9.2
27.1	18.9	13.6	31.4	24.2	12.3	35.4

3 Find Five–Number Summary

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Find the five–number summary for the given data.

1) The salaries (in millions of dollars) of the top 10 highest paid CEOs in the U.S. in 2006 according to Forbes business magazine.

249.42 230.55 139.96 135.53 122.67 80.73 75.33 71.84 69.66 68.95

2) The normal annual precipitation (in inches) is given below for 21 different U.S. cities.

39.1	32.9	18.5	35.6	27.1	27.8	8.6
23.5	42.6	34.7	20.2	12.0	5.1	13.9
22.6	10.9	16.4	25.4	17.2	14.7	51.7

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

3) The five–number summary of a dataset consists of the _____, _____, _____, _____, and _____.

Provide an appropriate response.

4) The article "Tobacco and Alcohol Use in G–Rated Children's Animated Films" investigated exposure to tobacco and alcohol use in all G–rated animated films released between 1937 and 1997 by five major film studios. Data on the total tobacco exposure time (in seconds) for films with tobacco use produced by Walt Disney, Inc., follow.

223	176	548	37	158	51	299	37	11
165	74	9	2	6	23	206	9	

Source: Journal of the American Medical Association (1999): 1131–1136.

Find the Five–Number Summary of Positions.

4 Construct/Interpret Box Plot

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Construct a boxplot as specified.

1) The weekly salaries (in dollars) of 24 randomly selected employees of a company are shown below. Construct a boxplot for the data set. What is the shape of the distribution?

310	320	450	460	470	500	520	540
580	600	650	700	710	840	870	900
1000	1200	1250	1300	1400	1720	2500	3700

Select the most appropriate answer.

2) Which of the following numerical summary measures cannot be easily approximated from a box plot?

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

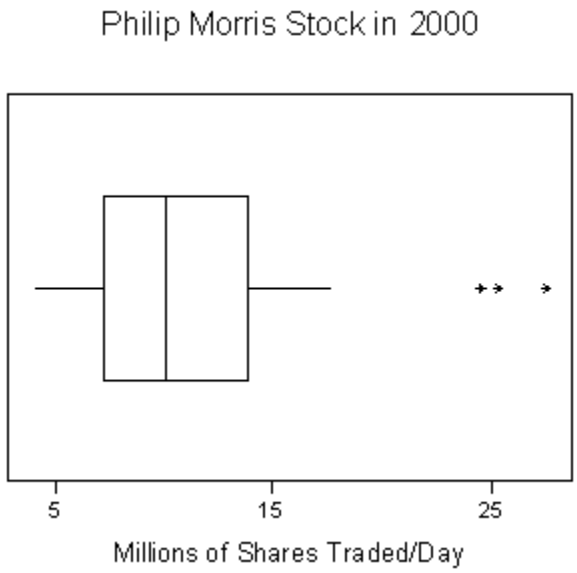
3) 1.The data below represent the number of inches of rain in Chicago, Illinois, during the month of April for 20 randomly selected years.

2.47	3.97	3.94	4.11	1.14
4.02	3.41	1.85	5.22	0.97
6.14	2.34	3.48	4.77	2.78
4.00	6.28	5.50	7.69	5.79

Source: NOAA, Climate Diagnostics Center

a. Construct a box plot for these data.
b. Describe the shape of this distribution.
c. Compute and interpret the standard deviation.

4) The box plot below represents the volume of Philip Morris stock traded for a random sample of 35 trading days in 2000. The volume of a stock is the number of shares traded on a given day.



Source: http://finance.yahoo.com

a. Approximately, what is the median for this dataset?
b. Are there any potential outliers in this dataset? If so, how many?
c. Describe the shape of the distribution. Would the standard deviation or the interquartile range be a better measure of spread for this dataset? Explain.

5 Use z–Score

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) Test scores for a history class had a mean of 79 with a standard deviation of 4.5. Test scores for a physics class had a mean of 69 with a standard deviation of 3.7. Suppose a student gets a 68 on the history test and a 87 on the physics test. Calculate the z–score for each test. On which test did the student perform better?

2) According to the Center for Disease Control growth charts of 2000, the weight at birth of males has a mean value of 3.53 kg with a standard deviation of 0.58. For a male child weighing 2.75 kg at birth, what is the corresponding z–score? (www.cdc.gov/growthcharts/)

3) According to the Center for Disease Control growth charts of 2000, the weight at birth of males has a mean value of 3.53 kg with a standard deviation of 0.58. What birth weight has a z–score of 0.81? (www.cdc.gov/growthcharts/)

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Fill in the blank.

4) The _____ for a data value is the number of standard deviations that it falls from the mean.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Select the most appropriate answer.

5) In human engineering and product design, it is important to consider the weights of people so that airplanes or elevators aren't overloaded. Based on data from the National Health Survey, the weight for adult males in the U.S. follows a bell–shaped distribution with a mean weight of 173 pounds and a standard deviation of 30 pounds. What proportion of these weights is between 203 pounds and 263 pounds?

6) In human engineering and product design, it is important to consider the weights of people so that airplanes or elevators are not overloaded. Based on data from the National Health Survey, the weight for adult males in the U.S. follows a bell–shaped distribution with a mean weight of 173 pounds and a standard deviation of 30 pounds. Using the z–score approach for detecting outliers, which of the following weights would represent potential outliers in the distribution of U.S. adult male weights?

Weights: 110 pounds, 157 pounds, 281 pounds

7) In human engineering and product design, it is important to consider the weights of people so that airplanes or elevators are not overloaded. Based on data from the National Health Survey, the distribution of weights for adult males in the U.S. has a mean weight of 173 pounds and a standard deviation of 30 pounds. Suppose the distribution of weights was skewed to the left. Which of the following values is most likely the value of the median weight?

8) According to the U.S. Bureau of the Census, County and City Data Book (1977), the area of New Jersey Counties, in square miles, ranges from 47 to 819 with Q1 =228, median=329 and Q3=476. The full data set follows.

47 103 130 192 221 228 234 267 307 312 329 362 365 423 468 476 500 527 569 642 819

In a boxplot, what would be the values to which the whiskers extend

9) According to the U.S. Bureau of the Census, County and City Data Book (1977), the area of New Jersey Counties, in square miles, ranges from 47 to 819 with Q1 =228, median=329 and Q3=476. The full data set follows.

47 103 130 192 221 228 234 267 307 312 329 362 365 423 468 476 500 527 569 642 819

According to the 1.5 × IQR criterion, are there any potential outliers in the data set?

10) he salaries of the top 10 highest paid CEOs in the U.S. ranged from 249.42 to 68.95 million dollars in 2006 according to Forbes business magazine. These data had Q1 =71.84, median=101.7 and Q3=139.96. The full data set is given below.

249.42 230.55 139.96 135.53 122.67 80.73 75.33 71.84 69.66 68.95

Using the 1.5 x IQR criterion, are there any potential outliers in the data set?

11) The salaries of the top 10 highest paid CEOs in the U.S. ranged from 249.42 to 68.95 million dollars in 2006 according to Forbes business magazine. These data had Q1 =71.84, median=101.7 and Q3=139.96. The full data set is given below.

249.42 230.55 139.96 135.53 122.67 80.73 75.33 71.84 69.66 68.95

In a boxplot, what would be the values to which the whiskers extend?

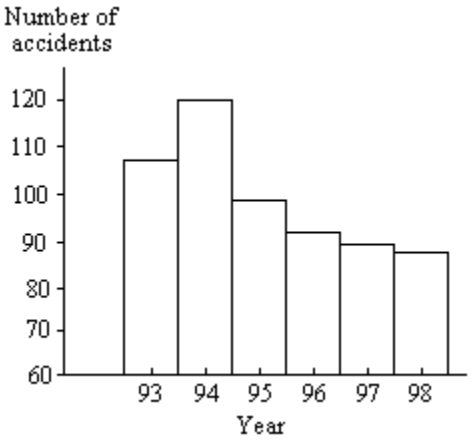
2.6 How Can Graphical Summaries Be Misused?

1 Misleading Summaries

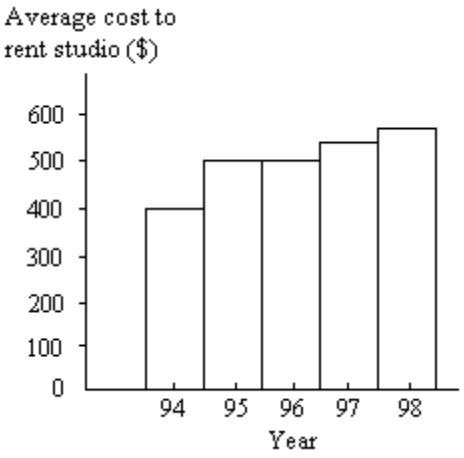
SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Provide an appropriate response.

1) The histogram below shows the number of car accidents occurring in one city in each of the years 1993 through 1998. The number of accidents dropped in 1995 after a new speed limit was imposed. Why is the graph misleading? How would you redesign the graph to be less misleading?

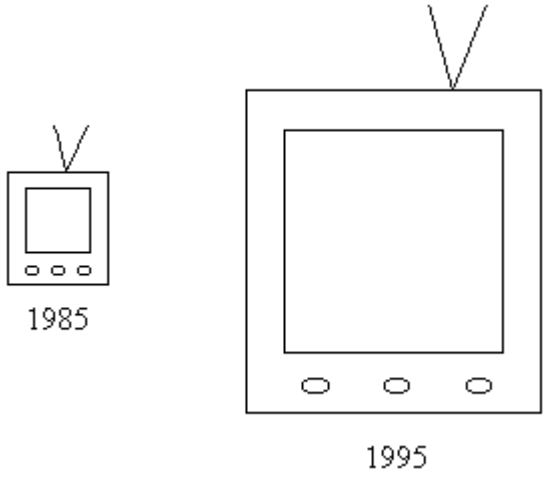


2) The bar graph below shows the average cost of renting a studio in one city in each of the years 1994 through 1998.



By what percentage does the average price increase from 1994 to 1995? Obtain a truncated version of the graph by sliding a piece of paper over the bottom of the graph so that the bars start at 300. In the truncated graph, by what percentage does the price appear to increase from 1994 to 1995? Why is the truncated graph misleading?

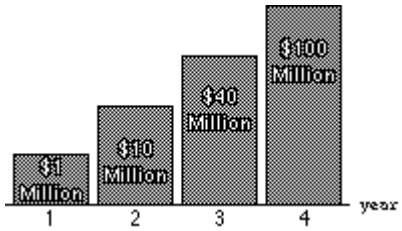
3) A television manufacturer sold three times as many televisions in 1995 as it did in 1985. To illustrate this fact, the manufacturer draws a pictogram as shown below. The television on the right is three times as tall and three times as wide as the television on the left.



Why is this pictogram misleading? What visual impression is portrayed by the pictogram?

Identify the abuse of statistics.

4) The graph shows the increases in a certain expenditure over a four-year period. What is wrong with the graph?



MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

5) Which of the following is not a guideline to use for constructing effective graphs?

Ch. 2 Exploring Data with Graphs and Numerical Summaries
Answer Key

2.1 What Are the Types of Data?

1 Categorical or Quantitative Variables

1) A
2) A
3) A
4) A
5) categorical
6) quantitative
7) A

2 Discrete or Continuous Variables

1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A

3 Interpret Frequency Table

1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A
9) A
10) a. stock performance
    b. categorical
    c. up
    d.

Stock performance	up	same	down
Count	0.525	0.175	0.300

11) a. number of children under five
    b. discrete
    c. 1
    d.

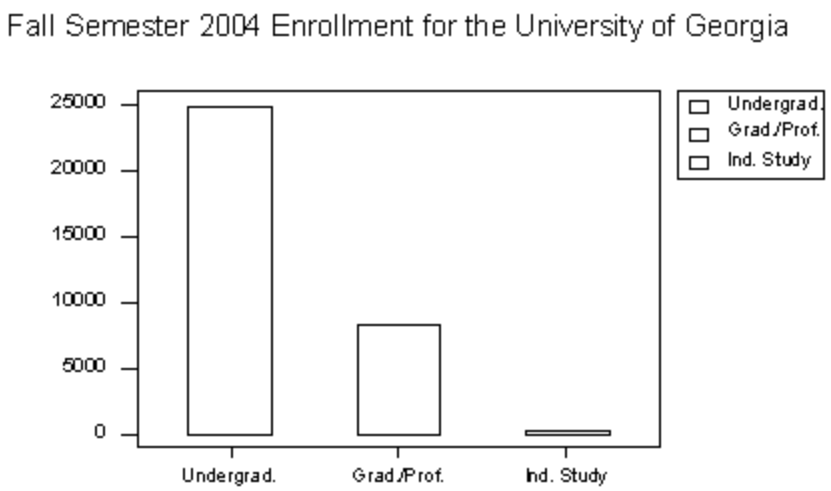
Number of children under five	0	1	2	3	4
Count	0.25	0.30	0.20	0.20	0.05

2.2 How Can We Describe Data Using Graphical Summaries?

1 Construct/Interpret Bar Chart

1) A
2) A
3) A
4) A
5) A
6) A

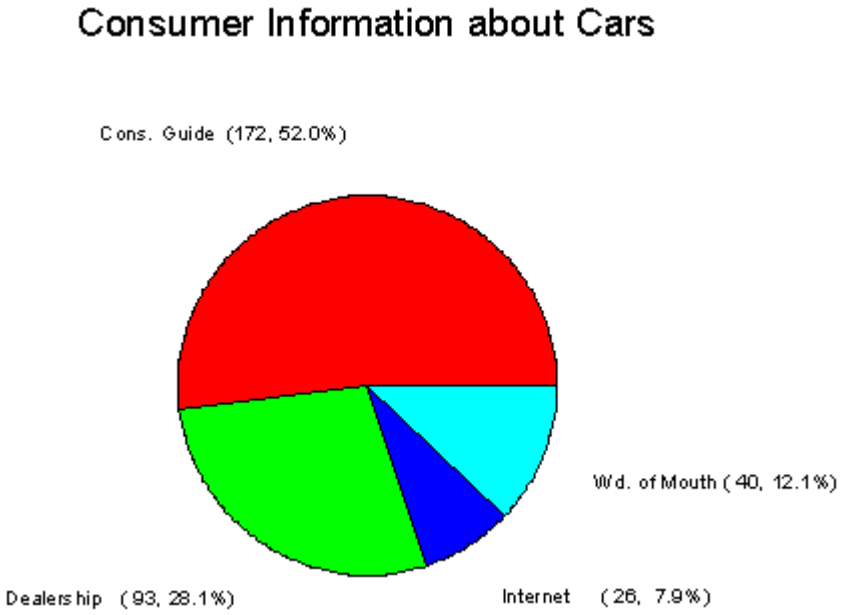
7) A
8) a.



b. No, both a dot plot and a stem–and–leaf plot are used on small quantitative datasets.

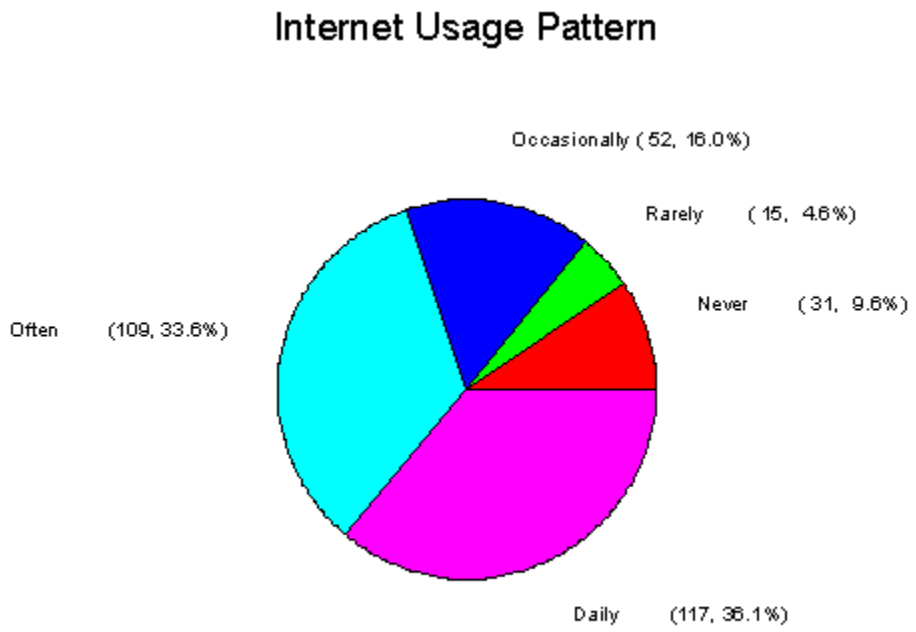
2 Construct/Interpret Pie Chart

1) A
2) A
3) A
4) A
5) a.



b. Since it is of interest to know which categories were more useful to consumers, ordering the categories as in a Pareto chart would be more appropriate than listing them alphabetically.

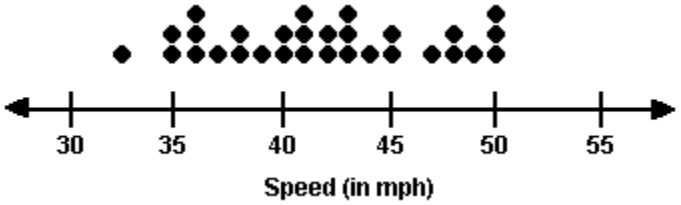
6) a.



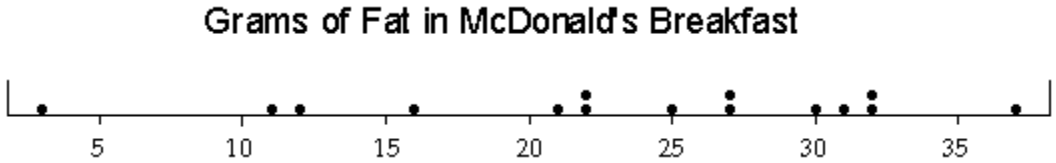
b. Since the categories of Internet usage pattern have a natural order from never to daily, it makes more sense to leave the categories in this natural order rather than ordering them from the tallest bar to the shortest bar.

3 Construct/Interpret Dot Plot

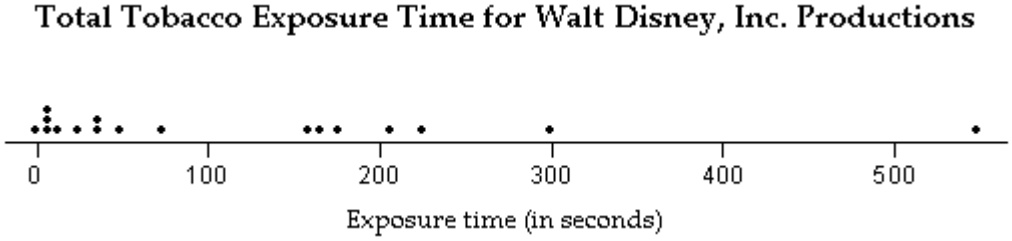
1) A
2) A
3) A
4) A
5)



6)

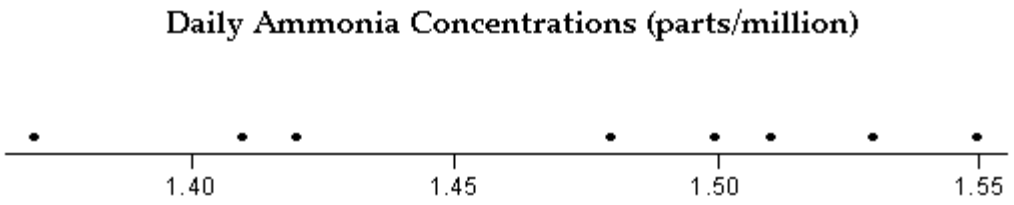


7)



This distribution appears to be skewed to the right.

8)



4 Construct/Interpret Stem–and–Leaf Plot

1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A
9) A
10) A
11) A
12)

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

5 Interpret Histogram

1) A
2) A
3) A
4) a. 0 to 0.49, 0.5 to 0.99, 1.0 to 1.49, 1.5 to 1.99, 2.0 to 2.49, 2.5 to 2.99, 3.0 to 3.49, 3.5 to 3.99, 4.0 to 4.49, 4.5 to 4.99; b. The distribution is skewed to the right. c. You can get the actual data values from a dot plot or stem–and–leaf plot. d. The shape would not change.
5) A
6) A

6 Construct/Interpret Time Plot

1) a.

b. There is a clear decreasing trend over time; c. No, a histogram would not depict the trend in this dataset.

2) A
3) A
4) A

7 Identify Shape of Distribution

1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A

8 Graphical Summaries

1) A
2) A
3) A
4) A
5) A
6) histogram
7) A
8) A
9) A
10) This clarifies what percent a slice represents and which of two slices is larger.

2.3 How Can We Describe the Center of Quantitative Data?

1 Measures of Center Analysis

1) A
2) A
3) A
4) A

5) A
6) A
7) A
8) A
9) mean; median
10) outliers

2 Find Mean

1) A
2) A
3) A
4) mean = 1.471

3 Find Median

1) A
2) A
3) A
4) median = 1.49
5) mean = 1.985, median = 1.85; The median seems more appropriate for this dataset, because this dataset is highly skewed to the right.

2.4 How Can We Describe the Spread of Quantitative Data?

1 Range

1) A
2) A
3) A
4) A
5) range

2 Standard deviation/variance

1) A
2) A
3) A
4) A
5) A
6) A

3 Empirical Rule

1) A
2) A
3) A
4) A
5) A

2.5 How Can Measures of Position Describe Spread?

1 Quartiles, IQR, Percentiles

1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A
9) A
10) A

2 Identify Outliers

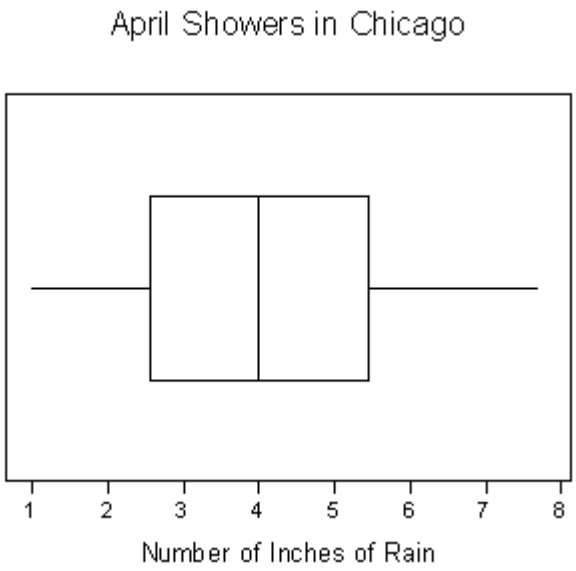
1) A
2) A

3 Find Five–Number Summary

1) A
2) A
3) minimum value; Q1; median; Q3; maximum value
4) minimum = 2 seconds, Q1 = 10 seconds, median = 51 seconds, Q3 = 191 seconds, and maximum = 548 seconds

4 Construct/Interpret Box Plot

1) A
2) A
3) a.



b. The distribution is approximately symmetrical; c. standard deviation = 1.779 inches; The typical distance the data falls from the mean is 1.779 inches.

4) a. median = about 10 million shares; b. yes, 3; c. The distribution is skewed to the right. The IQR would be a better measure of spread for this dataset, because it is highly skewed and contains 3 potential outliers. The standard deviation is not a resistant measure of spread.

5 Use z–Score

1) A
2) A
3) A
4) z–score
5) A
6) A
7) A
8) A
9) A
10) A
11) A

2.6 How Can Graphical Summaries Be Misused?

1 Misleading Summaries

1) Possible answer: The graph is misleading because it is truncated. The scale on the vertical axis should start at zero so that the bars will be in the correct proportions. A part of the vertical axis could be omitted but the symbol // should then be used to warn the reader of the modified axis.
2) Possible answer: The average price increases by 25% from 1994 to 1995. Using the truncated graph, the price appears to double from 1994 to 1995 (i.e. it appears to increase by 100%). Using the truncated graph, the differences between the bars seem bigger (relatively) than they really are.
3) Possible answer: The area of the television on the right is nine times (not three times) the area of the television on the left. The pictogram gives the visual impression that sales in 1995 were nine times the sales in 1985.

4) The bars are not drawn in the correct proportions.
5) A