Chapter 2 - Graphical Methods for Describing Data Distributions

True / False

1. A data set is discrete if the possible values are isolated points on the number line.
    a. True
    b. False
ANSWER: True

2. A data set consisting of many observations of a single characteristic is a categorical data set.
    a. True
    b. False
ANSWER: False

3. A data set is multivariate if it consists only of numeric variables.
    a. True
    b. False
ANSWER: False

4. Frequency distributions can only be used with categorical data.
    a. True
    b. False
ANSWER: False

5. The relative frequency for a particular category is the number of times the category appears in the data.
    a. True
    b. False
ANSWER: False

6. Bar charts should be used with categorical data.
    a. True
    b. False
ANSWER: True

7. Dotplots work best for small and moderate sized numerical data sets.
    a. True
    b. False
ANSWER: True

8. An outlier is an unusually small or large data value.
    a. True
    b. False
ANSWER: True

9. The quantity $\sqrt{\text{number of observations}}$ often gives a rough estimate of the appropriate number of intervals in a histogram.
    a. True

b. False

ANSWER: True

10. A curve with tails that decline more rapidly than the tails of a normal curve is called a heavy-tailed distribution.
   a. True
   b. False

ANSWER: False

11. The density of a class can be calculated by multiplying the relative frequency of the class times the class width.
   a. True
   b. False

ANSWER: False

12. For stem and leaf plots with single-digit leaves, commas must be used to separate the leaves.
   a. True
   b. False

ANSWER: False

13. One advantage of histograms is that they may be used for large data sets.
   a. True
   b. False

ANSWER: True

14. If the upper tail of a distribution stretches out farther than the lower tail, the distribution is negatively skewed.
   a. True
   b. False

ANSWER: False

15. In a scatter plot, both the horizontal and vertical axes must be set at zero.
   a. True
   b. False

ANSWER: False

16. A pie chart is most useful for numeric data.
   a. True
   b. False

ANSWER: False

Multiple Choice

17. A survey form solicited the following responses:

I). age of respondent
II). gender of respondent
III). level of job satisfaction (completely dissatisfied/somewhat dissatisfied/somewhat satisfied/completely satisfied)
IV). annual income

Chapter 2 - Graphical Methods for Describing Data Distributions

Which of the responses represent *categorical* data?
    a. I only
    b. II only
    c. III only
    d. II and III only
    e. All the responses are categorical.
ANSWER: d

18. The following relative frequency distribution summarizes student responses to the question "If you could make your college choice over, would you still choose to enroll at your current college?"

Response	Relative Frequency
Definitely yes	0.45
Probably yes	0.33
Probably no	0.13
Definitely no	0.09

Which of the following is the correct bar chart for the response data?

a.

b.

c.

d.

e.
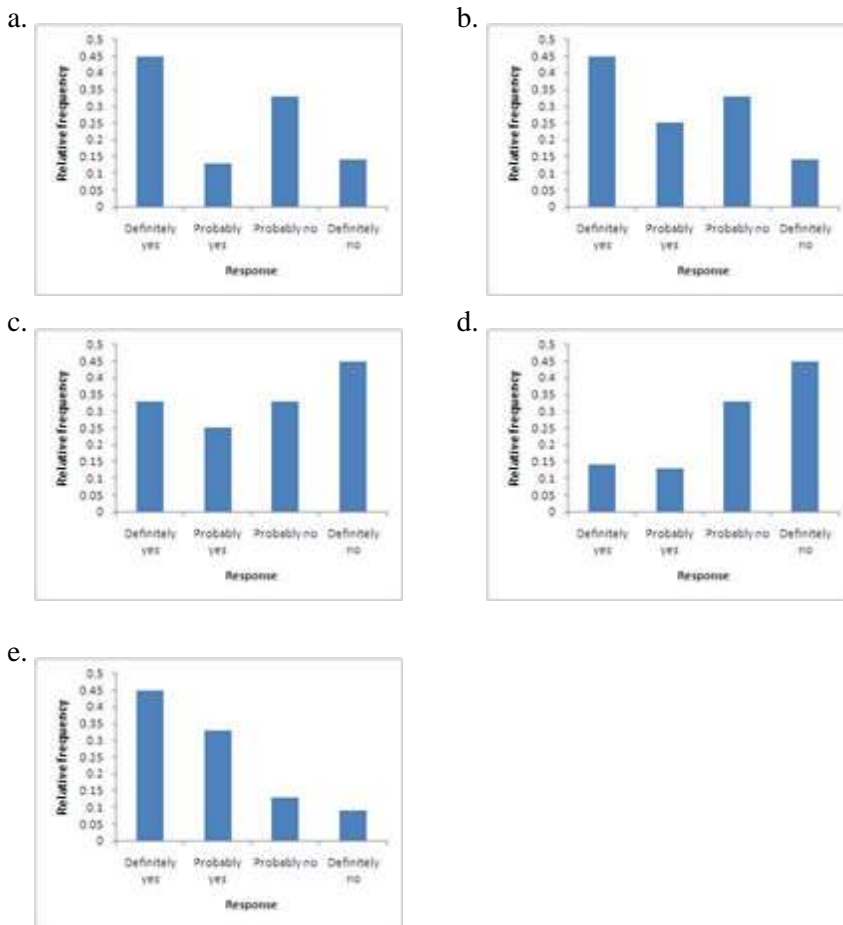
ANSWER: e

19. Which of the following variables are discrete?

I). the volume of liquid in a 16-ounce bottle of soda pop

Chapter 2 - Graphical Methods for Describing Data Distributions

II). the percentage of males 18-25 who participate in a contact sport
III). the number of broken eggs in a package of a dozen eggs
IV). a count of the statistics majors at a certain university

    a. II only

    b. II and III only

    c. III and IV only

    d. I only

   e. All of these variables are discrete.

ANSWER: c

20. A survey asked adult respondents how dependent they were on various electronic devices. The accompanying table summarizes the responses.
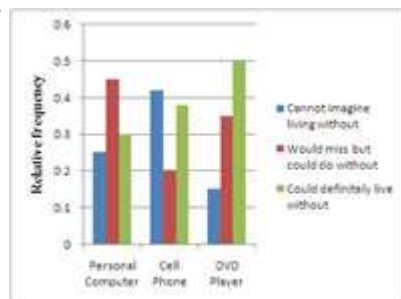
Response	Relative Frequency		
	Personal Computer	Cell Phone	DVD Player
Cannot imagine living without	0.25	0.42	0.15
Would miss but could do without	0.45	0.20	0.35
Could definitely live without	0.30	0.38	0.50

Select a comparative bar chart that shows the distributions of responses for the three different technologies.
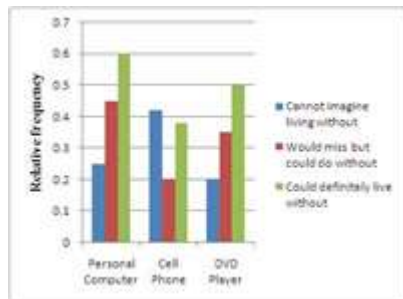
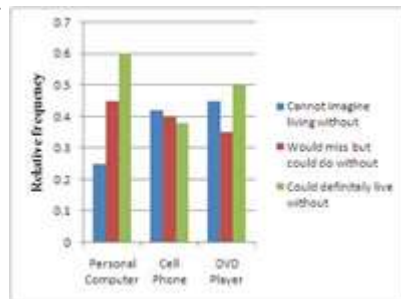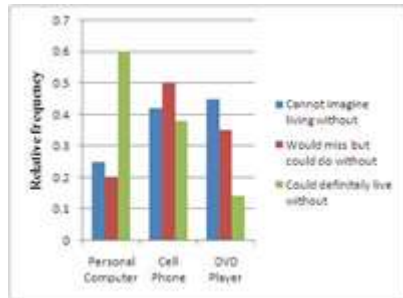   a.
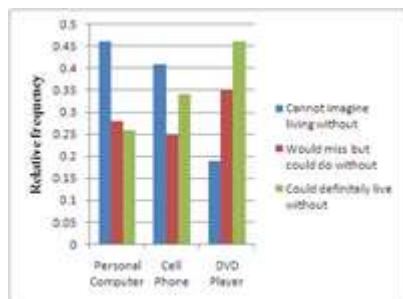
    b.

    c.

    d.

    e.

Chapter 2 - Graphical Methods for Describing Data Distributions

ANSWER: b

21. According to the stem-and-leaf display below, how many times does the number 58 appear in the data set?

Stem: Tens	Leaf: Ones
2	5
3	222355679
4	045
5	11225889
6	03689
7	69
8	04

  a. 0
  b. 1
  c. 2
  d. 5
  e. 58

ANSWER: c

22. Which of the following stem and leaf plots corresponds a data distribution that is positively (right) skewed?

I.

Leaf unit = 1	
2	159
3	
4	389
5	03346
6	004567
7	59
8	7

II.

Leaf unit = 1	
1	0124458
2	11135
3	29
4	239
5	3
6	36

III.

Leaf unit = 1	
1	6
2	
3	
4	
5	3457
6	58
7	0347
8	022346899

  a. I only
  b. II only
  c. III only
  d. I and III only
  e. None of the distributions is positively skewed.

ANSWER: b

Chapter 2 - Graphical Methods for Describing Data Distributions

23. The annual household incomes, in thousands of dollars, for 150 households in a small community are summarized in the histogram below.
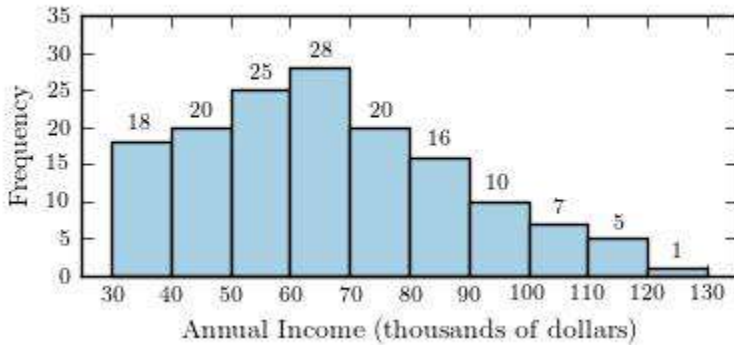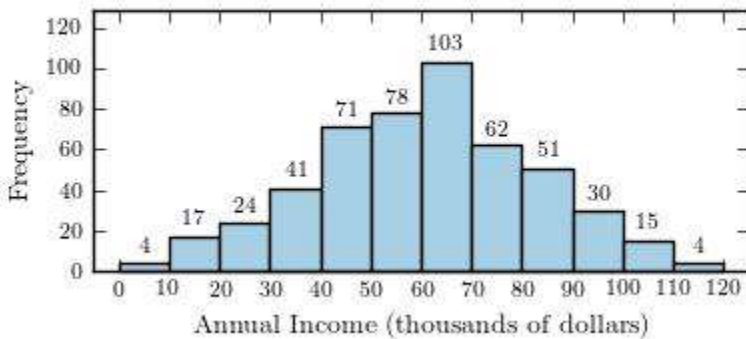

What proportion of households have annual incomes of $90,000 or more?

    a. 0.85

    b. 0.67

    c. 0.08

    d. 0.15

    e. 0.11

ANSWER: d

24. The annual household incomes, in thousands of dollars, for 500 households in a small community are summarized in the histogram below.

Which of the following values best describes a typical household income for this community?

    a. $40,000

    b. $50,000

    c. $60,000

    d. $70,000

    e. $80,000

ANSWER: c

25. The frequency distribution shown summarizes data on the percentage of college students enrolled in public institutions for the 50 U.S. states.

Percentage Enrolled	Frequency
40 to < 50	1

<u>**Chapter 2 - Graphical Methods for Describing Data Distributions**</u>

| 50 to < 60	5
60 to < 70	2
70 to < 80	11
80 to < 90	25
90 to < 100	6

What is the relative frequency for the 70 to < 80 interval?

    a. 0.05

    b. 0.10

    c. 0.22

    d. 11

    e. 50

ANSWER: c

26. The cumulative relative frequency plot shown was constructed using data on the percentage of college students enrolled in public institutions for the 50 U.S. states. Which of the following is closest to the approximate *proportion* of states that have more than 90 percent of college students enrolled in public institutions?
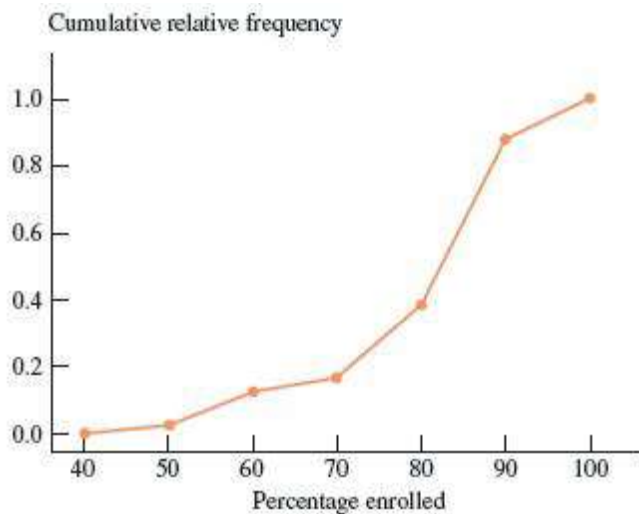
    a. 0.6

    b. 0.1

    c. 0.0

    d. 0.4

    e. 0.9

ANSWER: b

Subjective Short Answer

27. In a psychological study of distance perception, introductory psychology class volunteers were given 5, 10, or 15 extra credit points depending on their level of participation in the experiment. They walked along paths laid out along hallways in the psychology building, and were asked to judge the distance they had walked. Each path distance was between 100 and 200 feet in length, and contained from two to seven 90 degree turns. The investigators also kept track of class standing (freshman, sophomore, junior, or senior). For each of the variables in this study, determine whether it is categorical or numerical. For each numerical variable, determine whether it is discrete or continuous:

Chapter 2 - Graphical Methods for Describing Data Distributions

(a)   The true length of the path

(b)   The amount of credit given to the volunteers

(c)   The length of the path as judged by the students

(d)   The number of 90 degree turns

(e)   Class standing

ANSWER: a)   Numerical, continuous
            b)   Numerical, discrete
            c)   Numerical, continuous
            d)   Numerical, discrete
            e)   Categorical

28. In a study of how college students give directions from maps, 40 Introductory Psychology volunteers, 20 men and 20 women, were given the task of giving directions to another. Observations were made of subjects' direction-giving behavior in the experiment:

(a)   whether the map was available, or the directions were given from memory,
(b)   the sex of the direction giver,
(c)   the distances given as part of the directions,
(d)   the number of times directions such as North, Southwest, etc. were used, and
(e)   the frequency of errors in the directions
For each of the variables in this experiment, determine whether it is categorical or numerical. For each numerical variable, determine whether it is discrete or continuous:

(a)   whether the map was available, or the directions were given from memory

(b)   the sex of the direction giver

(c)   the distances given as part of the directions

(d)   the number of times directions such as North, Southwest, etc. were used

(e)   Frequency of errors in the directions
ANSWER: a)   Categorical
            b)   Categorical
            c)   Numerical, continuous
            d)   Numerical, discrete
            e)   Numerical, discrete

29. The most important right and responsibility of citizens in a democratic society is voting. The bar chart below shows, for different age groups in the 2000 presidential election in the United States, the percentage of individuals in that age group who are registered to vote.

Chapter 2 - Graphical Methods for Describing Data Distributions
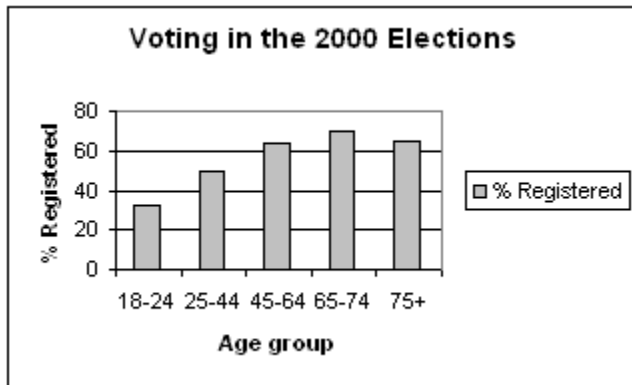
(a)   Describe in a few sentences the registration pattern that you see in the data above.

(b)   Some individuals argue that the reason for the current level of registration in the 18-24 year-old group is that many of them are college students and would have to vote absentee. In your opinion, would that explain the registration level for the 18 - 24 age-group? Why or why not?

ANSWER:

a)   The proportion of people registering is smallest for the 18-24 year-old group, and gets larger with age, until finally declining in the 75+ year-old group.

b)   Voting by absentee ballot may be more of a hassle b/c you have to get the ballot, but actual registration could be done at home before going to college. It would seem that unless the 18-24 year-olds decide for some reason they will not vote in their college years because of expected hassle of voting, the future voting would not be affecting the registration level.

30. Artificial nest have often been used when biologists study predation on birds. It is usually assumed that predators will respond to artificial nests in the same way they do to real birds' nests. A recent experiment was performed to check this assumption for predators of the American robin (*Turdus migratorius*) by comparing the egg loss in three experimental treatments: (1) natural robin's nests, (2) artificial wicker-basket nests placed 30 meters in a random direction from natural robin's nests, and (3) artificial wicker-basket nests placed in a 180 m by 240 m grid around robin's nests. The data below are <u>cumulative</u> counts of eggs predated (stolen or broken and eaten) at 5, 10, and 15 day points of the experiment.

Cumulative count of depredated eggs

Nest type	Original *n*	Day 5 check	Day 10 check	Day 15 check
Natural robin	27	12	17	19
30 m from robin	27	3	12	18
Grid	48	7	32	44

(a)   Construct a <u>bar chart</u> to display the cumulative counts for the natural robin nests at their 5-, 10-, and 15-day checks.
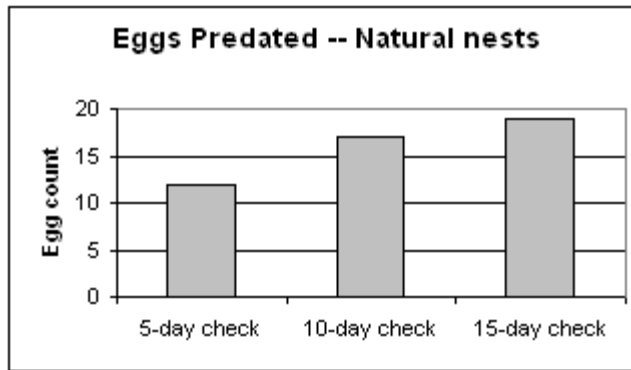
(b)   Fill in the <u>relative</u> frequency distribution (table) representing the cumulative counts for the artificial nests in the <u>grid</u>. (Do not convert the table to a graph.)

Nest type	Original *n*	Day 5 check	Day 10 check	Day 15 check
Natural robin	27			
30 m from robin	27			
Grid	48			

Chapter 2 - Graphical Methods for Describing Data Distributions

ANSWER:

a)

Eggs Predated -- Natural nests

b)

Nest type	Original n	Day 5 check	Day 10 check	Day 15 check
Natural robin	27			
30 m from robin	27			
Grid	48	**0.146**	**0.667**	**0.917**

31. Bird-plane collisions

Collisions between birds and airplanes are a major problem, threatening passenger safety and resulting in costly repairs. During the years 1990 - 1994, researchers gathered bird remains from runways at Lihue Airport on the island of Kauai, Hawaii. The data below are counts of birds found on the runway used for takeoffs and the runway used for landings from the 4 most frequently involved species:

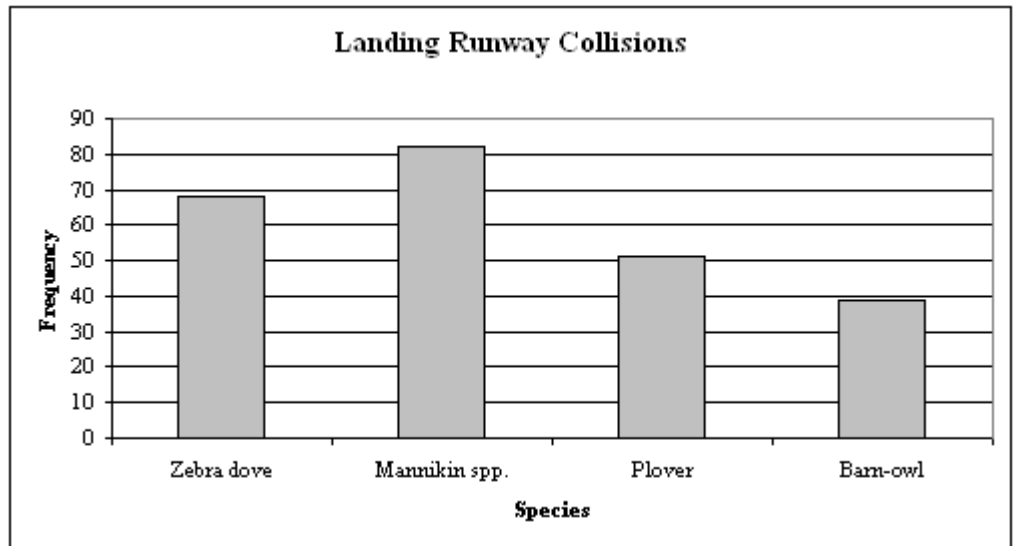| Species	Takeoff	Landing
Zebra dove	49	68
Mannikin spp.	33	82
Plover	29	51
Barn-owl	6	39

(a) Construct a <u>bar chart</u> to display the bird species data for the <u>landing</u> runway.

(b) Construct a <u>relative</u> frequency distribution (table) representing the bird species data for the <u>takeoff</u> runway.

Chapter 2 - Graphical Methods for Describing Data Distributions

ANSWER:

a)

b)

| Species	Rel. Freq. Takeoff
Zebra dove	.4188
Mannikin spp.	.2821
Plover	.2479
Barn-owl	.0513

32. A statewide survey released by the University of Iowa Injury Prevention Research Center was designed to assess compliance with Iowa seatbelt law. The table below shows the results for 5 Iowa cities.

City	% Age 2 – 5 properly restrained	% Age 6-10 properly restrained
Belle Plaine	84	93
Cedar Rapids	79	94
Guttenberg	58	68
Iowa City	92	99
Manchester	59	82

(a) Display these data in a comparative bar chart so that the different ages can be compared for the different cities.

(b) Write a brief description of the differences and similarities in the patterns you observe in your bar chart for part (a).

Chapter 2 - Graphical Methods for Describing Data Distributions

ANSWER:

(a)
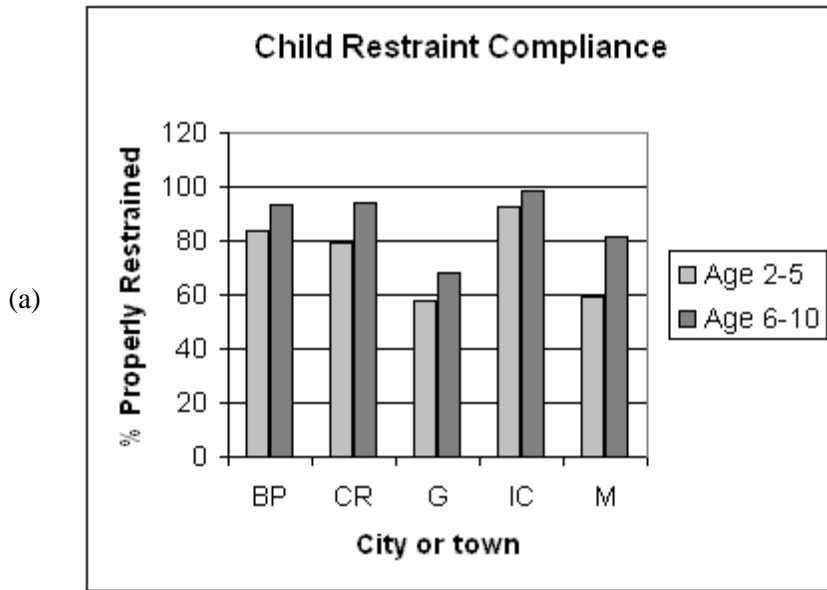
(b) The patterns across ages are the same for each of the cities--the 6-10 year age group has a higher level of compliance than the age 2 - 5 group. It appears that generally Belle Plaine, Cedar Rapids, and Iowa City have higher compliance rates than Guttenberg and Manchester.

33. One interesting aspect of surveys is that the order of the questions can affect the results. This is thought to occur because one question can alter the response to a subsequent question. In 1948 many people in the United States were concerned about the spread of communism as a political and economic philosophy. A survey given that year contained both of the following questions:

1. Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?

2. Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?

The following table contains the percentages of the Yes/No answers given by respondents when asked these questions in different orders.

Responses to reporter question in 1948

Question order	% Yes	% No
Question #1 asked first	37	63
Question #2 asked first	73	27

a) Construct a comparative bar chart for these response percentages.

b) Interpret the results of the 1948 survey. Do you think the order of the questions made a difference in the responses? Explain.

ANSWER: a) comparative bar chart:

Chapter 2 - Graphical Methods for Describing Data Distributions
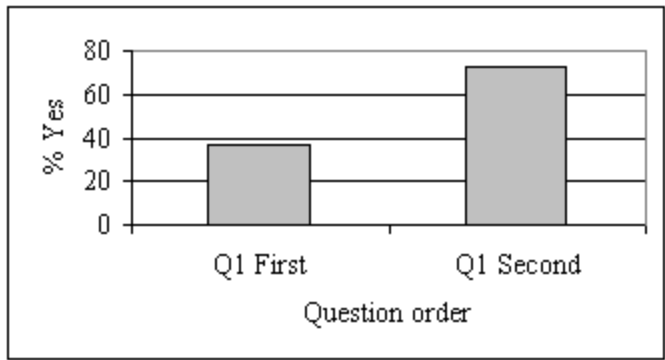
b)   Yes, the order made a difference. The rate of "yes" responses was very nearly doubled if the question was asked after the one about American reporters in Russia.

34. Biologists know that bats have different flight characteristics. Some bats exhibit fast, straight, and long flights; others are moderately slow, and highly maneuverable. The flight capabilities of bats are presumably a function of (a) the characteristics of the bat wings, and (b) the laws of physics. The data presented below are wing weights in grams for certain species of bats from Nicaragua and Costa Rica. Two categories of bat species are represented: fruit-eaters and insect-eaters.
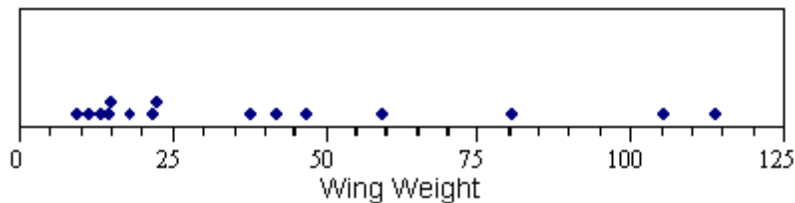
Fruit eaters:
9.9, 11.5, 47.0, 59.6, 18.2, 22.1, 22.9, 15.0, 15.4, 13.3, 38.3, 42.2, 114.3, 105.4, 80.9
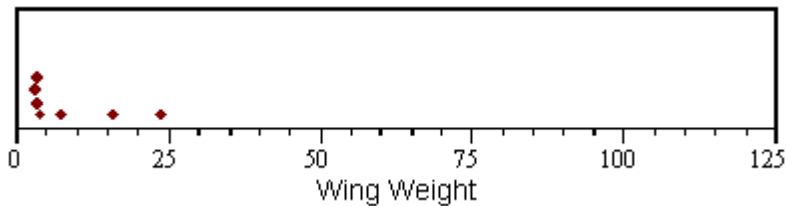
Insect eaters:
23.8, 16.1, 4.2, 3.9, 7.5, 3.5, 3.9

(a)   Construct two dotplots that will allow you to easily compare the distributions of the wing weights for the two categories of bats, fruit-eaters and insect-eaters.

(b)   Write a short paragraph describing the features of the two plots; what do these plots tell you about the differences in wing weight for fruit- and insect-eaters?

ANSWER: a)   Fruit Eaters wing weight:

Insect Eaters wing weight:

b)   The fruit eating bats have wings that are generally heavier than those of the insect eating bats. They are also much more spread out in weight, ranging from 9.9 to 114.3 grams. The insect-eating bats only have a range of 3.5 to 23.8 grams.

Chapter 2 - Graphical Methods for Describing Data Distributions

Exhibit 2-1
In a story written in December 2002, the *Des Moines Register* reported on the average number of years of teaching experience for the elementary schools in Des Moines, Iowa. The data, by elementary school, appears in the table below.

School	Years Exper	School	Years Exper	School	Years Exper	School	Years Exper
Adams	13.8	Hanawalt	12.7	McKee	17.5	Phillips	10.5
Capitol View	9.7	Hillis	17	McKinley	5.6	Pleasant Hill	12.8
Cattell	8.5	Howe	14.9	Mitchell	6.2	River Woods	5.9
Douglas	13.8	Hubbell	15.1	Monroe	16.2	Stowe	16.5
Downtown	7.0	Jackson	16.8	Moore	16.6	Studebaker	18.4
Edmunds	8.4	Jefferson	19.2	Morris	9.9	Wallace	4.8
Findley	14.9	King	7.0	Moulton	8.6	Willard	9.7
Garton	9.3	Longfellow	9.8	Oak Park	11.4	Windson	17.7
Granger	17.2	Lovejoy	9.9	Park Avenue	11.4	Woodlawn	17
Greenwood	14	Madison	20.8	Perkins	11.2	Wright	12.9

35.
(a) Construct a histogram for these data, using the following class intervals:

0 to < 4
5 to < 8, etc.

(b) Describe the distribution of the average number of years of teaching experience, and note any unusual features.

ANSWER:

a)

b) This is a fairly symmetrical distribution with no apparent outliers. If anything it is slightly skewed toward the higher values.

Exhibit 2-2
In a story written in December 2002, the *Des Moines Register* reported on the percentages of students who qualify for Free/Reduced lunch programs in the elementary schools in Des Moines, Iowa. The data, by elementary school, appears in the table below.

School	% F/R	School	% F/R	School	% F/R	School	% F/R
Adams	64	Hanawalt	14	McKee	55	Phillips	27
Capitol View	78	Hillis	35	McKinley	78	Pleasant Hill	17
Cattell	55	Howe	49	Mitchell	64	River Woods	48
Douglas	45	Hubbell	31	Monroe	52	Stowe	56
Downtown	8	Jackson	60	Moore	42	Studebaker	30
Edmunds	83	Jefferson	5	Morris	54	Wallace	80

Chapter 2 - Graphical Methods for Describing Data Distributions

| Findley	64	King	71	Moulton	88	Willard	82
Garton	53	Longfellow	84	Oak Park	59	Windson	37
Granger	52	Lovejoy	66	Park Avenue	38	Woodlawn	40
Greenwood	38	Madison	58	Perkins	54	Wright	29

36. Refer to Exhibit 2-2.

(a)   Construct a histogram for these data, using the following class intervals:

    0 to < 20
    20 to < 40, etc.

(b)   Describe the distribution of Free/Reduced eligibility percents, and note any unusual features.

ANSWER:

a)
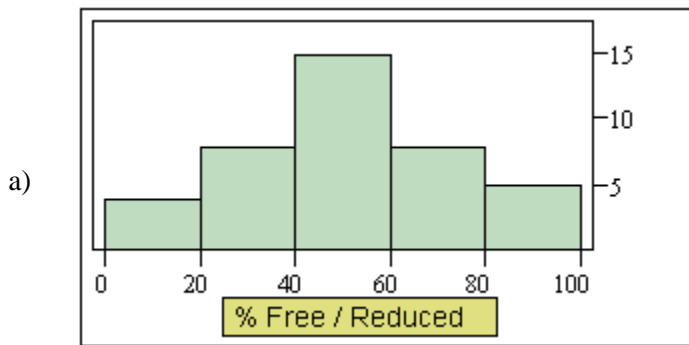
b)   This distribution is very symmetrical, with no outliers or any skewness. It could be a normal distribution.

37. Ruminants are hoofed animals such as cattle and sheep. Two categories of ruminants are browsers (leaf eaters) and grazers (grass eaters). Some biologists believe that the incisor arcade, that part of the mouth between the incisors, is shaped differently for browsers and grazers, and that might explain their food selection. The "curvature" of an incisor arcade is a measure of its shape, with large values indicating a long, thin mouth, and small values a shorter, wider mouth. The back-to-back stem plot below compares the curvatures of incisor arcades of a sample of species of browsers and grazers in the ruminant family.

Curvature, Browsers vs. Grazers

Browsers   Grazers
N = 18   *N* = 54

```
|               1L    |
|               1H    | 9
|               2L    | 1
4 |              3L    | 22244
99885 |          3H    | 55566667778999999
410 |            4L    | 022223
99987 |          4H    | 55556666677778889
321 |            5L    | 03
|               5H    | 9
|               6L    | 0
|               6H    |
|               7L    |
```

Chapter 2 - Graphical Methods for Describing Data Distributions

0 | 7H |

(a) Briefly describe each of the two curvature distributions, browsers and grazers, noting any unusual features of the data.

(b) The theory of some biologists is that browsers have thinner mouths because the browse plants frequently have thorns and spines; browsers have evolved to have thinner mouths to get beyond the thorns and spines to the food part of the plant. Is the data in the back-to-back stem plot consistent with this theory? What aspect of the data would you consider when answering this question?

ANSWER:

a) The distribution of browsers is centered in the low 40 range. Other than one prominent outlier at 70, the distribution is approximately symmetrical. The distribution of browsers is likewise centered in the low 40 range and is quite symmetrical, but more spread out than the browsers if the outlier is excluded.

b) The evidence to support this theory is not very strong. The browsers may have a slightly higher average curvature, but the two distributions overlap quite a lot, indicating that the differences are modest at best.

38. An experiment was performed to determine which brand of comparably priced light bulbs, A or B, lasts longer on average. Ten bulbs of each brand were tested, and their lifetimes recorded in hours. The data are shown below:

Bulb A: 88, 97, 110, 122, 128, 129, 135, 141, 150, 165
Bulb B: 23, 55, 93, 106, 111, 115, 123, 128, 153, 171

(a) Display these data using a comparative (back-to-back) stemplot.

(b) What are the differences and similarities in the lifetime distributions for the two samples?

(c) If your preference for brand A or B must be determined from the information in parts (a) and (b), which brand would you buy? Explain your choice using appropriate statistical terminology.

ANSWER: a)   Stemplot: Bulb A vs. Bulb B

Bulb A (hrs)	Stem	Bulb B (hrs)	
	17	1	
5	16		
0	15	3	
1	14		
5	13		Stem: tens
9 8 2	12	3 8	Leaf: ones
0	11	1 5	
	10	6	
7	9	3	
8	8		
	7		
	6		
	5		
	4		
	3		
	2	3	
	1		

Chapter 2 - Graphical Methods for Describing Data Distributions

|0 |

  b) The two distributions are similar in being reasonably symmetric. The distribution for Bulb A is less widely spread out, and the values are generally a little higher than for Bulb B.

  c) I would buy Bulb A, since the center of the distribution is higher and the variability is lower, so I think I could count on getting a quality bulb with A more so than with B.

39. The stem plot below displays the weights (in ounces) of a random sample of tomatoes grown on a local farm

Weights of Tomatoes (oz)

```
2L |        4
2H |
3L |        2
3H |        89              stem: ones
4L |        13              leaf: tenths
4H |        5589
5L |        11122334
5H |        668999
6L |        223344
6H |        5556
```

(a) Briefly describe the distribution of tomato weights.

(b) What is the weight of the heaviest tomato in the sample?

ANSWER: a) This is moderately skewed in the negative direction, with most of the weights bunched in the higher 5 categories and fewer in the lower 5 weight categories.

    b) The heaviest tomato is 6.6 ounces.

40. Bat characteristics

Biologists know that bats have different flight characteristics. Some bats exhibit fast, straight, and long flights; others are moderately slow, and highly maneuverable. The flight capabilities of bats are presumably a function of (a) the characteristics of the bat wings, and (b) the laws of physics. The data presented below are wing spans in centimeters for certain species of bats from Nicaragua and Costa Rica. Two categories of bat species are represented: fruit-eaters and insect-eaters.

Fruit eaters:
25.2, 27.2, 42.0, 44.8, 31.3, 32.5, 32.0, 28.1, 30.7, 27.0, 41.1, 41.6, 58.4, 58.7, 53.9
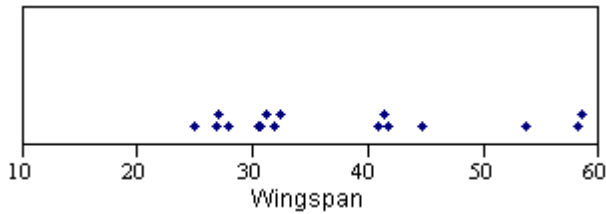
Insect eaters:
32.8, 27.4, 21.0, 23.9, 27.5, 22.4, 18.7

(a) Construct two dotplots that will allow you to easily compare the distributions of the wing spans for the two categories of bats, fruit-eaters and insect-eaters.

(b) Write a short paragraph describing the features of the two plots; what do these plots tell you about the differences in wing span for fruit- and insect-eaters?

ANSWER: a) Fruit Eaters wingspan:

Chapter 2 - Graphical Methods for Describing Data Distributions


Insect Eaters wingspan:
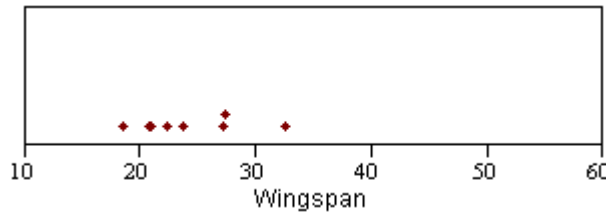
b) The wingspans for fruit-eaters are generally longer than for the insect-eating bats. The wingspans are also more spread out for the fruit-eaters, from 25.2 to 58.7 cm, compared to 18.7 to 32.8 for the insect-eaters.
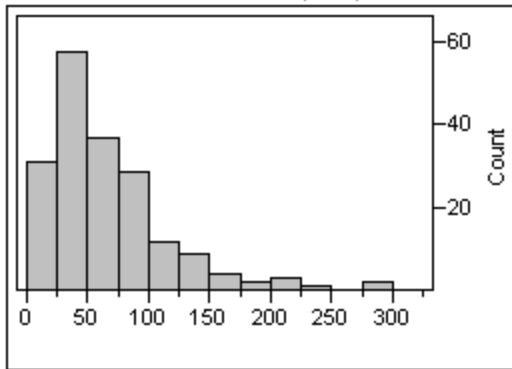
Exhibit 2-4
In the article, "Scaling of Mammalian Long Bones: Small and Large Mammals Compared," the average circumferences (distances around) in mm of the humerus, femur, and tibia of different species of mammals were reported. These are the long bones that primarily support the body mass in mammals. Each data value is the mean circumference for the available measures for a single species.
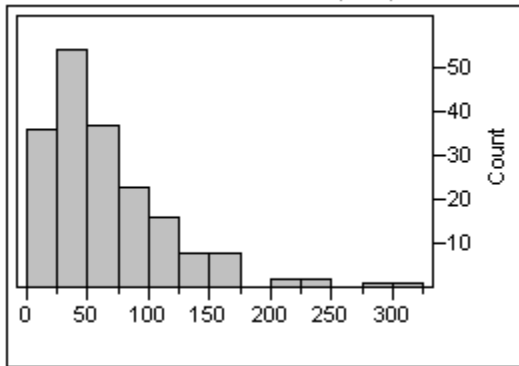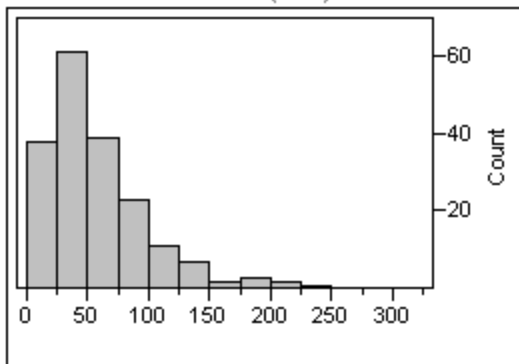
Chapter 2 - Graphical Methods for Describing Data Distributions

41. Refer to Exhibit 2-4.

Describe the differences and similarities in the distributions of these different mammalian bones.

ANSWER: The shapes of these distributions are very similar, with a peak between 25 and 50 and positively skewed. The humerus and femur extend a little higher than the tibia, to about 300 vs. 200.

42. Refer to Exhibit 2-4.

Biologists speculate that the different amounts of weight might be supported by different sized bones. Based on the histograms, does it seem like one of these bones has on average a greater or lesser circumferences? Explain.

ANSWER: The differences are not great, since the large parts of the distributions are very similar. Perhaps for the largest animals the humerus and femur support more of the weight, indicated by the fact that these distributions extend to higher values than does the tibia distribution.

43. In birds, the muscles that pull the wing down, and thus cause the actual flight, are the "large pectorals." In the article, "Dimensional Relationships for Flying Animals," the author reports the following pectoral muscle weight in grams for birds of prey (*Rapaces diurnes voiliers.*)

958, 715, 299, 476, 223, 270, 194, 226, 113.5, 79.8,
149, 84.2, 73.8, 53.8, 36, 48, 151

(a)   Construct a stem and leaf plot of these data.

(b)   In a few sentences, describe the distribution of the pectoral muscle weights.

ANSWER: a)   stem and leaf:

Pectoral muscle weights (grams):

9		58
8		
7		15
6		
5		
4		76
3		
2		23,26,70,99
1		13,49,51,94
0		36,48,54,74,80,84

stem: hundred
leaf: ones

b)   This is a skewed distribution, with most weights under 200 grams, but extending up almost to 1000 grams. The highest values are extreme outliers.

44. In the article, "Dimensional Relationships for Flying Animals," the wing areas in square centimeters of gliding birds of prey are reported. Night hunters are distinguished from the birds that hunt during the daylight hours, and the wing areas for samples of each type of bird are displayed in the following table:

Wing area of gliding birds of prey (sq. cm.)

Night hunters:	3715	1082	1396	405	1163	1304	459
Day hunters:	1317	822	530	2321	708	703	1285

Chapter 2 - Graphical Methods for Describing Data Distributions

  558   438

a) Construct a <u>comparative</u> (back to back) stem & leaf plot for the wing area for the two types of birds.

(b) Biologists speculate that since birds could be detected at a distance by their prey during the day, the species that hunt during the day should be smaller on average than those that hunt at night. Are the data as displayed in part (a) consistent with this theory? What specific aspects of the plot in part (a) support your answer?

ANSWER: a)   Back to back stem and leaf:

Day hunter wings (cm^2)   Night hunter wings (cm^2)

stem: thousands	3H	715
leaf: units	3L	
	2H	
321	2L	
	1H	
317,285	1L	082,163,304,396
822,708,703,558,530	0H	
438	0L	405,459

b) Yes, the majority of day hunters are below 1000 cm^2, while for night hunters the majority are above 1000 cm^2.

45. Iowa is an agricultural state with a large number of rivers. The use of chemical pesticides and increasing size of livestock herds has created a concern for the water quality in the state. A researcher recently surveyed the rivers of Iowa and analyzed the concentrations of certain chemicals at different locations in the watersheds of Iowa. The cumulative relative frequency plot below displays the distribution of Nitrogen concentrations along the Cedar River.

Cumulative frequency Distribution
mg/L Nitrogen
Cedar River
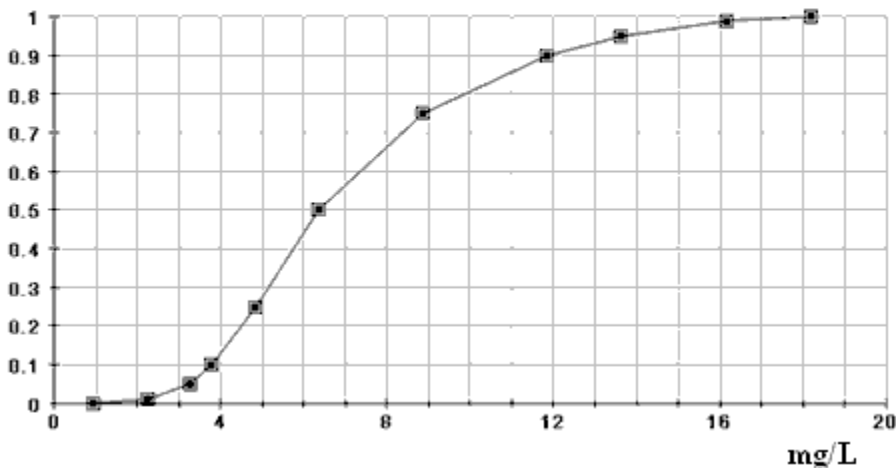
mg/L

(a) Approximately what percent of the Nitrogen concentration measurements are less than 10 mg/L?

Chapter 2 - Graphical Methods for Describing Data Distributions

(b)   Approximately what percent of the Nitrogen concentration measurements are 12 mg/L or more?

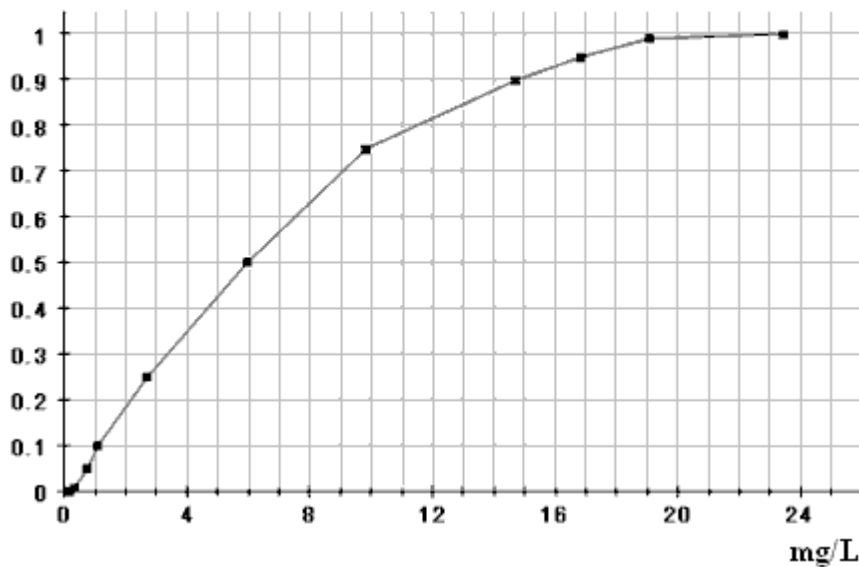(c)   50% of the Nitrogen concentration measurements lie below what approximate value?

ANSWER: a)   Approximately 80%

b)   Approximately 10%

c) $6 - 6.5$ mg / L

46. Iowa is an agricultural state with a large number of rivers. The use of chemical pesticides and increasing size of livestock herds has created a concern for the water quality in the state. A researcher recently surveyed the rivers of Iowa and analyzed the concentrations of certain chemicals at different locations in the watersheds of Iowa. The cumulative relative frequency plot below displays the distribution of Nitrogen concentrations along the Skunk River.

Cumulative frequency Distribution
mg/L Nitrogen
Skunk River

(a)   Approximately what percent of the Nitrogen concentration measurements are less than 10 mg/L?

(b)   Approximately what percent of the Nitrogen concentration measurements are 12 mg/L or more?

(c)   50% of the Nitrogen concentration measurements lie below what approximate value?

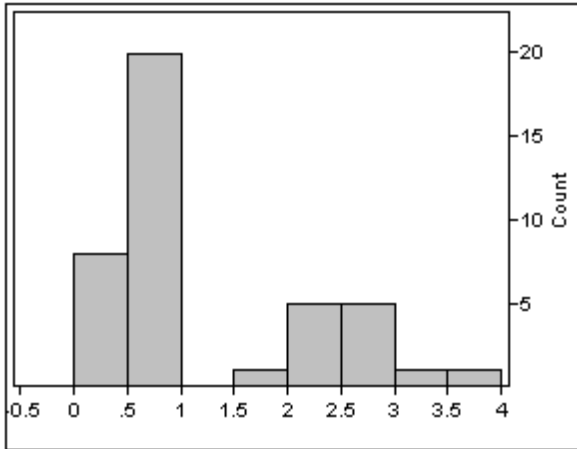*ANSWER:* a)   Approximately 75%

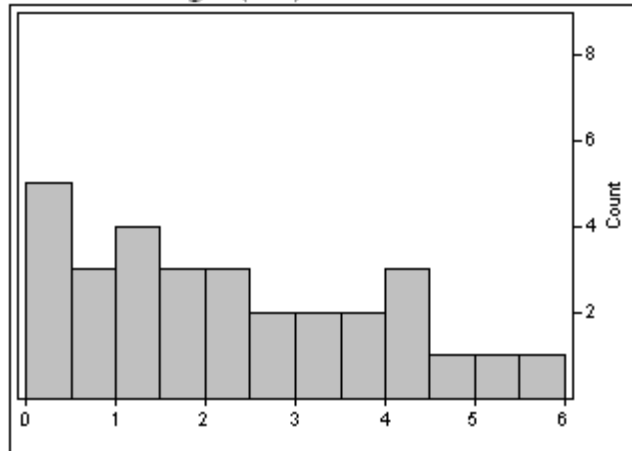b)   Approximately 20%

c)   6 mg / L

47. In 1986 marine biologists reported the discovery of a new species of fish in the Western North Atlantic. Data on the head length ($n = 41$) and pelvic fin length ($n = 30$) are shown below.

Chapter 2 - Graphical Methods for Describing Data Distributions

Head Length (mm)

Pelvic Fin Length (mm)

a) Describe in appropriate statistical terms, the distributions of these two variables.

b) The distribution of head lengths certainly appears to be rather strange. What do you think might account for the shape of this distribution?

ANSWER:

a) The distribution of head length is bimodal, with clumps centered at .5mm and 2.5mm, and a gap between these two clumps. The pelvic fin length is skewed to the right over a range from 0 to 6 mm.

b) Generally one would not expect bimodal measurements unless perhaps there were a mixture of two subpopulations. In this case perhaps males are larger or smaller, or there may be two subspecies of fish that make up the identified species.

Exhibit 2-3
There has been a recent increase in the popularity of caffeine-containing "energy drinks" or "functional beverages"-- substances that are considered to be food that may provide some health benefit. However, caffeine in these drinks presents an important health concern, especially for women of childbearing age. Studies have shown an increased risk of spontaneous abortion and low birth weight children to women consuming more than 150 mg/day of caffeine. In a recent study, eighteen 12-oz servings of different brands of carbonated sodas were analyzed to ascertain their caffeine content, with the results presented below. The data are in mg/12-oz serving; if no caffeine was detected, this is reported as "*

Chapter 2 - Graphical Methods for Describing Data Distributions

mg/serving."

Caffeine (mg) in 12 oz servings of Carbonated Sodas

29.5	38.2	39.6	*	29.5	31.7
27.4	45.4	48.2	36.0	33.8	*
19.4	*	*	18.0	34.6	*

48. Refer to Exhibit 2-3.

(a)   Construct a stem and leaf plot of the caffeine-content for the <u>detected</u> amounts of caffeine.

(b)   In a few sentences, describe the distribution of the caffeine-content. How would your description change if the "none detected" were treated as 0.0 mg/serving?

ANSWER: (a)   stem and leaf:

Caffeine (mg)

```
1 |
1 |                      80,94
2 |                                  Stem: tens
2 |                      74,95,95    Leaf: units
3 |                      17,38,46
3 |                      60,82,96
4 |
4 |                      54,82
```

(b)   This distribution is fairly uniformly distributed with small gaps near each end.

49. Refer to Exhibit 2-3.

a)   Construct a <u>comparative</u> (back to back) stem & leaf plot for the container vs. fountain amounts of caffeine. (Do not use the data for carbonated sodas where caffeine was not detected.) Describe how the two distributions compare.

(b)   What do you think might explain the differences between the two distributions?

ANSWER: a)   Back to back stem and leaf:

Fountain Container

```
                    |  1   |
                    |  1   | 80,94
                    |  2   |
                    |  2   | 74,95,95        Stem: tens
   41,30,26,11,11,07 |  3   | 17,38,46        Leaf: units
         63,63,52 |  3   | 60,82,96
                    |  4   |
                    |  4   | 54,82
```

b)   There does not appear to be an appreciable difference in location. The difference in variability perhaps indicates a difference in

quality control over the fountain syrup process than the container process. Perhaps there are more steps in the container process, or perhaps the raw materials are more variable in the container process.

50. Artificial bird nests have been used extensively when studying predator behavior. The assumption is that predators will find and respond to artificial nests in the same way they do to the natural nests. In a study to test this hypothesis the predation of 27 American Robin (*Turdus migratorius*) nests and 27 artificial wicker-basket nests were compared. Artificial nests were placed 30m from existing robin's nests in a randomly chosen direction. The natural nests and artificial nests were checked at 5-day intervals, and the number of nests (out of 27) that had been disturbed by predators in the interval was noted. The results are as follows:

Numbers of real and artificial Robin nests predated

Nest type	5-day check	10-day check	15-day check
Natural	12	17	19
Artificial	3	12	18

(a) Construct a time series plot for the number of natural nests disturbed by predators. Use the same set of axes to plot the number of artificial nests disturbed by predators.

(b) Does it appear that the artificial and natural nests have the same pattern of predation? Justify your answer by appealing to specific aspects of your plot in part (a).

ANSWER:

a)
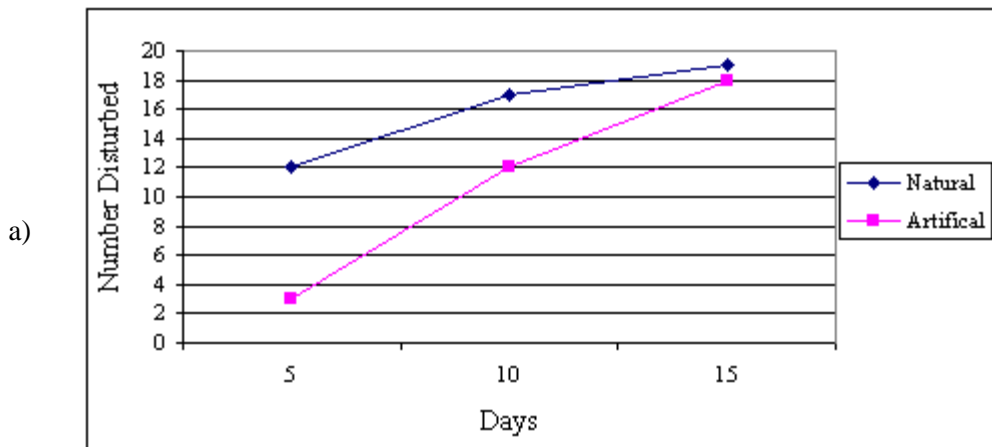
b) The artificial nests appear to be less likely to disturbed by predators. At 5 days far fewer have been disturbed, but over the course of the 15 days they almost catch up to the rate for the natural nests.

51. Records of the missions of Baja California in the form of baptisms and burial records provide an ideal opportunity to examine the consequences of what is known as the "Colombian Exchange," the movement of organisms large and small between the Old and New Worlds after the appearance of Columbus in 1492. The data below are two time series of the numbers of baptisms at two Central Baja missions, San Jose de Comondu and San Gertrudis. The data are for the decade 1770 - 1779.

Chapter 2 - Graphical Methods for Describing Data Distributions
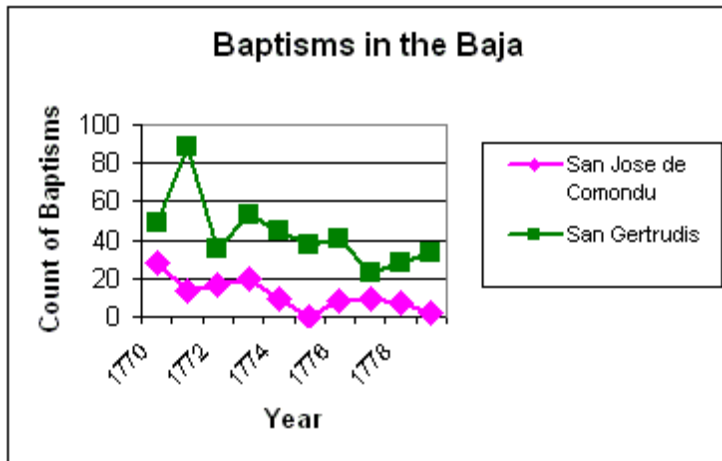
(a) How would you generally describe the trend of baptisms for the two missions over this period? Do the missions appear to show similar trends? Justify your response by appealing to the graph.

(b) One of the influences of the meeting of the Old and New Worlds was the transmission of diseases to the native populations, who had developed no natural immunity from European diseases. If there were anecdotal evidence for an outbreak of disease during this time, which mission and what year seems to be the most likely candidate? Again, justify your response by appealing to the graph.

ANSWER:

(a) Generally the trends seem to be parallel, except for an apparent greater variability at San Gertrudis in the 1770 - 1773 period. The two curves run roughly parallel except for those initial years of the decade.

(b) The most likely candidate is the San Gertrudis mission in 1772. There is a precipitous drop in the number of baptisms in the 1771/1772 time frame, which would be consistent with a population loss due to disease. (Another possibility is that so many baptisms occurred in 1771 there weren't many people left over for the 1772 baptisms.)

52. In a story written in December 2002, the *Des Moines Register* reported on the percentages of students who qualify for Free/Reduced lunch programs in the elementary schools in Des Moines, Iowa. The data, by elementary school, appears in the table below.

School	% F/R	School	% F/R	School	% F/R	School	% F/R
Adams	64	Hanawalt	14	McKee	55	Phillips	27
Capitol View	78	Hillis	35	McKinley	78	Pleasant Hill	17
Cattell	55	Howe	49	Mitchell	64	River Woods	48
Douglas	45	Hubbell	31	Monroe	52	Stowe	56
Downtown	8	Jackson	60	Moore	42	Studebaker	30
Edmunds	83	Jefferson	5	Morris	54	Wallace	80
Findley	64	King	71	Moulton	88	Willard	82
Garton	53	Longfellow	84	Oak Park	59	Windson	37
Granger	52	Lovejoy	66	Park Avenue	38	Woodlawn	40
Greenwood	38	Madison	58	Perkins	54	Wright	29

The data presented in the table about the Des Moines, Iowa, elementary schools was from an article that related the percent of children eligible for Free and Reduced lunch and teacher experience. The percent of children eligible for Free/Reduced lunch is generally regarded as an indicator of the socio-economic status of a neighborhood. The concern of the *Des Moines Register* in the article was that the least experienced teachers typically teach students from low-income

Chapter 2 - Graphical Methods for Describing Data Distributions

families. The accompanying scatter plot shows the data from the article.

YearsExp Vs. FreeReduced%
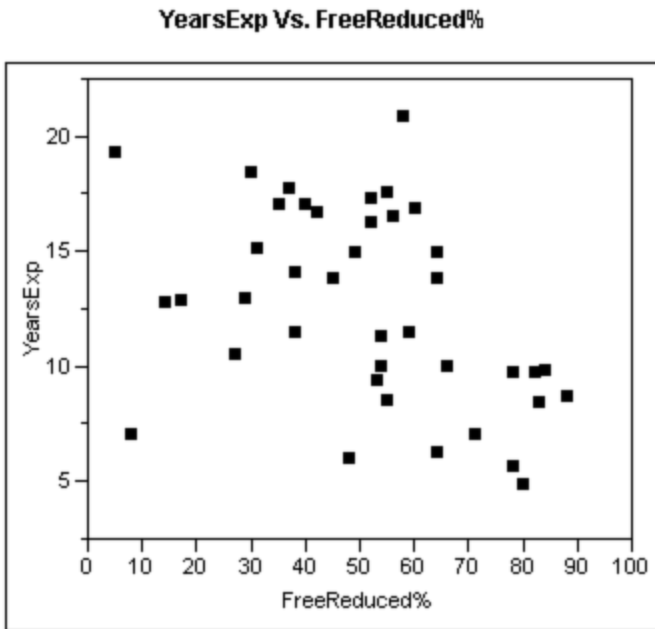
(a) If the *Des Moines Register* is correct, what sort of pattern should "ideally" appear in the scatter plot?

(b) Does this pattern emerge in these data? (Hint: Consider separately those schools with less than 75% eligibility, and those with greater than 75% eligibility for Free/Reduced lunch.)

ANSWER:

a) If the newspaper is correct, then schools with lower-income families would have both a higher rate of free lunches and a lower level of teacher experience. This would show in the graph as a decreasing trend: as the percentage of free lunches increases, the average years experience should decrease.

b) This is shown in the graph. Those schools with more than 75% eligibility for free lunches have less experience (no more than 10 years), while the schools that have less than 75% eligibility have average experience values up to more than 20 years. However, this trend is only moderate at best.

53. Investigators writing in the *Journal of Early Adolescence* were interested in the communication about sex and related topics between parents and 12-14 year olds. The investigators asked students and parents whether or not they had communicated with their child/parent about topics such as pregnancy, how to say "no," and methods of contraception. A scatter plot of their data is presented below. Each point represents one topic listed by the investigators. For example, 72.4% of parents and 57.6% of adolescents said the parents had discussed childbirth. This would be recorded as the point (57.6, 72.4).

Chapter 2 - Graphical Methods for Describing Data Distributions

Parents vs. Adolescents reporting of Discussion:
Sex Education Topics (% Yes)
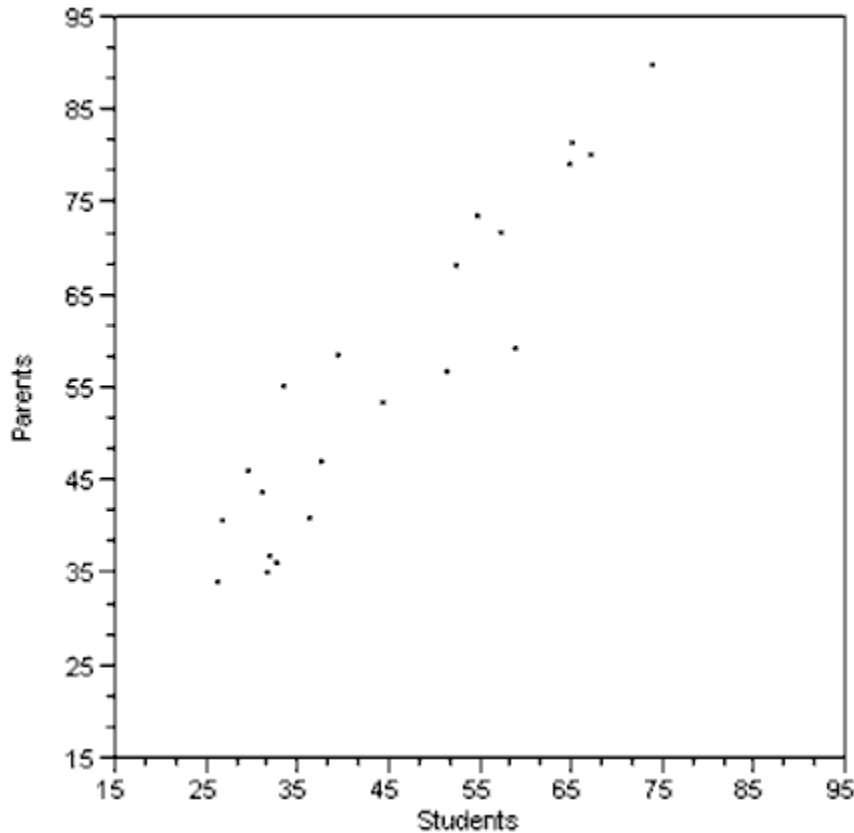
(a) On the graph above, sketch the line $y = x$, representing a line of "perfect agreement" between the parents and adolescents.

(b) What explanation do you have for the placement of the points in the scatter plot relative to the line you sketched in part (a)?

ANSWER: a) Draw diagonal line from (15,15) at lower left to (95,95) at upper right.

b) All the points are above the line, indicating that for all topics, parents are generally saying that they have talked about the topics more so than indicated by the adolescents' response.

54. Investigators studying the relation between gender and the interests of early adolescents asked boys and girls in grades 5-9 to indicate on a 7-point scale their degree of interest in topics such as Life, Sports, and Cars.

The data in the scatter plot represent average levels of interest for boys and girls in grades 5 - 9. Each point represents one topic listed by the investigators. For example, boys average rating for "Opposite sex" as a 6.2, while girls on average rated their interest in this topic as 6.0 on the 7 point interest scale. This is recorded as the point (6.2, 6.0).

Chapter 2 - Graphical Methods for Describing Data Distributions
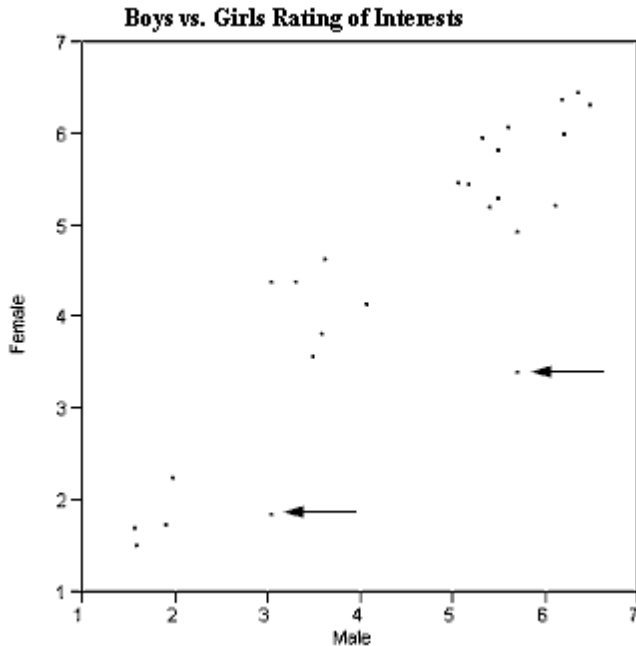
There are three interesting features of this scatter plot that you are to interpret below. What can you infer from each of these features of the scatterplot?

(a)   The points generally seem to scatter around the line, $y = x$.

(b)   There seem to be three clusters of points, one in the lower left, one in the middle, and one in the upper right of the graph.

(c)   There are two points (as indicated by the arrows) that do not seem to fit the overall pattern. What can be said about the topics these points represent?

ANSWER: a)   (Draw diagonal line $y = x$ from (1,1) at lower left to (7,7) at upper right.) This means that the average interest for topics is about the same for boys and girls.

   b)   Boys and girls seem to agree on most areas of low interest (lower left cluster), medium interest (middle cluster) and high interest (upper right cluster).

   c)   These points indicate that there are two topics for which the interest levels for boys and girls differ sharply. For both of these topics, boys indicated a higher interest than girls.

55. The pie chart below summarizes the results of a survey of 300 randomly selected students at a particular high school. The investigators asked about soft drink preferences at a local high school.

<u>Chapter 2 - Graphical Methods for Describing Data Distributions</u>
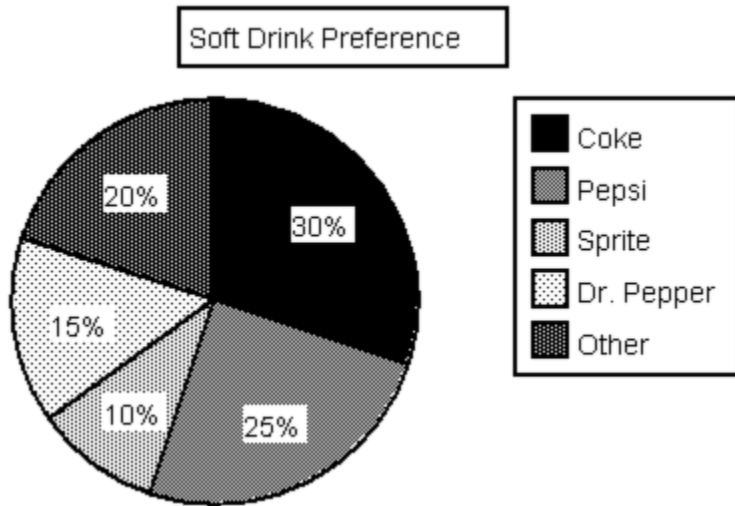
(a)   How many out of the 300 students indicated a preference for Dr. Pepper?

(b)   Write a few sentences summarizing the soft drink preference for this sample of students.

(c)   Using the information in the pie chart above, sketch a relative frequency bar chart for these data.

ANSWER: a)   45 students (.15 × 300)

  b)   More students preferred Coke than any other drink, and colas (Coke and Pepsi) account for more than half of the preferred drinks.

  c)   Relative frequency bar chart:
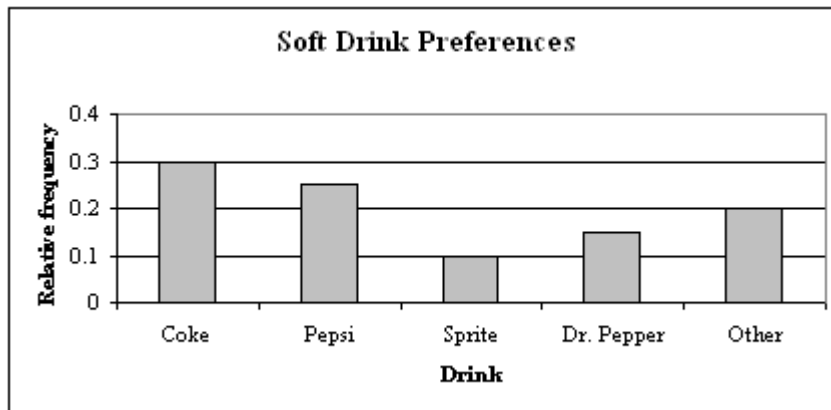
56. One of the "media messages" sent by television is that appearance is a very important aspect of one's being. These messages are thought to spur young women to have unrealistic expectations about beauty, thus spending billions of dollars on cosmetics and toiletries. Some of these messages were studied during the 1999-2000 prime-time television season, when investigators kept track of the "appearance" comments. Some of their data, comments by males, classified by gender of recipient and type of appearance comment, is reproduced in the table below.

Insults, Compliments, and Neutral Appearance Comments
(Males commenting)

	Males	**Males**

Chapter 2 - Graphical Methods for Describing Data Distributions

Nature of comment	about males	about females
Compliments	52	106
Neutral comments	60	35
Insults	41	15
Totals	153	156

(a) Construct two pie charts, one for comments about males, one for comments about females, to summarize the data above.

(b) Using the information in the table above and the pie charts in part (a), display the given information in a comparative bar chart.

In the same study of appearance comments mentioned, comments by females were tabulated. That data appears below:

Insults, Compliments, and Neutral Appearance Comments
(Females commenting)

Nature of comment	Females about males	Females about females
Compliments	49	79
Neutral comments	26	57
Insults	21	32
Totals	96	168

(c) Display these data in a <u>segmented</u> bar chart to compare the distribution of comments in these two samples.
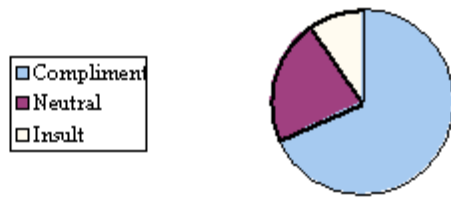
(d) Based on the information from the graphs in parts (a) - (c), how do the distributions differ when men rather than women are <u>making</u> the comments, and how do they differ when men rather than women are the <u>recipients</u> of the comments?

ANSWER: a) Comparative pie charts for comments about males and females:
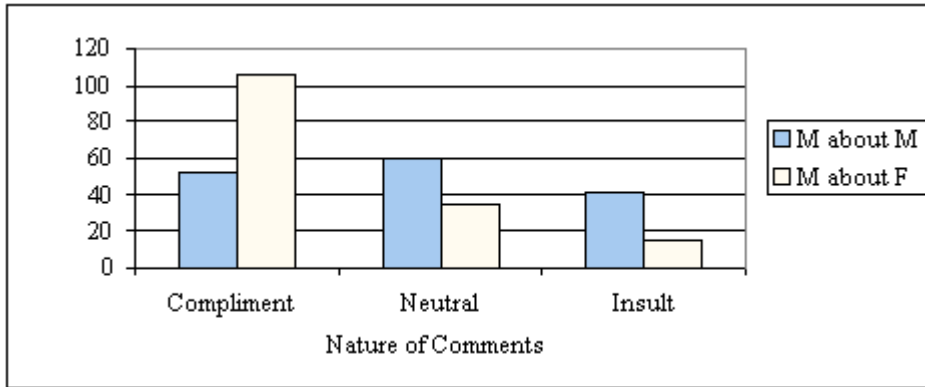

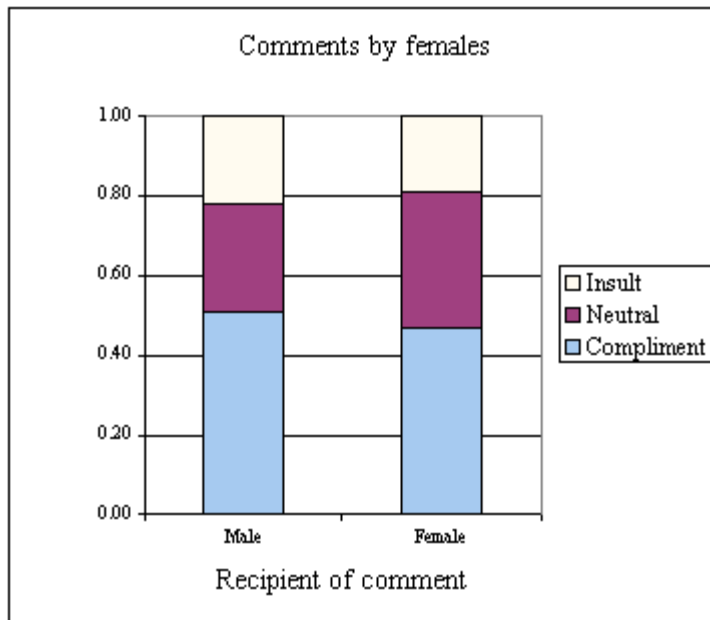Males about males   Males about females

☐ Compliment
■ Neutral
☐ Insult

b) Comparative bar chart for comments about males and females:

Chapter 2 - Graphical Methods for Describing Data Distributions

c)

For the gender of the <u>maker</u> of the comments: Men's comments differ sharply depending upon the gender of the recipient. Men making comments are much more likely to be complimentary to a female, and rarely offer insulting comments, whereas men are much more likely to be critical of other men. Women who make comments do so with about the same pattern whether the recipient is male or female.

d)

For the gender of the <u>recipient</u> of the comments: Women recipients are much more likely to get compliments than an insult. For male recipients, the three categories are about evenly distributed, meaning a male can expect an insult about as frequently as a compliment.

57. In a survey of 30 working adults who earn less than $6,000 per year, investigators asked about job satisfaction and income. The responses are summarized in the following table:
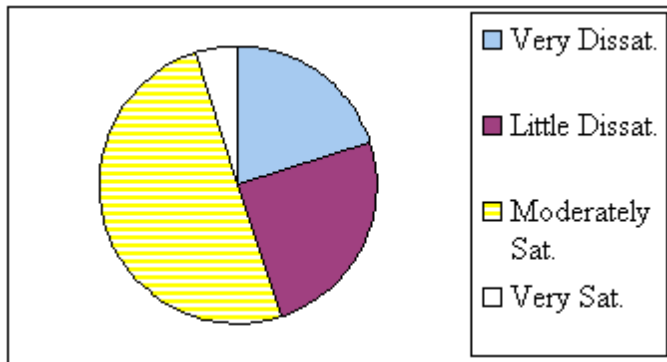
Job Satisfaction of < $6,000 annual income

| Very Dissatisfied	Little Dissatisfied	Moderately Dissatisfied	Very Dissatisfied
60	75	150	15

Construct and sketch a pie chart for these data.

Chapter 2 - Graphical Methods for Describing Data Distributions

ANSWER: pie chart:

58. In recent years there has been a change in the grocery shopping habits of shoppers. "Non-traditional" stores such as gas stations and Wal-Marts are now selling groceries. The stacked bar chart below shows the percentages of market share for Traditional, Non-traditional, and Convenience stores for 1988 and 2005.
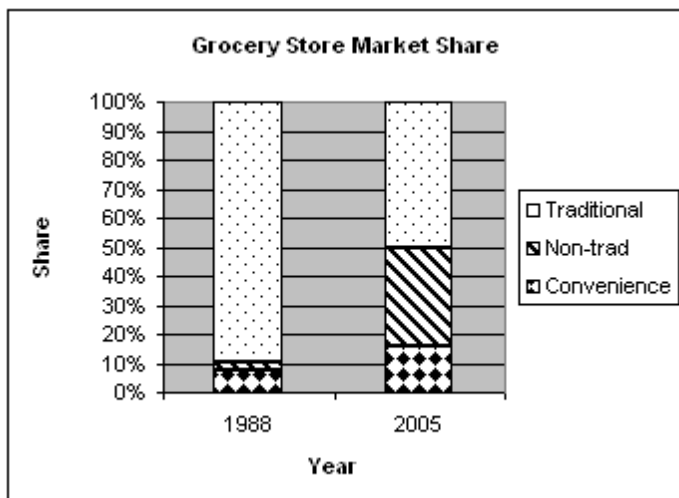
(a)   In a few sentences describe how the proportions of market share have changed for traditional grocery stores from 1988 to 2005.

(b)   The market share for which type of store has grown the most, compared to their market share in 1988. What aspect of the chart shows this?

ANSWER:

     a)   The market share for the traditional grocery stores has declined markedly from about 90% to 50%. Most of this market share went to the non-traditional grocery scores, with convenience stores going from a bit under 10% to about 15% market share.

     b)   The non-traditional stores have grown the most. The ratio of lengths (or areas) in the graphs for the non-traditional stores for the two years appears to be at least 5.

59. Each person in a sample of 800 students who identified themselves as Democrats was asked about his or her mother's political party. The responses are summarized in the following table:

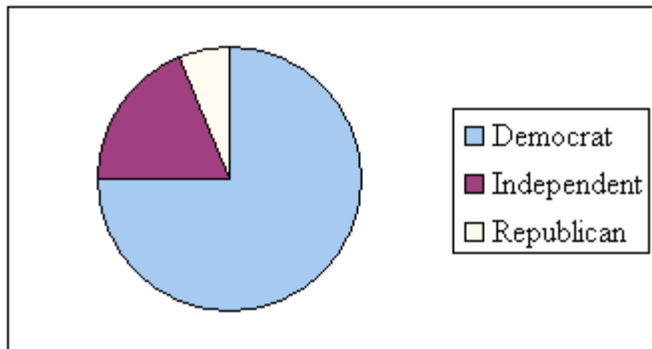**Mother's political identification of student Democrats**

Democrat	Independent	Republican

Chapter 2 - Graphical Methods for Describing Data Distributions

| 600 | 150 | 50 |

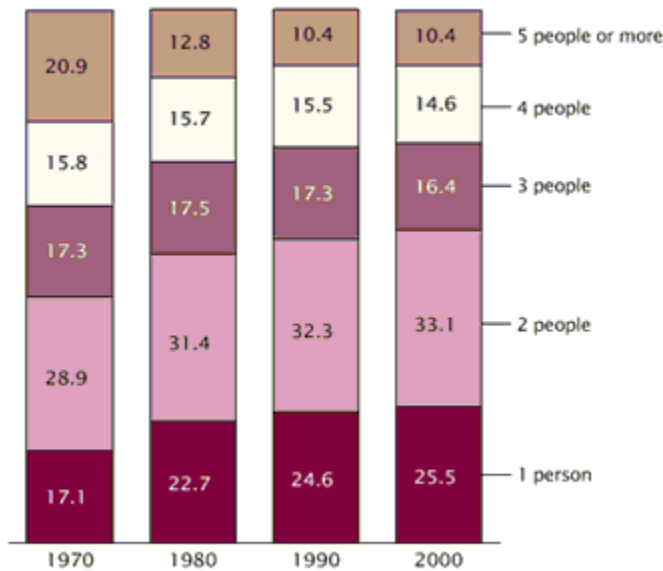Construct and sketch a pie chart for these data.

ANSWER: pie chart:

60. As part of the United States Census, data is also collected on the number of persons in each household. The census data for four decades is summarized below.

Households by Size: Selected Years, 1970 to 2000
(Percent distribution)

Source: U.S. Census Bureau, Current Population Survey, March Supplements: 1970 to 2000.

(a) In a few sentences describe how the proportion of households with 4 people has changed from 1970 to 2000.

(b) What size of household appears to have decreased the most from 1970 to 2000?

ANSWER:
a) The proportion of households with 4 people has decreased slightly, from 15.5% of households to 14.6%.

b) The largest households, with 5 or more members, decreased sharply, from 20.9% to 10.4% during the 30 years. This is a decrease of more than 50%. The other households either grew in size or only decreased slightly.

Chapter 2 - Graphical Methods for Describing Data Distributions

Multi-Mode

61. A statistics company made a survey about using smartphones or regular cell phones versus the age group in one U.S. State. For each of the six age groups (younger than 18 years old, 18- to 24 years old, 25- to 39 years old, 40- to 59 years old, 60- to 69 years old, 70 years old and older) one was picked up hundred people were selected, which possessed either a smartphone, or a regular cell phone, or no possessed a cell phone at all.

Choose the more most appropriate graphical display for the statistical survey described above.

    a. Bar Chart

    b. Comparative Bar Chart

    c. Dotplot

    d. Stem-and-Leaf Display

    e. Histogram

ANSWER: b

62. According to the data from fec.gov, since the year 1980, in Texas candidates from the Republican Party won the presidential elections. Following table represents the results of elections since 1984.

Year	Presidential Candidate from the Republicans	Percent
1984	Ronald Reagan	63.61%
1988	George H.W. Bush	55.95%
1992	George H.W. Bush	40.56%
1996	Bob Dole	48.76%
2000	George W. Bush	59.30%
2004	George W. Bush	61.09%
2008	John S. McCain	55.45%
2012	Willard Mitt Romney	57.17%
2016	Donald J. Trump	52.23%

Choose the most appropriate graphical display for the results of elections in Texas.

    a. Bar Chart

    b. Comparative Bar Chart

    c. Time Series Plot

    d. Stem-and-Leaf Display

    e. Histogram

ANSWER: c

63. Professor wants to identify the relationship between the test scores in her class and the number of hours spent by student preparing to the test. She collected information about preparing time and the test scores in the table below.

#students	Time, h	Test score, max. 100 points
1	1.5	62
2	5.5	75
3	4.0	89

Chapter 2 - Graphical Methods for Describing Data Distributions

4	5.0	91
5	7.0	83
6	6.5	74
7	2.5	66
8	2.5	76
9	0.5	45
10	3.5	74
11	1.5	71
12	3.0	89
13	2.0	91
14	2.5	90
15	0.5	52
16	1.0	56
17	3.5	88
18	1.0	75
19	4.0	98
20	7.5	81
21	3.0	91
22	1.5	80

Choose the most appropriate graphical display for professor to see the relationship between preparing time and the test scores.

    a. Bar Chart

    b. Comparative Bar Chart

    c. Time Series Plot

    d. Scatterplot

    e. Histogram

ANSWER: d