

discussion of the Anscombe data in Section 12.9 serves as a strong reinforcement of the importance of residual analysis.

If time is available, you may wish to discuss the Durbin-Watson statistic for autocorrelation. Be sure to discuss how to find the critical values from the table of the D statistic and the fact that sometimes the results will be inconclusive.

Once the model fit has been found to be appropriate, inferences in regression can be made. First cover the t or F test for the slope by referring to the Excel or Minitab results. Here, the p -value approach is usually beneficial. Then, if time permits, you can discuss the confidence interval estimate for the mean and the prediction interval for the individual value.

The *Managing Ashland MultiComm Services* case, the Digital case, and the *Brynne Packaging* case each involves a simple linear regression analysis of a set of data.

Chapter 13

If time is available in the course, you can now move on to multiple regression. You should point out that Microsoft Excel or Minitab needs to be used to perform the computations in multiple regression. Once you have the results, you need to focus on the interpretation of the regression coefficients and how the interpretation differs between simple linear regression and multiple regression. Mention the aspects of multiple regression that are similar in interpretation to those in simple regression -- prediction, residual analysis, coefficient of determination, and standard error of the estimate.

If sufficient time is available, you can move on to the dummy variable model. With dummy variables, be sure to mention that the categories must be coded as 0 and 1. In addition, indicate the importance of determining whether there is an interaction between the dummy variable and the other independent variables. Further discussion can include interaction terms in regression models.

Both the *Managing Ashland MultiComm Services* case and the Digital case involve developing a multiple regression model that includes dummy variables.

To perform multiple regression, you can use In-Depth Excel, the Analysis ToolPak, PHStat, or Minitab. To perform logistic regression, you can use In-Depth Excel, PHStat, or Minitab.

Online Chapter 14

This chapter, which is only online and not in the text, can be downloaded from the text web site. In order to fully understand the role of statistics in quality management, the themes of quality management and Six Sigma need to be mentioned. Although students may wonder why this is either being discussed in a statistics class (or why they are reading non-statistical material), they usually enjoy learning about this subject since it provides a rationale for how the statistics course relates to management.

You may want to begin the discussion of control charts by demonstrating the Red Bead experiment. Tell the students that two broad categories of control charts will be considered attribute charts in Sections 14.2 and 14.4 and variables charts in Section 14.5.

Once this introduction has been completed, an overview of the theory of control charts can be undertaken. Begin by referring to the normal distribution and mention Shewhart's concern about committing errors in determining special causes. Tell the students that setting the limits at three standard deviation units away from the mean is done to insure that there is only a small chance that a stable process will have special cause signals that appear and cannot be explained. Continue the discussion by noting that the integer value 3 made computations simpler in an era prior to the availability of calculators and computers, and that experience has shown that this serves the purpose of keeping false alarms to a minimum.

Once these topics have been discussed, you are ready to begin covering specific control charts. The choice of where to start is an individual one. The simplest approach is to begin with the p chart and refer to the Red Bead experiment and then use other examples such as those shown in Section 14.2. Be sure that students are aware that Excel, PHStat, or Minitab can be used to construct the p chart. If time permits, you may wish to also cover the c chart. If you choose to do so, be sure to focus on the fact that the variable involved represents the number of nonconformities per unit (an area of opportunity).

The discussion of variables charts should begin with a review of the distinction between attribute and variables charts. Briefly discuss the decisions that need to be made when sample sizes are to be determined and subgroups are to be formed. Be sure to emphasize the fact that variables charts are usually done in pairs, one for the variability and the other for the mean. Emphasize the notion that if the variability chart is out of control, you will be unable to meaningfully interpret the chart for the mean. Again, note that Excel, PHStat, or Minitab can be used to construct both R and \bar{X} charts.

If time allows, you may wish to discuss the topic of process capability. This topic reinforces any previous coverage of the normal distribution. Be sure to go over the distinction between control limits and specification limits and the differences between the various capability statistics.

The themes of quality management and the inclusion of a discussion of the work of Deming and Shewhart allow you to distinguish between common causes of variation and special causes of variation. Perhaps the best way to reinforce this is by conducting the Red Bead experiment (see Section 14.3). This experiment allows the student to see the distinction between the two types of variation. The amount of time spent on Sections 14.7 and 14.8 is a matter of instructor discretion. Some may wish to just list the fourteen points and have students read the section, while others will want to cover the points in detail. Regardless of which approach is taken, in order to emphasize the importance of statistics, the Shewhart-Deming PDSA cycle needs to be mentioned since the study stage typically involves the use of statistical methods. In addition, points 6 (institute training on the job) and 13 (encourage education and self-improvement for everyone) underscore the importance of everyone within an organization being familiar with the basic statistical methods required to manage a process. Students find the experiment of counting F s (see Figure 14.9) particularly intriguing since they can't believe that they have messed up such a seemingly easy set of directions.

The importance of statistics can be reinforced by briefly covering the Six Sigma[®], an approach that is being used by many large corporations. Go over the DMAIC model and compare it to Deming's 14 points.

The *Harnswell Company Sewing Machine Company* case contains several phases and uses R and \bar{X} charts. The *Managing Ashland MultiComm Services* case also has several phases and uses the p chart and R and \bar{X} charts.

CHAPTER 1

- 1.1 The type of beverage sold yields categorical or “qualitative” responses.
The type of beverage sold yields distinct categories in which no ordering is implied.
- 1.2 Three sizes of U.S. businesses are classified into distinct categories—small, medium, and large—in which order is implied.
- 1.3 The time it takes to download a video from the Internet is a continuous numerical or “quantitative” variable because time can have any value from 0 to any reasonable unit of time.
- 1.4 (a) The number of cellphones is a numerical variable that is discrete because the outcome is a count.
(b) Monthly data usage is a numerical variable that is continuous because any value within a range of values can occur.
(c) Number of text messages exchanged per month is a numerical variable that is discrete because the outcome is a count.
(d) Voice usage per month is a numerical variable that is continuous because any value within a range of values can occur.
(e) Whether a cellphone is used for email is a categorical variable because the answer can be only yes or no.
- 1.5 (a) numerical, continuous
(b) numerical, discrete
(c) categorical
(d) categorical
- 1.6 (a) Categorical
(b) Numerical, continuous
(c) Categorical
(d) Numerical, discrete
(e) Categorical
- 1.7 (a) numerical, continuous *
(b) categorical
(c) categorical
(d) numerical, discrete
*Some researchers consider money as a discrete numerical variable because it can be “counted.”
- 1.8 (a) numerical, continuous *
(b) numerical, discrete
(c) numerical, continuous *
(d) categorical
*Some researchers consider money as a discrete numerical variable because it can be “counted.”

32 Chapter 1: Defining and Collecting Data

- 1.9 (a) Income may be considered discrete if we “count” our money. It may be considered continuous if we “measure” our money; we are only limited by the way a country's monetary system treats its currency.
(b) The first format is preferred because the responses represent data measured on a higher scale.
- 1.10 The underlying variable, ability of the students, may be continuous, but the measuring device, the test, does not have enough precision to distinguish between the two students.
- 1.11 (a) The population is “all working women from the metropolitan area.” A systematic or random sample could be taken of women from the metropolitan area. The director might wish to collect both numerical and categorical data.
(b) Three categorical questions might be occupation, marital status, type of clothing. Numerical questions might be age, average monthly hours shopping for clothing, income.
- 1.12 The answer depends on the chosen data set.
- 1.13 The answer depends on the specific story.
- 1.14 The answer depends on the specific story.
- 1.15 The transportation engineers and planners should use primary data collected through an observational study of the driving characteristics of drivers over the course of a month.
- 1.16 The information presented there is based mainly on a mixture of data distributed by an organization and data collected by ongoing business activities.
- 1.17 (a) 001 (b) 040 (c) 902
- 1.18 Sample without replacement: Read from left to right in 3-digit sequences and continue unfinished sequences from end of row to beginning of next row.
Row 05: 338 505 855 551 438 855 077 186 579 488 767 833 170
Rows 05-06: 897
Row 06: 340 033 648 847 204 334 639 193 639 411 095 924
Rows 06-07: 707
Row 07: 054 329 776 100 871 007 255 980 646 886 823 920 461
Row 08: 893 829 380 900 796 959 453 410 181 277 660 908 887
Rows 08-09: 237
Row 09: 818 721 426 714 050 785 223 801 670 353 362 449
Rows 09-10: 406
Note: All sequences above 902 and duplicates are discarded.
- 1.19 (a) Row 29: 12 47 83 76 22 99 65 93 10 65 83 61 36 98 89 58 86 92 71
Note: All sequences above 93 and all repeating sequences are discarded.
(b) Row 29: 12 47 83 76 22 99 65 93 10 65 83 61 36 98 89 58 86
Note: All sequences above 93 are discarded. Elements 65 and 83 are repeated.

- 1.20 A simple random sample would be less practical for personal interviews because of travel costs (unless interviewees are paid to attend a central interviewing location).
- 1.21 This is a probability sample because the selection is based on chance. It is not a simple random sample because A is more likely to be selected than B or C.
- 1.22 Here all members of the population are equally likely to be selected and the sample selection mechanism is based on chance. But not every sample of size 2 has the same chance of being selected. For example the sample "B and C" is impossible.
- 1.23 (a) Since a complete roster of full-time students exists, a simple random sample of 200 students could be taken. If student satisfaction with the quality of campus life randomly fluctuates across the student body, a systematic 1-in-20 sample could also be taken from the population frame. If student satisfaction with the quality of life may differ by gender and by experience/class level, a stratified sample using eight strata, female freshmen through female seniors and male freshmen through male seniors, could be selected. If student satisfaction with the quality of life is thought to fluctuate as much within clusters as between them, a cluster sample could be taken.
- (b) A simple random sample is one of the simplest to select. The population frame is the registrar's file of 4,000 student names.
- (c) A systematic sample is easier to select by hand from the registrar's records than a simple random sample, since an initial person at random is selected and then every 20th person thereafter would be sampled. The systematic sample would have the additional benefit that the alphabetic distribution of sampled students' names would be more comparable to the alphabetic distribution of student names in the campus population.
- (d) If rosters by gender and class designations are readily available, a stratified sample should be taken. Since student satisfaction with the quality of life may indeed differ by gender and class level, the use of a stratified sampling design will not only ensure all strata are represented in the sample, it will also generate a more representative sample and produce estimates of the population parameter that have greater precision.
- (e) If all 4,000 full-time students reside in one of 10 on-campus residence halls which fully integrate students by gender and by class, a cluster sample should be taken. A cluster could be defined as an entire residence hall, and the students of a single randomly selected residence hall could be sampled. Since each dormitory has 400 students, a systematic sample of 200 students can then be selected from the chosen cluster of 400 students. Alternately, a cluster could be defined as a floor of one of the 10 dormitories. Suppose there are four floors in each dormitory with 100 students on each floor. Two floors could be randomly sampled to produce the required 200 student sample. Selection of an entire dormitory may make distribution and collection of the survey easier to accomplish. In contrast, if there is some variable other than gender or class that differs across dormitories, sampling by floor may produce a more representative sample.

34 Chapter 1: Defining and Collecting Data

- 1.24 (a) Row 16: 2323 6737 5131 8888 1718 0654 6832 4647 6510 4877
Row 17: 4579 4269 2615 1308 2455 7830 5550 5852 5514 7182
Row 18: 0989 3205 0514 2256 8514 4642 7567 8896 2977 8822
Row 19: 5438 2745 9891 4991 4523 6847 9276 8646 1628 3554
Row 20: 9475 0899 2337 0892 0048 8033 6945 9826 9403 6858
Row 21: 7029 7341 3553 1403 3340 4205 0823 4144 1048 2949
Row 22: 8515 7479 5432 9792 6575 5760 0408 8112 2507 3742
Row 23: 1110 0023 4012 8607 4697 9664 4894 3928 7072 5815
Row 24: 3687 1507 7530 5925 7143 1738 1688 5625 8533 5041
Row 25: 2391 3483 5763 3081 6090 5169 0546
Note: All sequences above 5000 are discarded. There were no repeating sequences.
- (b) 089 189 289 389 489 589 689 789 889 989
1089 1189 1289 1389 1489 1589 1689 1789 1889 1989
2089 2189 2289 2389 2489 2589 2689 2789 2889 2989
3089 3189 3289 3389 3489 3589 3689 3789 3889 3989
4089 4189 4289 4389 4489 4589 4689 4789 4889 4989
- (c) With the single exception of invoice #0989, the invoices selected in the simple random sample are not the same as those selected in the systematic sample. It would be highly unlikely that a random process would select the same units as a systematic process.
- 1.25 (a) A stratified sample should be taken so that each of the three strata will be proportionately represented.
- (b) The number of observations in each of the three strata out of the total of 1,000 should reflect the proportion of the three categories in the customer database. For example, $3,500/10,000 = 35\%$ so 35% of 1,000 = 350 customers should be selected from the prospective buyers; similarly $4,500/10,000 = 45\%$ so 450 customers should be selected from the first time buyers, and $2,000/10,000 = 20\%$ so 200 customers from the repeat buyers.
- (c) It is not simple random sampling because, unlike the simple random sampling, it ensures proportionate representation across the entire population.
- 1.26 Before accepting the results of a survey of college students, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were they clear, accurate, unbiased, valid? What operational definition of “vast majority” was used? What was the response rate? What was the sample size?
- 1.27 (a) Possible coverage error: Only employees in a specific division of the company were sampled.
- (b) Possible nonresponse error: No attempt is made to contact nonrespondents to urge them to complete the evaluation of job satisfaction.
- (c) Possible sampling error: The sample statistics obtained from the sample will not be equal to the parameters of interest in the population.
- (d) Possible measurement error: Ambiguous wording in questions asked on the questionnaire.

- 1.28 The results are based on an online survey. If the frame is supposed to be small business owners, how is the population defined? This is a self-selecting sample of people who responded online, so there is an undefined nonresponse error. Sampling error cannot be determined since this is not a random sample.
- 1.29 Before accepting the results of the survey, you might want to know, for example:
Who funded the study? Why was it conducted? What was the population from which the sample was selected? What was the frame being used? What sampling design was used?
What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What other questions were asked? Were they clear, accurate, unbiased, and valid? What was the response rate? What was the margin of error? What was the sample size?
- 1.30 Before accepting the results of the survey, you might want to know, for example: Who funded the study? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What other questions were asked? Were the questions clear, accurate, unbiased, and valid? What was the response rate? What was the margin of error? What was the sample size? What frame was used?
- 1.31 A population contains all the items of interest whereas a sample contains only a portion of the items in the population.
- 1.32 A statistic is a summary measure describing a sample whereas a parameter is a summary measure describing an entire population.
- 1.33 Categorical random variables yield categorical responses such as yes or no answers. Numerical random variables yield numerical responses such as your height in inches.
- 1.34 Discrete random variables produce numerical responses that arise from a counting process. Continuous random variables produce numerical responses that arise from a measuring process.
- 1.35 Items or individuals in a probability sampling are selected based on known probabilities while items or individuals in a nonprobability samplings are selected without knowing their probabilities of selection.
- 1.36 Microsoft Excel could be used to perform various statistical computations that were possible only with a slide-rule or hand-held calculator in the old days.
- 1.37 (a) The population of interest was 18-54 year olds who currently own a smartphone and/or tablet, and who use and do not use these devices to shop.
(b) The sample was the 1,003 18-54 year olds who currently own a smartphone and/or tablet, who use and do not use these devices to shop, and who responded to the study.
(c) A parameter of interest is the proportion of all tablet users in the population who use their device to purchase product and services.
(d) A statistic used to estimate the parameter of interest in (c) is the proportion of tablet users in the sample who use their device to purchase product and services.

36 Chapter 1: Defining and Collecting Data

- 1.38 The answers to this question depend on which article and its corresponding data set is being selected.
- 1.39 (a) The population of interest was supply chain executives in a wide range of industries representing a mix of company sizes from across three global regions: Asia, Europe, and the Americas.
(b) The sample was the 503 supply chain executives in a wide range of industries representing a mix of company sizes from across three global regions: Asia, Europe, and the Americas surveyed by PwC from May to July 2012.
(c) A parameter of interest is the proportion of supply chain executives in the population who acknowledge that supply chain is seen as a strategic asset in their company.
(d) A statistic used to estimate the parameter of interest in (c) is the proportion of supply chain executives in the sample who acknowledge that supply chain is seen as a strategic asset in their company.
- 1.40 The answers to this question depend on which data set is being selected.
- 1.41 (a) Categorical variable: Which of the following best describes this firm's primary business?
(b) Numerical variable: On average, what percent of total monthly revenues are e-commerce revenues?
- 1.42 (a) The population of interest was the collection of all the 10,000 benefitted employees at the University of Utah when the study was conducted.
(b) The sample consisted of the 3,095 benefitted employees participated in the study.
(c) gender: categorical; age: numerical; education level: numerical; marital status: categorical; household income: numerical; employment category: categorical
- 1.43 (a) (i)categorical (iii) numerical, discrete
(ii)categorical (iv) categorical
(b) The answers will vary.
(c) The answers will vary.

CHAPTER 2

2.1	(a)	Category	Frequency	Percentage
		A	13	26%
		B	28	56
		C	9	18

(b) Category “B” is the majority.

2.2 (a) Table frequencies for all student responses

Student Major Categories

Gender	A	C	M	Totals
Male	14	9	2	25
Female	6	6	3	15
Totals	20	15	5	40

(b) Table percentages based on overall student responses

Student Major Categories

Gender	A	C	M	Totals
Male	35.0%	22.5%	5.0%	62.5%
Female	15.0%	15.0%	7.5%	37.5%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on row percentages

Student Major Categories

Gender	A	C	M	Totals
Male	56.0%	36.0%	8.0%	100.0%
Female	40.0%	40.0%	20.0%	100.0%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on column percentages

Student Major Categories

Gender	A	C	M	Totals
Male	70.0%	60.0%	40.0%	62.5%
Female	30.0%	40.0%	60.0%	37.5%
Totals	100.0%	100.0%	100.0%	100.0%

2.3 (a) You can conclude that Android smartphones have seen steady increase in market shares while BlackBerry and Other OS smartphones have seen steady decrease in market shares since 2011. Android smartphones dominated the market in all those three years.

(b) The iOS smartphones have overtaken Other OS smartphones and owned the second largest market share since 2012. The Microsoft smartphones have arisen to the third place in terms of market share in 2013 from the fifth place position in 2011 while the Other OS smartphones have dropped from the second place in 2011 to the last place in 2013.

38 Chapter 2: Organizing and Visualizing Variables

- 2.4 (a) The percentage of complaints for each automaker:

Automaker	Frequency	Percentage	Cumulative Pct.
General Motors	551	18.91%	18.91%
Other	516	17.71%	36.62%
Nissan Motors Corporation	467	16.03%	52.64%
Ford Motor Company	440	15.10%	67.74%
Chrysler LLC	439	15.07%	82.81%
Toyota Motor Sales	332	11.39%	94.20%
American Honda	169	5.80%	100.00%

- (b) General Motors has the most complaints, followed by Other, Nissan Motors Corporation, Ford Motor Company, Chrysler LLC, Toyota Motor Sales and American Honda.

- (c) The percentage of complaints for each category:

Category	Frequency	Percentage	Cumulative Pct.
Powertrain	1148	42.82%	42.82%
Steering	397	14.81%	57.63%
Interior Electronics/Hardware	279	10.41%	68.03%
Fuel/Emission/Exhaust System	240	8.95%	76.99%
Airbags and Seatbelts	201	7.50%	84.48%
Body and Glass	182	6.79%	91.27%
Brakes	163	6.08%	97.35%
Tires and Wheels	71	2.65%	100.00%

- (d) Powertrain has the most complaints, followed by steering, interior electronics/hardware, fuel/emission/exhaust system, airbags and seatbelts, body and glass, brakes, and, finally, tires and wheels.

- 2.5 (a) The percentage of values for each factor:

Most Important Factor	Frequency	Percentage	Cumulative Pct.
Product	464	35.80%	35.80%
Leadership	400	30.86%	66.67%
Marketing	346	26.70%	93.36%
Technology	86	6.64%	100.00%

- (b) Product is the most influencing factor in successful start-ups, followed by Leadership, Marketing and Technology.

- 2.6 (a)

Region	Oil Production (millions of barrels a day)	Percentage
Iran	2.69	3.27%
Saudi Arabia	9.58	11.66%
Other OPEC countries	17.93	21.82%
Non-OPEC countries	51.99	63.26%
Total	82.19	100.00%

- (b) More than half the oil produced is from non-OPEC countries. About 22% is produced by OPEC countries other than Iran and Saudi Arabia.

- 2.7 (a) The percentage of values for each response need:

Needs	Frequency	Percentage	Cumulative Pct.
Easier-to-use analytic tools	127	30.98%	30.98%
Improved ability to present and interpret data	123	30.00%	60.98%
Improved ability to predict impacts of my actions	49	11.95%	72.93%
Faster access to data	41	10.00%	82.93%
Improved relationships to the business line organizations	37	9.02%	91.95%
Improved ability to plan actions	33	8.05%	100.00%

- (b) “Easier-to-use analytic tools” is the most frequently mentioned need, followed by “Improved ability to present and interpret data”, “Improved ability to predict impacts of my actions”, “Faster access to data”, “Improved relationships to the business line organizations” and “Improved ability to plan actions”.

- 2.8 (a) Table of total percentages

ENJOY SHOPPING FOR CLOTHING FOR YOURSELF	GENDER		Total
	Male	Female	
Yes	22%	25%	47%
No	28%	25%	53%
Total	50%	50%	100%

Table of row percentages

ENJOY SHOPPING FOR CLOTHING FOR YOURSELF	GENDER		Total
	Male	Female	
Yes	46%	54%	100%
No	53%	47%	100%
Total	50%	50%	100%

Table of column percentages

ENJOY SHOPPING FOR CLOTHING FOR YOURSELF	GENDER		Total
	Male	Female	
Yes	44%	51%	47%
No	56%	49%	53%
Total	100%	100%	100%

- (b) A higher percentage of females enjoy shopping for clothing for themselves.

40 Chapter 2: Organizing and Visualizing Variables

2.9 (a)

Table of total percentages:

	Project Owner's Backing History		
Project Outcomes	Backing History	No Backing History	Total
Successful	26%	28%	54%
Not successful	16%	30%	46%
Total	42%	58%	100%

Table of row percentages:

	Project Owner's Backing History		
Project Outcomes	Backing History	No Backing History	Total
Successful	48%	52%	100%
Not successful	35%	65%	100%
Total	42%	58%	100%

Table of column percentages:

	Project Owner's Backing History		
Project Outcomes	Backing History	No Backing History	Total
Successful	62%	49%	54%
Not successful	38%	51%	46%
Total	100%	100%	100%

- (b) The column percentages is most informative for these data as they show that among those owners with backing history, 62% are successful while only 49% are successful among those with no backing history.
- (c) The percentage of success among owners with backing history is higher than those with no backing history.

2.10 Social recommendations had very little impact on correct recall. Those who arrived at the link from a recommendation had a correct recall of 73.07% as compared to those who arrived at the link from browsing who had a correct recall of 67.96%.

2.11 Ordered array: 63 64 68 71 75 88 94

2.12 Ordered array: 73 78 78 78 85 88 91

- 2.13 (a) $(17 + 7) / 70 = 34.29\%$ of small businesses pay less than 21% of the employee monthly health-care premium.
- (b) $(7 + 4) / 70 = 15.71\%$ of small businesses pay between 21% and 75% of the employee monthly health-care premium.
- (c) $(35) / 70 = 50.00\%$ of small businesses pay more than 75% of the employee monthly health-care premium.

2.14 (a) 0 but less than 5 million, 5 million but less than 10 million, 10 million but less than 15 million, 15 million but less than 20 million, 20 million but less than 25 million, 25 million but less than 30 million.

(b) 5 million

(c) 2.5 million, 7.5 million, 12.5 million, 17.5 million, 22.5 million, and 27.5 million.

- 2.15 (a) Ordered array: Cost(\$) 203.06, 208.48, 212.16, 227.36, 240.04, 249.22, 262.40, 263.10, 266.40, 268.28, 271.74, 273.98, 280.98, 295.40, 308.18, 309.30, 319.10, 321.18, 321.63, 324.08, 336.05, 338.00, 344.92, 382.00, 395.20, 434.96, 456.60, 472.20, 542.00, 659.92,

- (b) PHStat output:

Bin Cell	Frequency	Percentage	Cumulative Pctage.
200 but less than 270	10	0.3333333	33.33%
270 but less than 340	12	0.4	73.33%
340 but less than 410	3	0.1	83.33%
410 but less than 480	3	0.1	93.33%
480 but less than 550	1	0.0333333	96.67%
550 but less than 620	0	0	96.67%
620 but less than 690	1	0.0333333	100.00%

- (c) The costs of attending a basketball game is concentrating around \$305 for twelve of the teams have costs between \$270 and \$340.

- 2.16 (a)
- | Electricity Costs | Frequency | Percentage |
|-------------------|-----------|------------|
| \$80 to \$99 | 4 | 8% |
| \$100 to \$119 | 7 | 14 |
| \$120 to \$139 | 9 | 18 |
| \$140 to \$159 | 13 | 26 |
| \$160 to \$179 | 9 | 18 |
| \$180 to \$199 | 5 | 10 |
| \$200 to \$219 | 3 | 6 |

- (b)

<i>Electricity Costs</i>	<i>Frequency</i>	<i>Percentage</i>	<i>Cumulative %</i>
\$99	4	8%	8%
\$119	7	14%	22%
\$139	9	18%	40%
\$159	13	26%	66%
\$179	9	18%	84%
\$199	5	10%	94%
\$219	3	6%	100%

- (c) The majority of utility charges are clustered between \$120 and \$180.

42 Chapter 2: Organizing and Visualizing Variables

2.17 (a), (b) Annual Time Sitting in Traffic (hours)

Bin Cell	Frequency	Percentage	Cumulative Pctage.
15 but less than 20	1	3.23%	3.23%
20 but less than 25	4	12.90%	16.13%
25 but less than 30	4	12.90%	29.03%
30 but less than 35	2	6.45%	35.48%
35 but less than 40	7	22.58%	58.06%
40 but less than 45	3	9.68%	67.74%
45 but less than 50	4	12.90%	80.65%
50 but less than 55	2	6.45%	87.10%
55 but less than 60	1	3.23%	90.32%
60 but less than 65	1	3.23%	93.55%
65 but less than 70	0	0.00%	93.55%
70 but less than 75	2	6.45%	100.00%

Cost of Sitting in Traffic(\$)

Bin Cell	Frequency	Percentage	Cumulative Pctage.
300 but less than 450	4	12.90%	12.90%
450 but less than 600	6	19.35%	32.26%
600 but less than 750	6	19.35%	51.61%
750 but less than 900	5	16.13%	67.74%
900 but less than 1050	6	19.35%	87.10%
1050 but less than 1200	2	6.45%	93.55%
1200 but less than 1350	1	3.23%	96.77%
1350 but less than 1550	0	0.00%	96.77%
1550 but less than 1650	1	3.23%	100.00%

- (c) The annual time sitting in traffic is concentrated around 37.5 hours with a few spending as much as around 72.5 hours.
- (d) The cost of sitting in traffic per year is concentrated around \$675 with one costing as much as \$1,575.

2.18 (a), (b)

Bin Cell	Frequency	Percentage	Cumulative Pctage.
695 but less than 705	3	2.10%	2.10%
705 but less than 715	12	8.39%	10.49%
715 but less than 725	12	8.39%	18.88%
715 but less than 735	19	13.29%	32.17%
735 but less than 745	18	12.59%	44.76%
745 but less than 755	24	16.78%	61.54%
755 but less than 765	22	15.38%	76.92%
765 but less than 775	20	13.99%	90.91%
775 but less than 785	10	6.99%	97.90%
795 but less than 795	3	2.10%	100.00%

- (c) The average credit scores are concentrated around 750.

2.19 (a), (b)

<i>Bin</i>	<i>Frequency</i>	<i>Percentage</i>	<i>Cumulative %</i>
-0.00350 but less than -0.00201	13	13.00%	13.00%
-0.00200 but less than -0.00051	26	26.00%	39.00%
-0.00050 but less than 0.00099	32	32.00%	71.00%
0.00100 but less than 0.00249	20	20.00%	91.00%
0.00250 but less than 0.00399	8	8.00%	99.00%
0.004 but less than 0.00549	1	1.00%	100.00%

- (c) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.20 (a), (b)

<i>Bin</i>	<i>Frequency</i>	<i>Percentage</i>	<i>Cumulative %</i>
8.310 -- 8.329	3	6.12%	6.12%
8.330 -- 8.349	2	4.08%	10.20%
8.350 -- 8.369	1	2.04%	12.24%
8.370 -- 8.389	4	8.16%	20.41%
8.390 -- 8.409	4	8.16%	28.57%
8.410 -- 8.429	15	30.61%	59.18%
8.430 -- 8.449	7	14.29%	73.47%
8.450 -- 8.469	5	10.20%	83.67%
8.470 -- 8.489	5	10.20%	93.88%
8.490 -- 8.509	3	6.12%	100.00%

- (c) All the troughs will meet the company's requirements of between 8.31 and 8.61 inches wide.

2.21 (a), (b)

<i>Strength</i>	<i>Frequency</i>	<i>Percentage</i>	<i>Cumulative Percentage</i>
1500 -- 1549	1	3.33%	3.33%
1550 -- 1599	2	6.67%	10.00%
1600 -- 1649	2	6.67%	16.67%
1650 -- 1699	7	23.33%	40.00%
1700 -- 1749	5	16.67%	56.67%
1750 -- 1799	7	23.33%	80.00%
1800 -- 1849	3	10.00%	90.00%
1850 -- 1899	3	10.00%	100.00%

- (c) The strength of all the insulators meets the company's requirement of at least 1500 lbs.

44 Chapter 2: Organizing and Visualizing Variables

2.22 (a), (b) Manufacturer A:

Bin Cell	Frequency	Percentage	Cumulative Pctage.
6,500 but less than 7,500	3	7.50%	7.50%
7,500 but less than 8,500	5	12.50%	20.00%
8,500 but less than 9,500	20	50.00%	70.00%
9,500 but less than 10,500	9	22.50%	92.50%
10,500 but less than 11,500	3	7.50%	100.00%

Manufacturer B:

Bin Cell	Frequency	Percentage	Cumulative Pctage.
7,500 but less than 8,500	2	5.00%	5.00%
8,500 but less than 9,500	8	20.00%	25.00%
9,500 but less than 10,500	16	40.00%	65.00%
10,500 but less than 11,500	9	22.50%	87.50%
11,500 but less than 12,500	5	12.50%	100.00%

- (c) Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of its bulbs lasted less than 10,500 hours, contrasted with 70% of Manufacturer A's bulbs, which lasted less than 9,500 hours. None of Manufacturer A's bulbs lasted more than 11,499 hours, but 12.5% of Manufacturer B's bulbs lasted between 11,500 and 12,499 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 7,500 hours, whereas all of Manufacturer B's bulbs lasted at least 7,500 hours

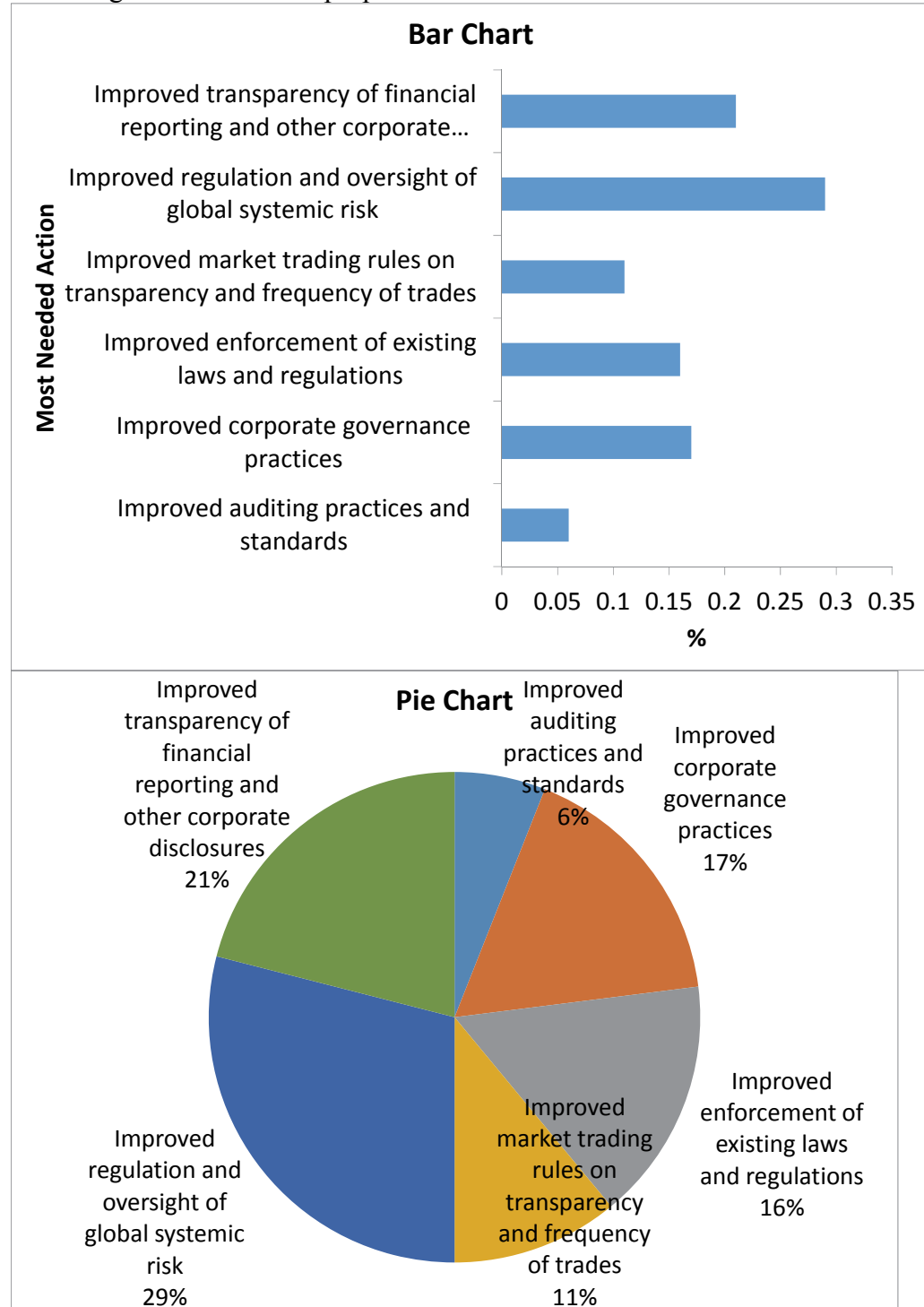
2.23 (a)

Amount of Soft Drink	Frequency	Percentage
1.850 – 1.899	1	2%
1.900 – 1.949	5	10
1.950 – 1.999	18	36
2.000 – 2.049	19	38
2.050 – 2.099	6	12
2.100 – 2.149	1	2
Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100

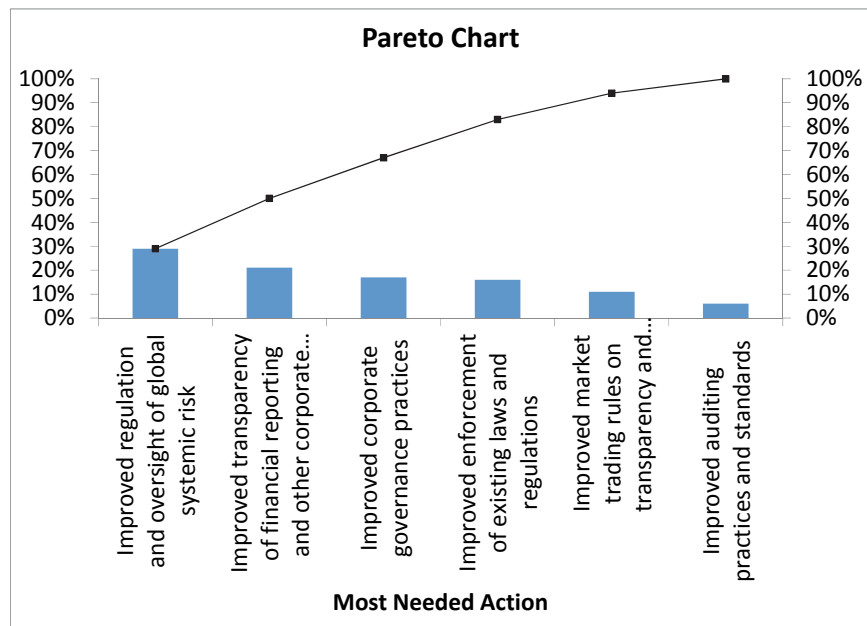
- (b) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.24 (a)

Percentages in decimals as proportions

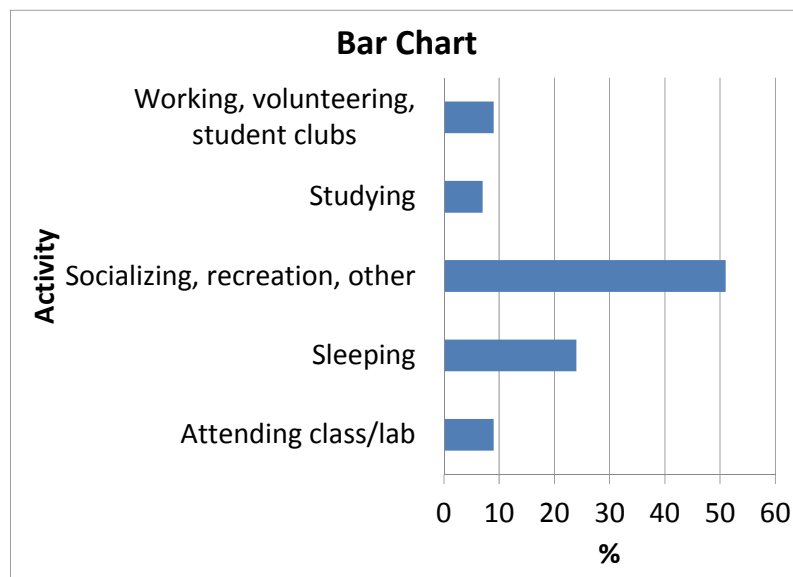


2.24 (a)
cont.

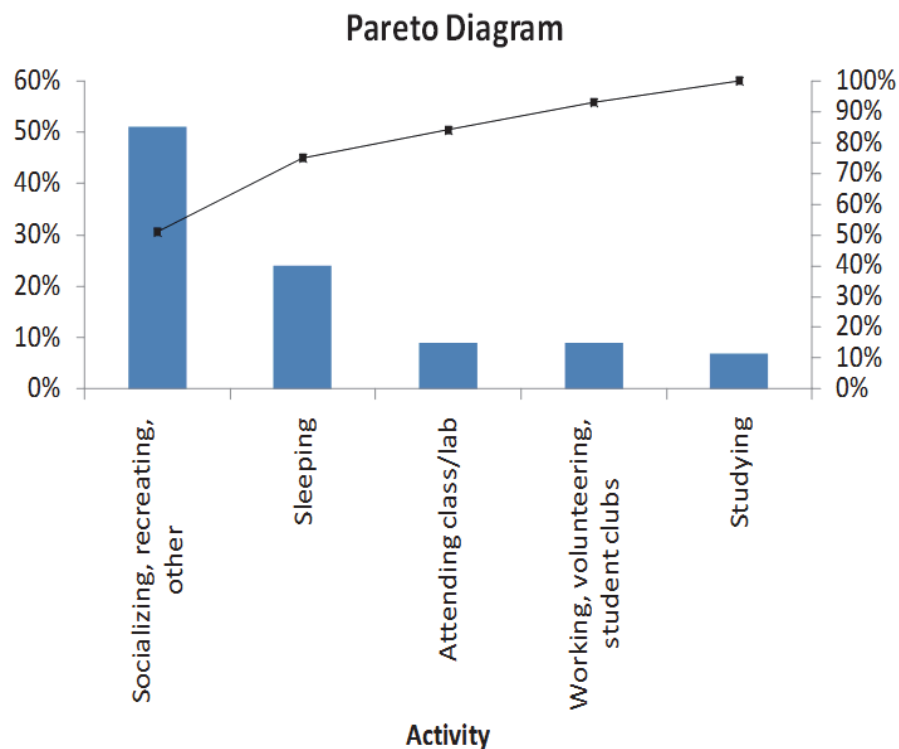
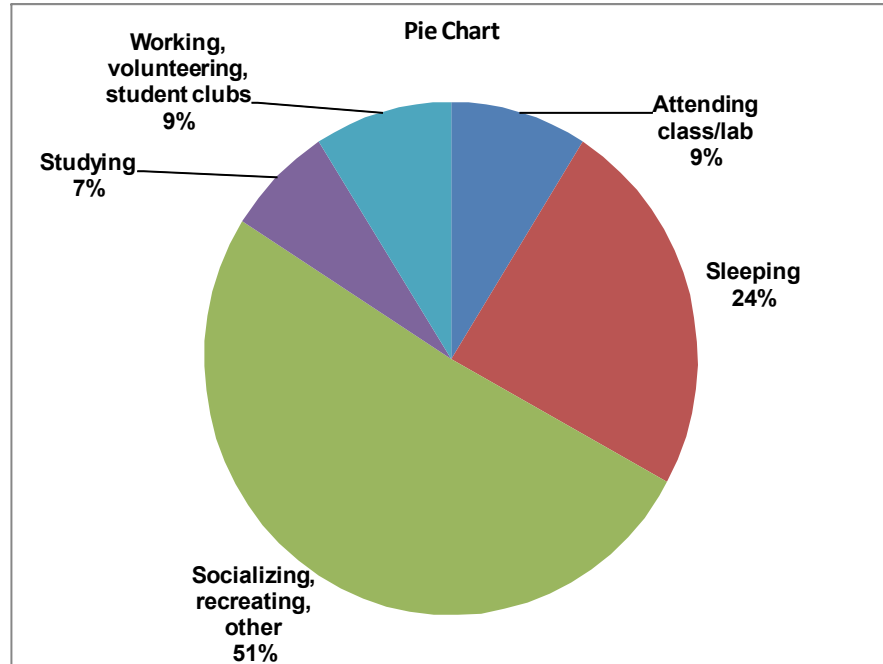


- (b) The Pareto diagram is better than the pie chart to portray these data because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.
- (c) You can conclude that “improved regulation and oversight of global systemic risk” accounts for the largest percentage (29%) of the most needed action to improve investor trust and market integrity.

2.25 (a)

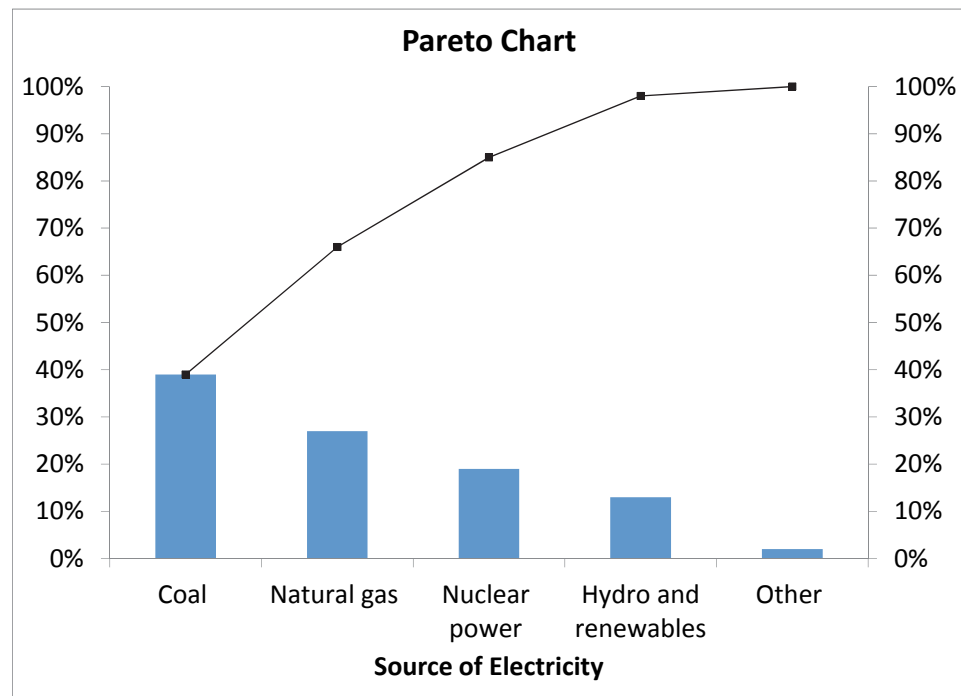


2.25 (a)
cont.



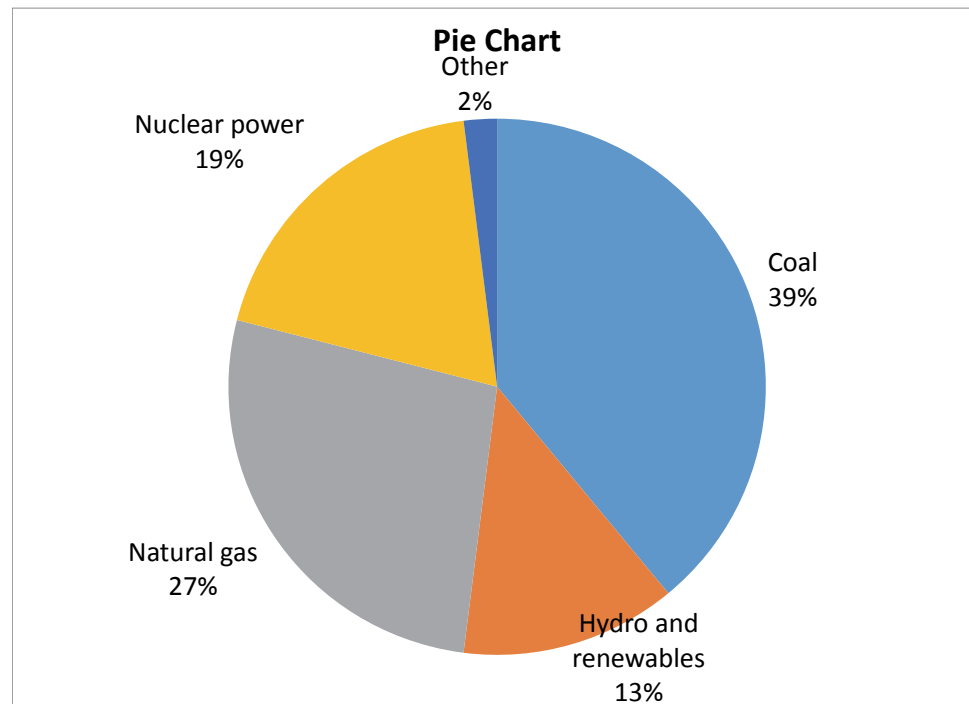
- (b) The Pareto diagram is better than the pie chart or the bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.
- (c) From the Pareto diagram, it is obvious that slightly more than 50% of them were socializing, recreating or performing other activities.

2.26 (a)



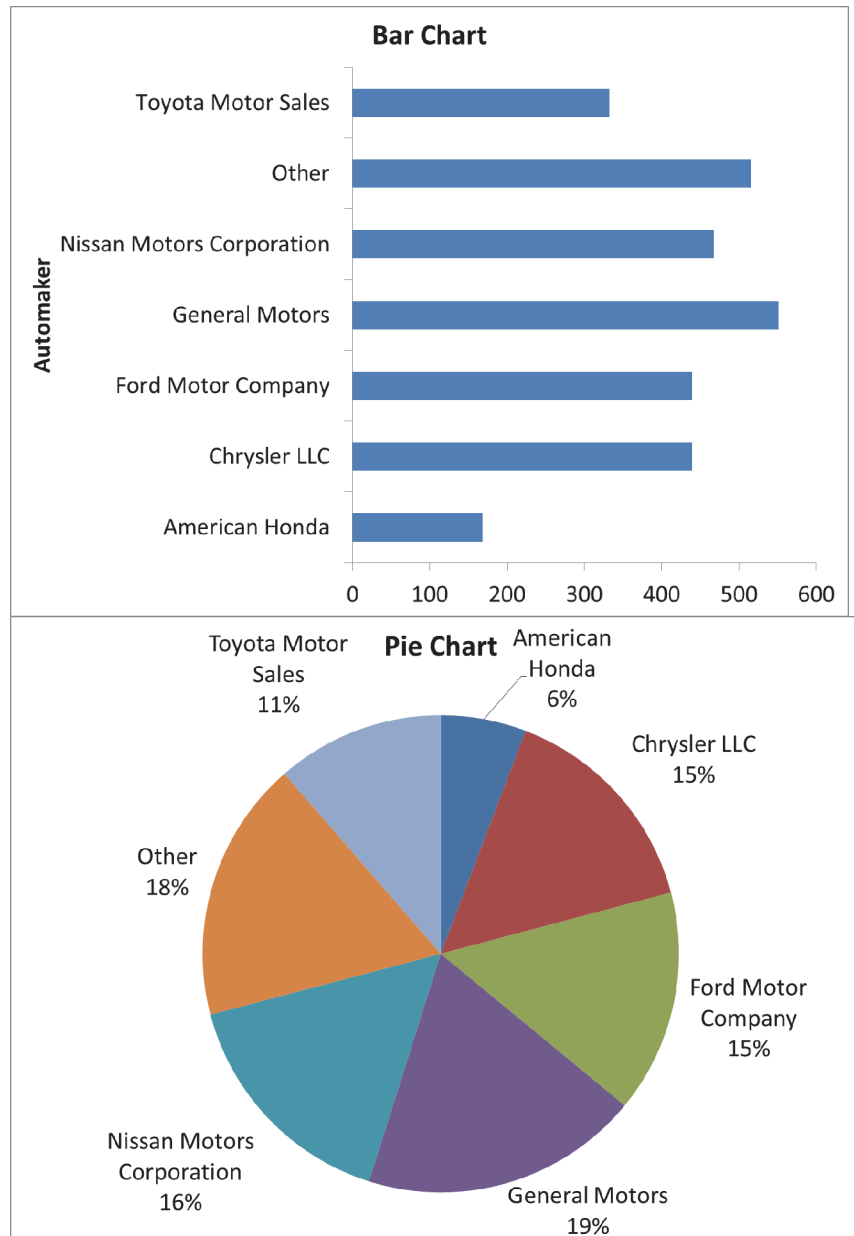
(b) Eighty-five percent of power is derived from coal, natural gas, or nuclear power.

(c)



(d) The Pareto diagram is better than the pie chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

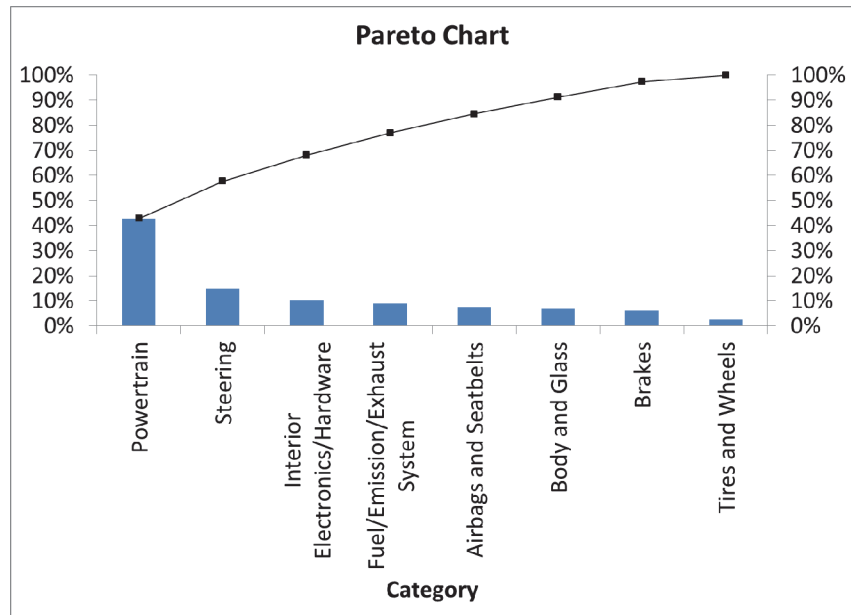
2.27 (a)



(b) The bar chart is more suitable if the purpose is to compare the categories. The pie chart is more suitable if the main objective is to investigate the portion of the whole that is in a particular category. *

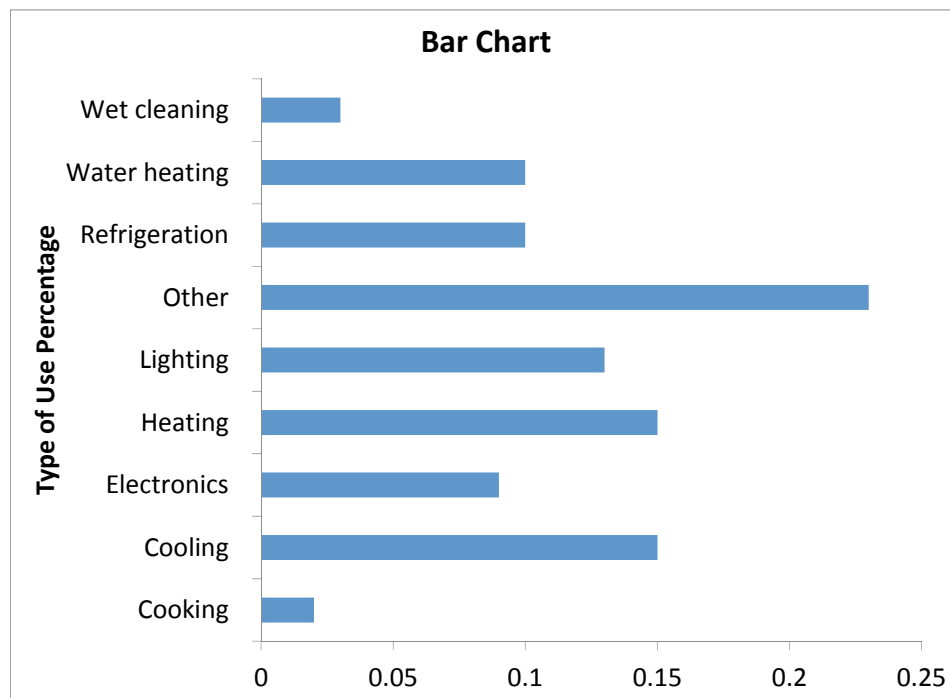
* Note: This is one of the many possible solutions for the question.

2.27 (c)
cont.

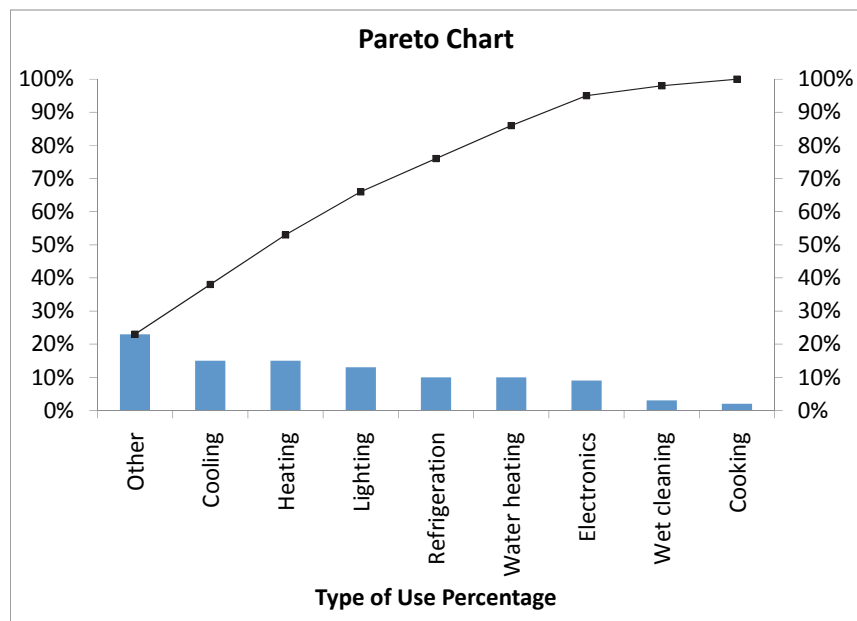
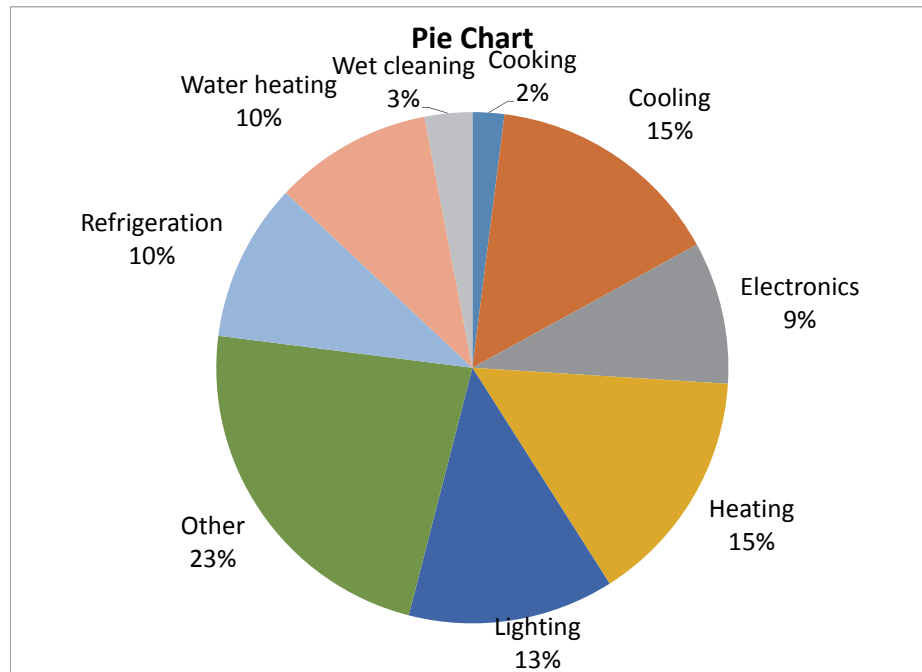


- (d) The “vital few” reasons for the categories of complaints are “powertrain”, “steering”, and “interior electronics/hardware” which account for more than 68% of the complaints. The remaining reasons are the “trivial many” which make up less than 32% of the complaints.

2.28 (a)

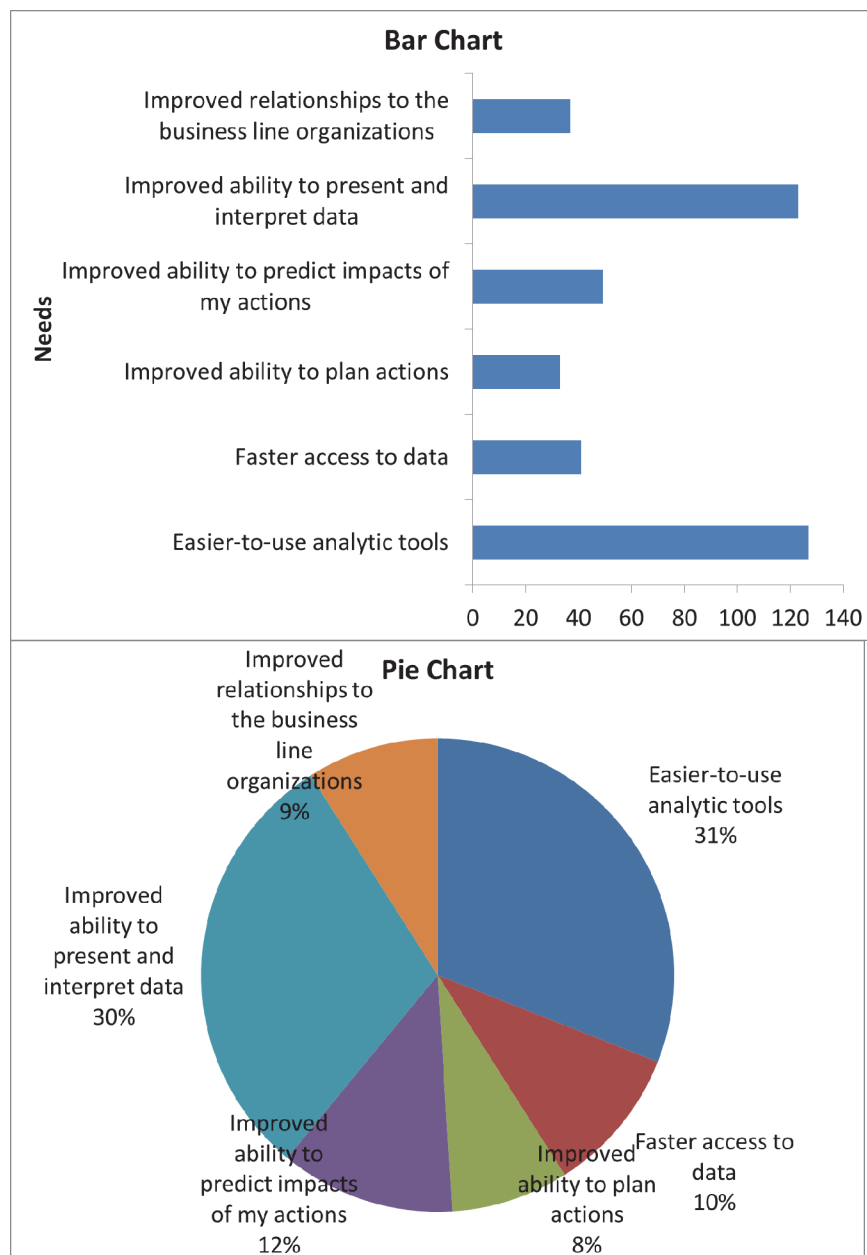


2.28 (a)
cont.



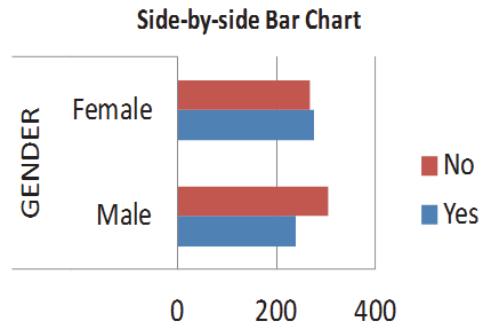
- (b) The Pareto diagram is better than the pie chart and bar chart because it not only sorts the frequencies in descending order; it also provides the cumulative polygon on the same scale.
- (c) Other, cooling, heating and lighting accounted for 66% of the residential electricity consumption in the United States.

2.29 (a)



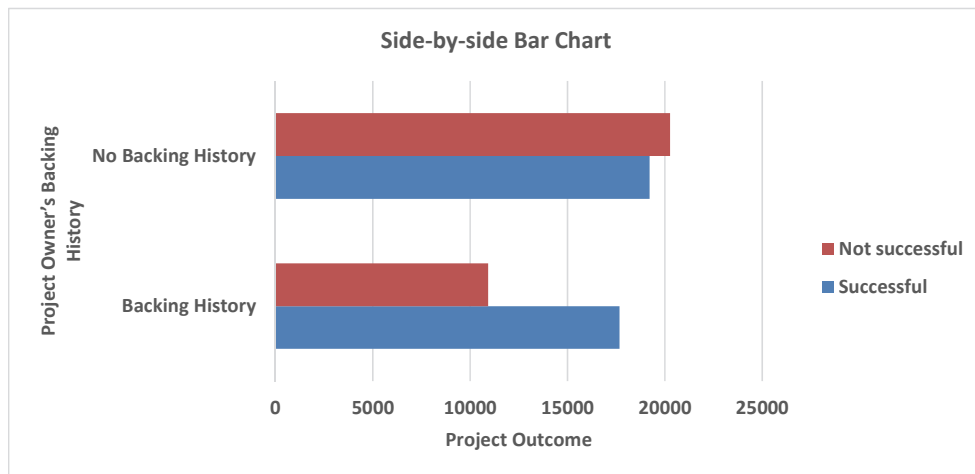
- (b) The highest percentage of needs for employer success with human resource metrics and reports comes from “easier-to-use analytic tools” at 30.98%, followed by “improved ability to present and interpret data” at 30%, “improved ability to predict impacts of my actions” at 11.95%, “faster access to data” at 10%, “improved relationships to the business line organizations” at 9.02% and “improved ability to plan actions” at 8.05%.

2.30 (a)



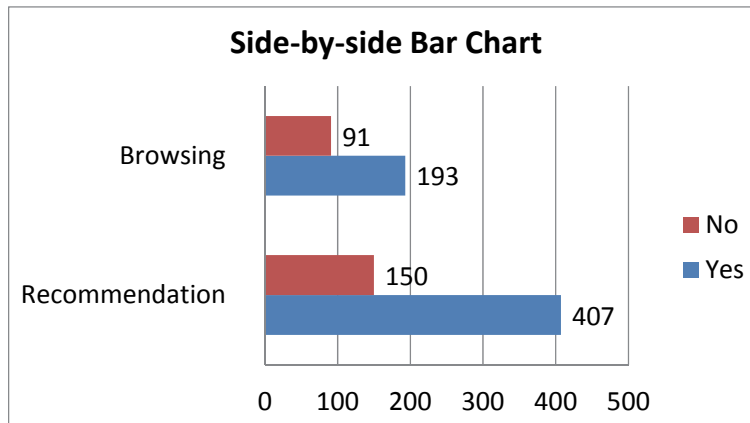
(b) A higher percentage of females enjoy shopping for clothing.

2.31 (a)



(b) The percentage of success among owners with backing history is higher than those with no backing history.

2.32 (a)



(b) Social recommendations had very little impact on correct recall.

54 Chapter 2: Organizing and Visualizing Variables

2.33 Stem-and-leaf of Finance Scores

```

5      34
6      9
7      4
9      38

```

2.34 Ordered array: 50 74 74 76 81 89 92

2.35 (a) Ordered array: 9.1 9.4 9.7 10.0 10.2 10.2 10.3 10.8 11.1 11.2
11.5 11.5 11.6 11.6 11.7 11.7 12.2 12.2 12.3
12.4 12.8 12.9 13.0 13.2

- (b) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.
- (c) The most likely gasoline purchase is between 11 and 11.7 gallons.
- (d) Yes, the third row is the most frequently occurring stem in the display and it is located in the center of the distribution.

2.36 (a)

		Stem-and-Leaf Display	
		Stem unit: 100	
Statistics		2	0 1 1 3 4 5 6 6 7 7 7 8
Sample Size	30	3	0 1 1 2 2 2 2 4 4 4 8
Mean	326.2640	4	0 3 6 7
Median	308.7400	5	4
Std. Deviation	102.7341	6	6
Minimum	203.0600		
Maximum	659.9200		

- (b) The costs are concentrated around \$200 and \$300.

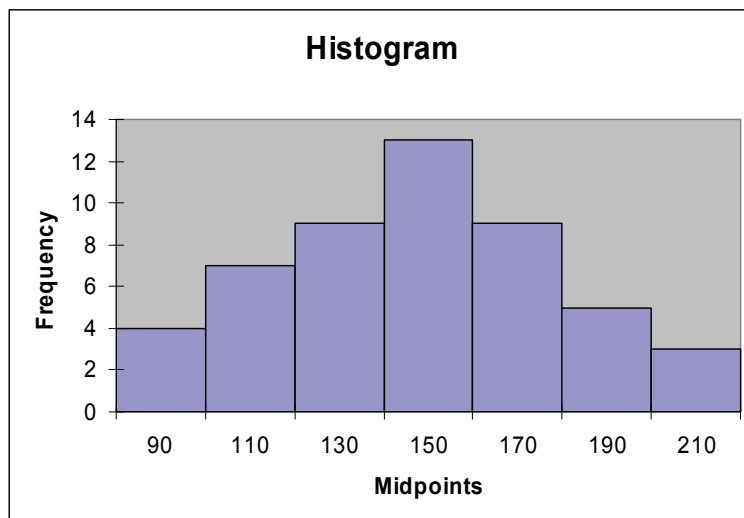
2.37 (a) Ordered array: 1.5, 3.2, 4.6, 7.1, 8.9, 9.0, 9.4, 9.9, 10.0, 10.1, 10.8, 11.5, 11.7, 11.8, 13.8, 14.0, 14.0, 16.1, 17.7, 26.3, 31.2, 32.5, 74.5, 91.6, 113.3, 127.4

2.37 (b)
cont.

Stem-and-Leaf Display			
Stem unit 10			
Statistics		0	2357999
Sample Size	26	1	000122244468
Mean	26.61154	2	6
Median	11.75	3	13
Std. Deviation	34.42669	4	
Minimum	1.5	5	
Maximum	127.4	6	
		7	5
		8	
		9	2
		10	
		11	3
		12	7

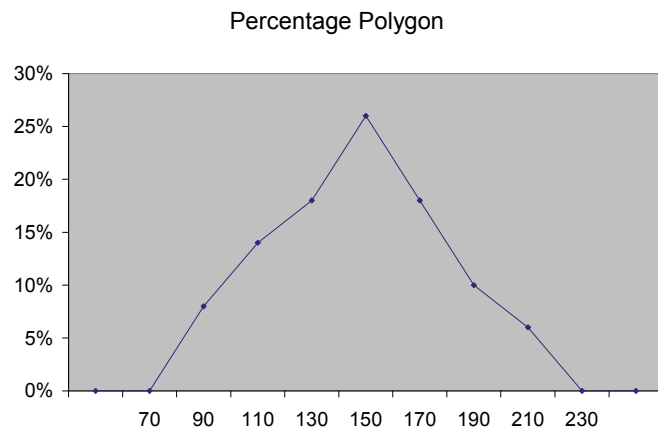
- (c) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.
- (d) The amount of caffeine in energy drinks is concentrated around 1.0 mg/oz.

2.38 (a)

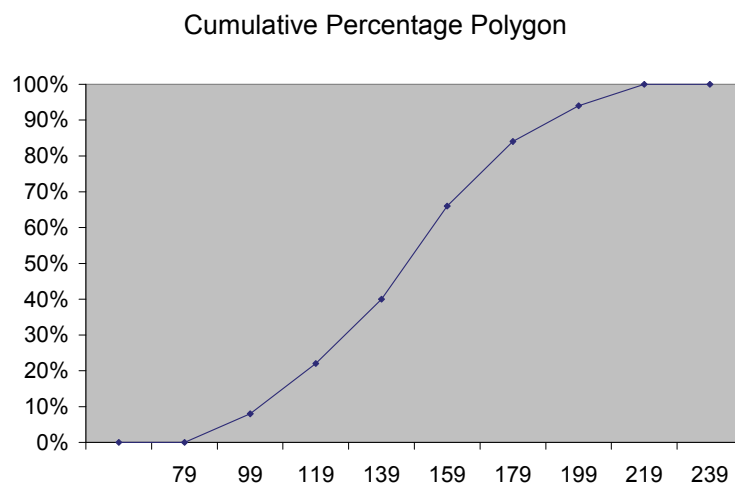


56 Chapter 2: Organizing and Visualizing Variables

2.38 (a)
cont.



(b)

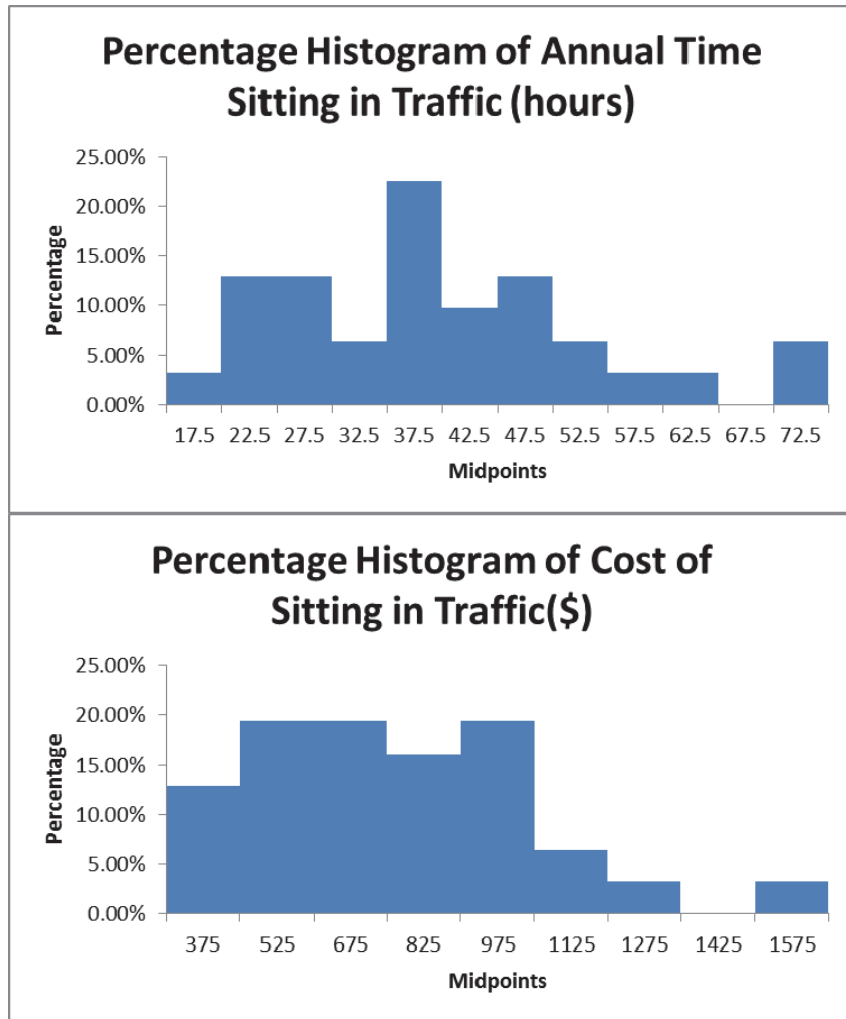


(c) The majority of utility charges are clustered between \$120 and \$180.

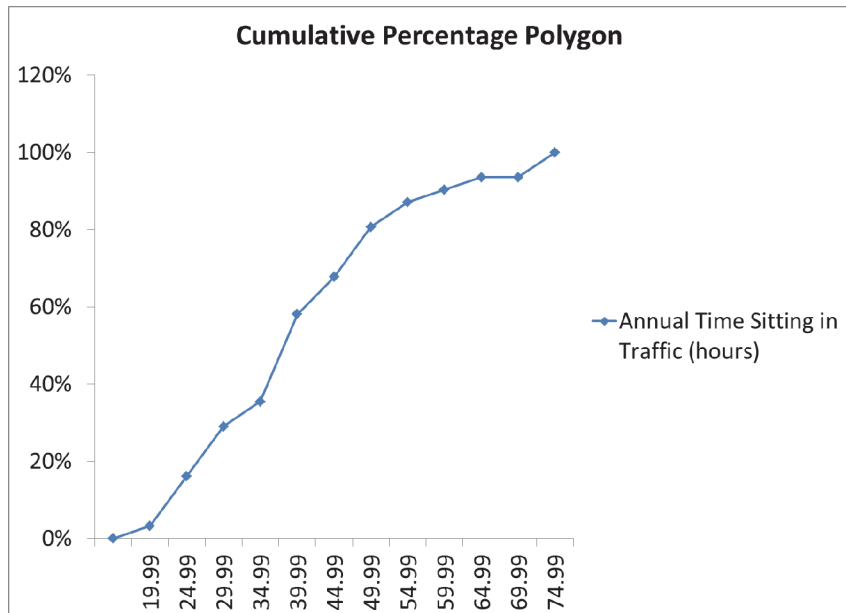
2.39 The costs of attending a baseball game is concentrating around \$200. There are a few outliers in the right tail with two teams having a cost higher than \$320.

2.40 Property taxes seem concentrated between \$1,000 and \$1,500 and also between \$500 and \$1,000 per capita. There were more states with property taxes per capita below \$1,500 than above \$1,500.

2.41 (a)

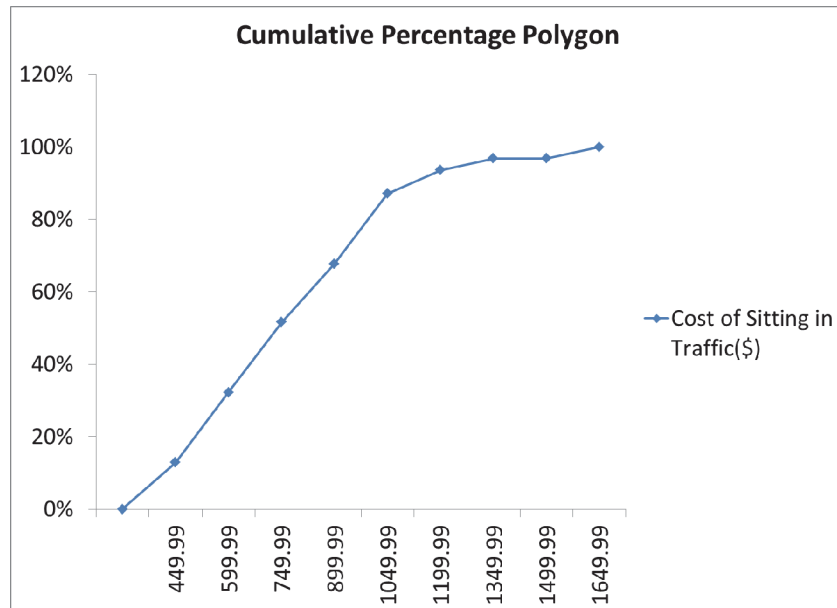


(b)



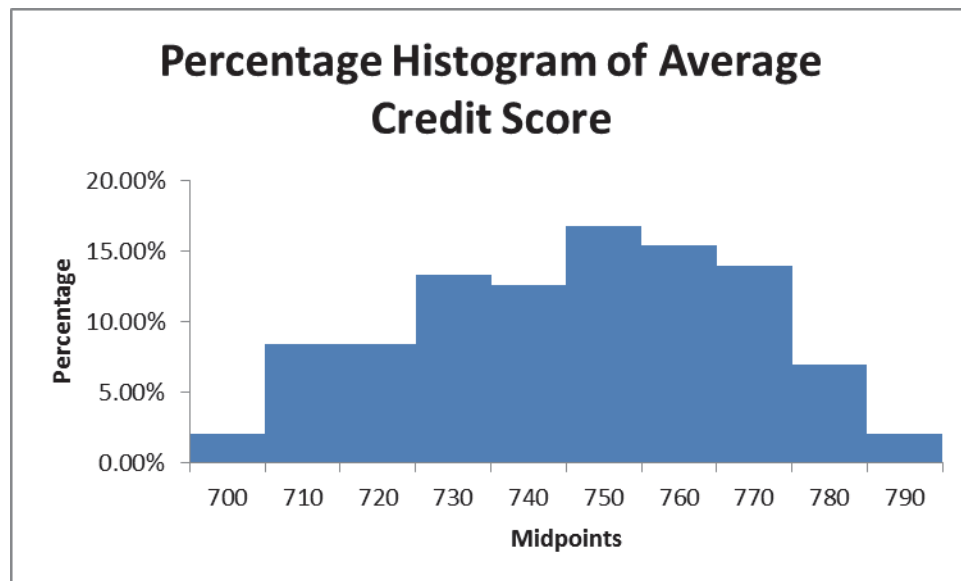
58 Chapter 2: Organizing and Visualizing Variables

2.41 (b)
cont.

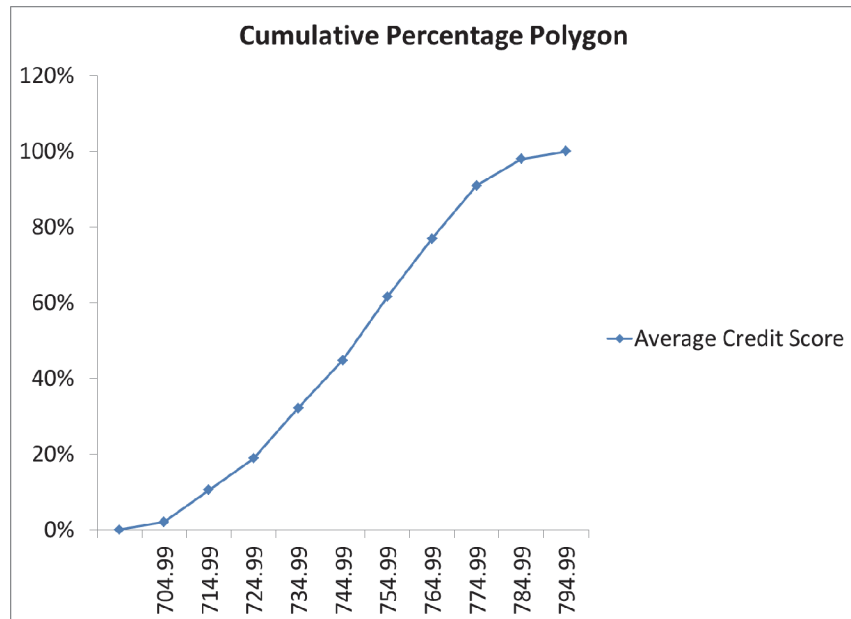


- (c) The annual time sitting in traffic is concentrated around 37.5 hours with a few spending as much as around 72.5 hours.
- (d) The cost of sitting in traffic per year is concentrated around \$675 with one costing as much as \$1,575.

2.42 (a)

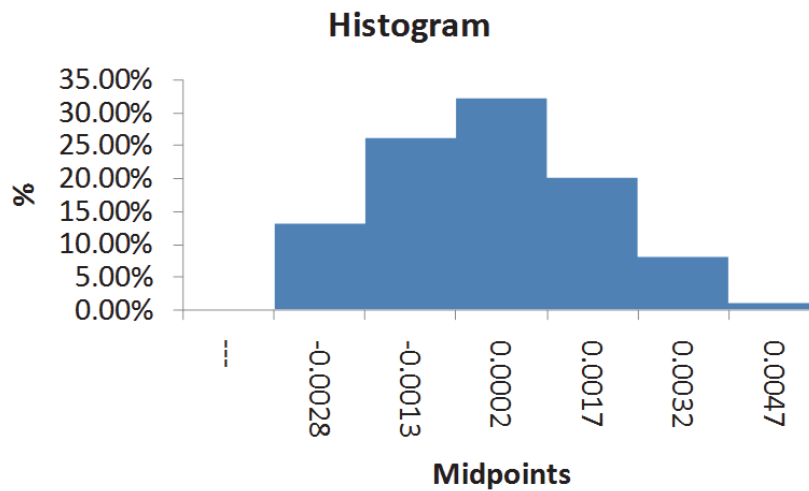


2.42 (b)
cont.



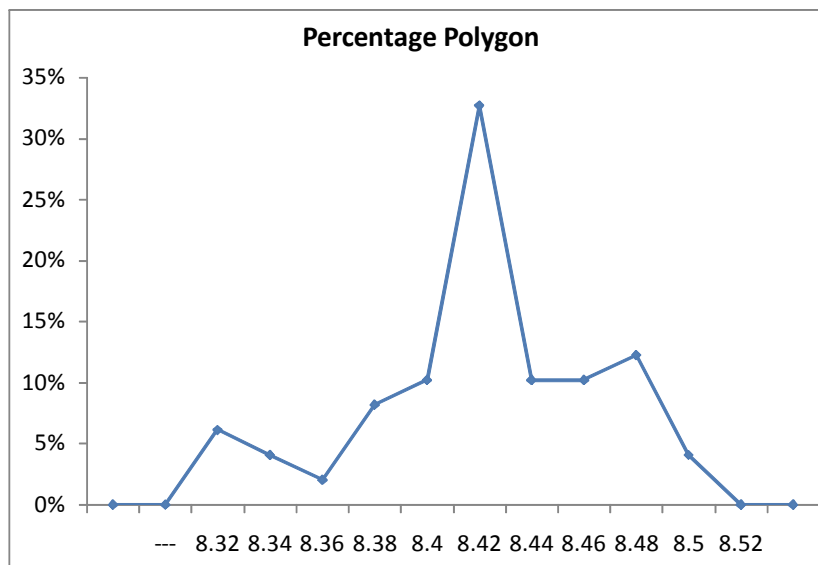
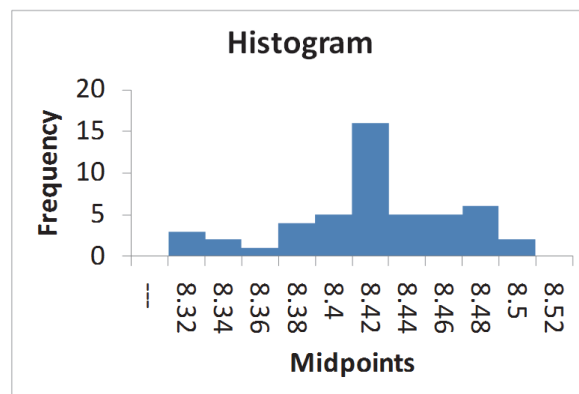
(c) The average credit scores are concentrated around 750.

2.43 (a)

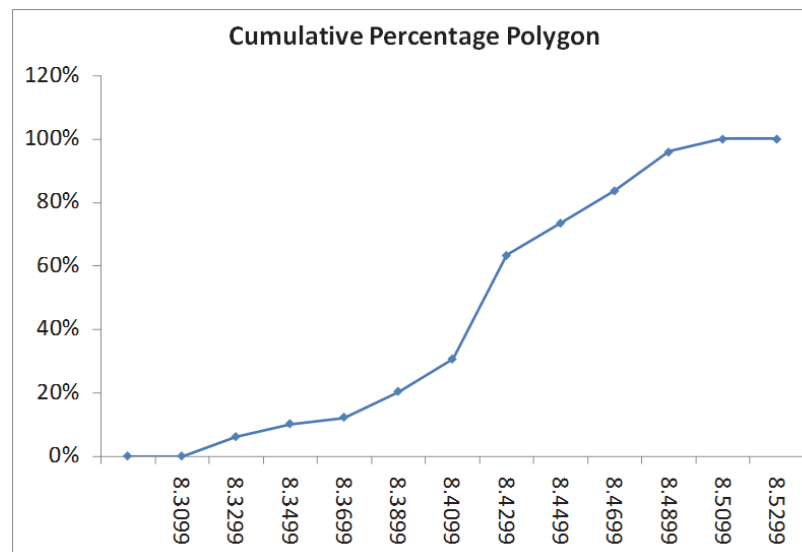


(b) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.44 (a)

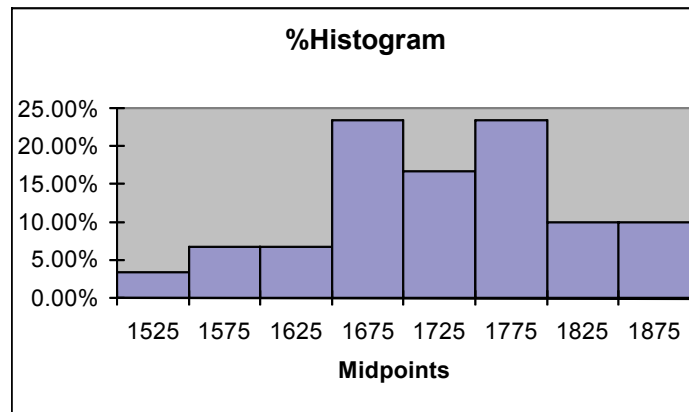


(b)

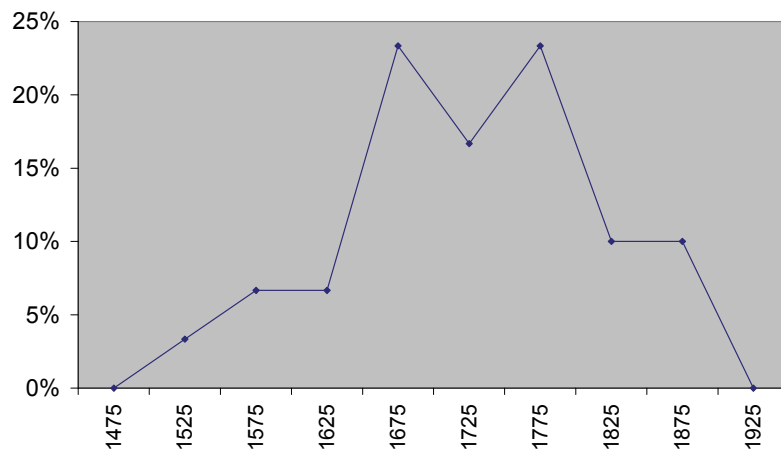


(c) All the troughs will meet the company's requirements of between 8.31 and 8.61 inches wide.

2.45 (a)

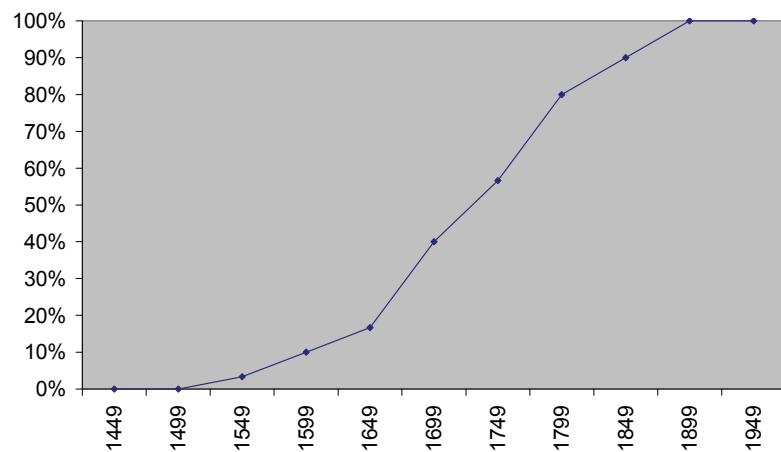


Percentage Polygon



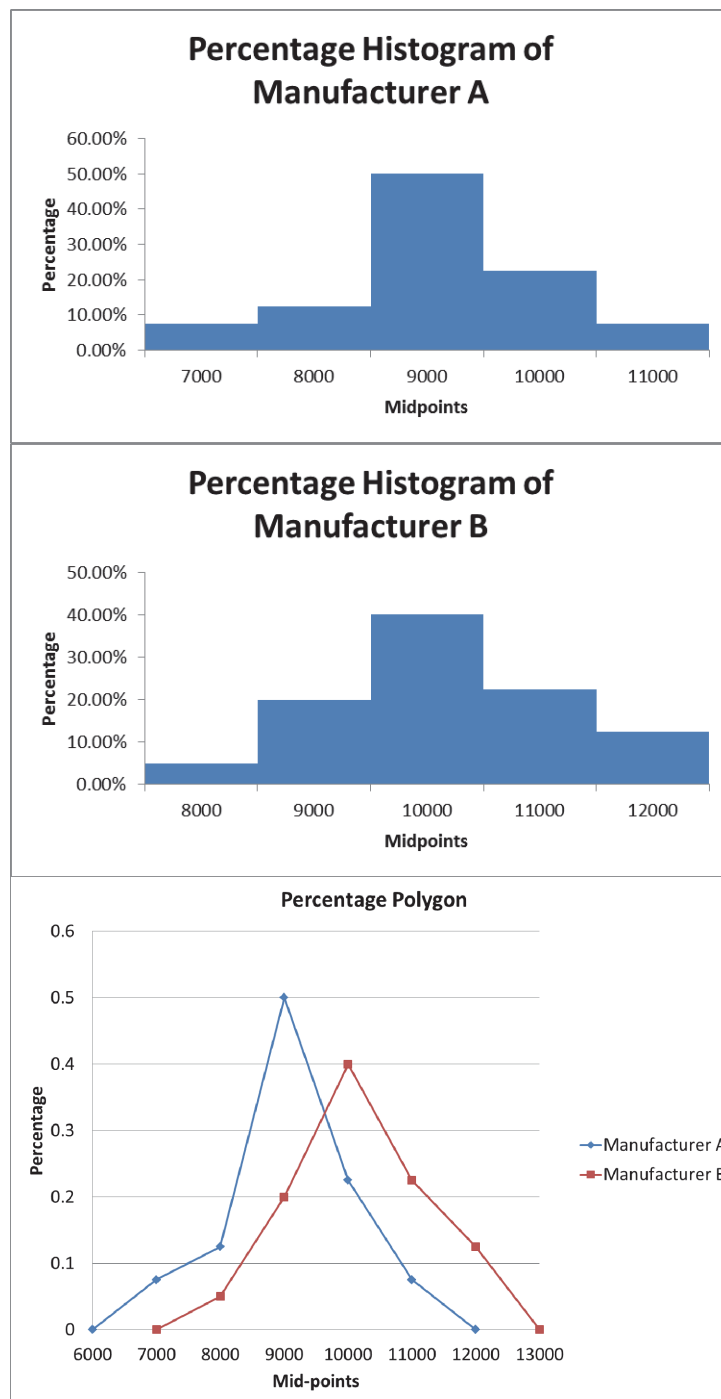
(b)

Cumulative Percentage Polygon

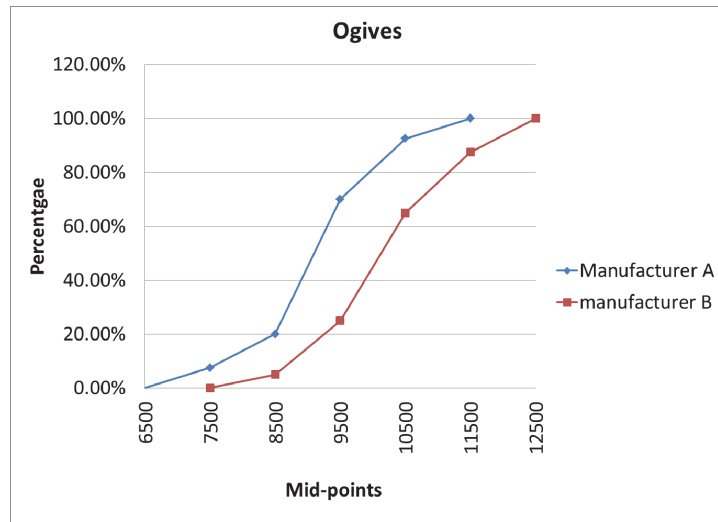


(c) The strength of all the insulators meets the company's requirement of at least 1500 lbs.

2.46 (a)

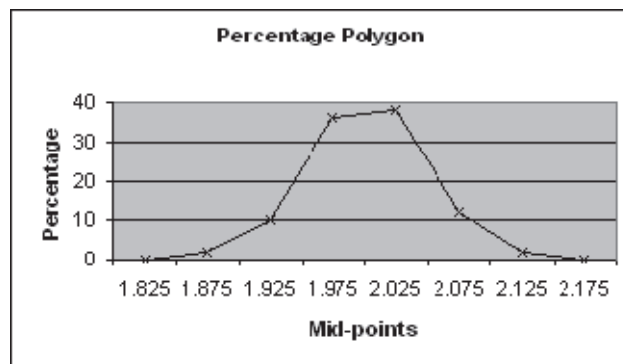
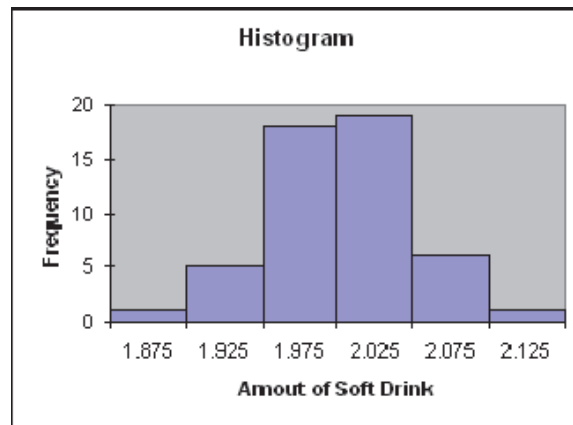


2.46 (b)
cont.



- (c) Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of their bulbs lasted 10499 hours or less contrasted with 70% of Manufacturer A's bulbs which lasted 9499 hours or less. None of Manufacturer A's bulbs lasted more than 11499 hours, but 12.5% of Manufacturer B's bulbs lasted between 11500 and 12499 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 7500 hours, while all of Manufacturer B's bulbs lasted at least 7500 hours.

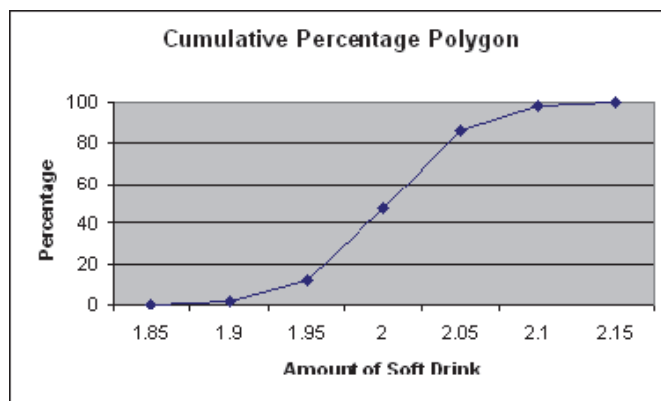
2.47 (a)



64 Chapter 2: Organizing and Visualizing Variables

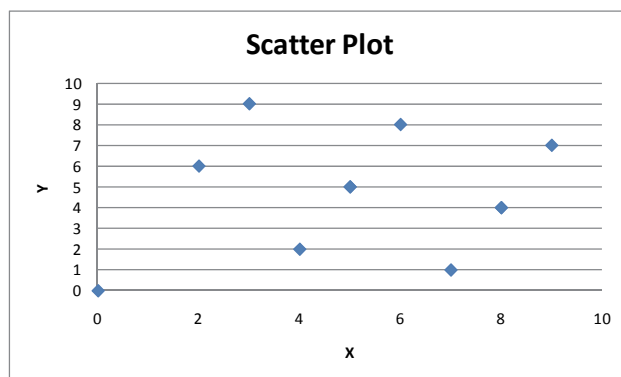
2.47 (b)

Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100



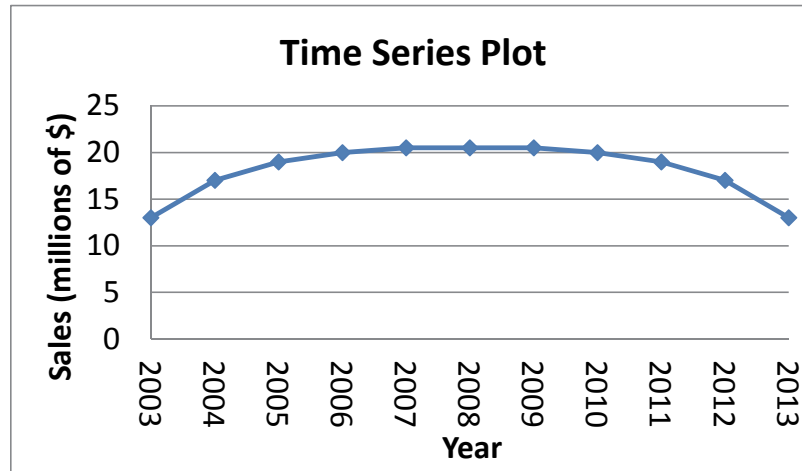
- (c) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.48 (a)



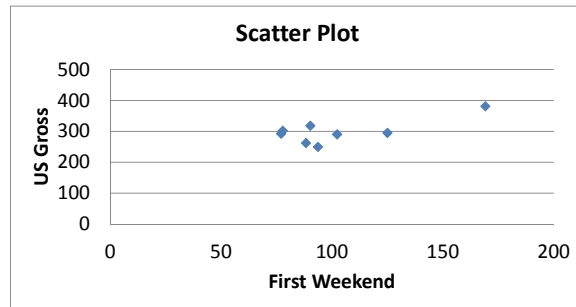
- (b) There is no relationship between X and Y .

2.49 (a)

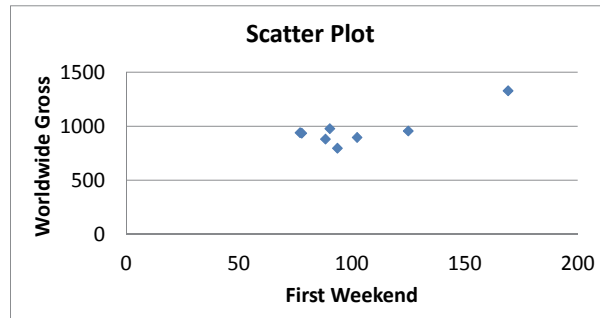


(b) Annual sales appear to be increasing in the earlier years before 2006 but start to decline after 2008.

2.50 (a)

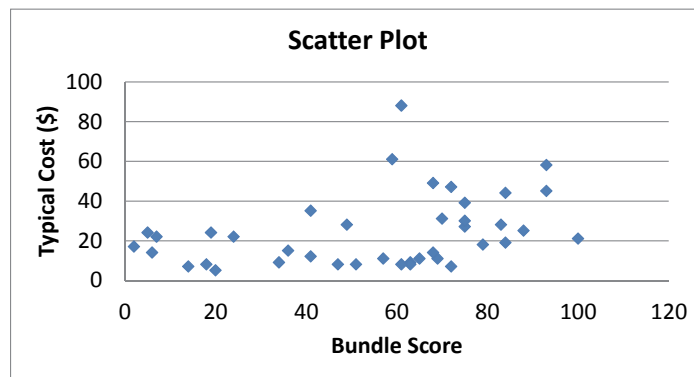


(b)



(c) There appears to be a linear relationship between the first weekend gross and either the U.S. gross or the worldwide gross of Harry Potter movies. However, this relationship is greatly affected by the results of the last movie, *Deathly Hallows, Part II*.

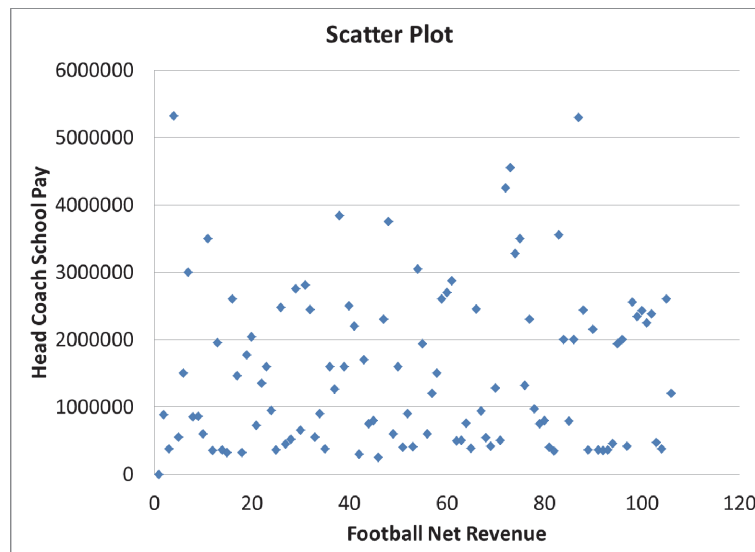
2.51 (a)



(b) There appears to be a positive relationship between Bundle score and typical cost.

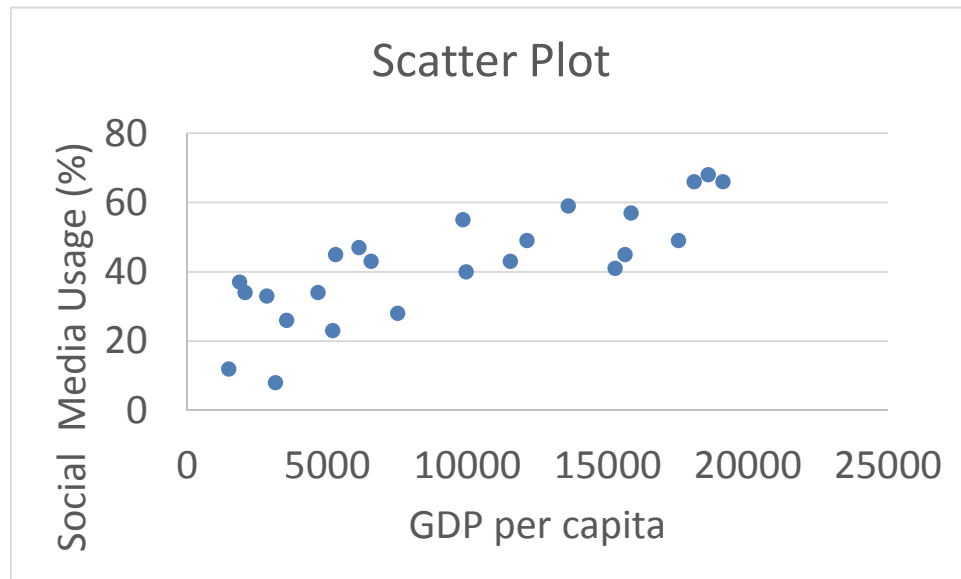
2.52 (a) Yes, schools with higher revenues will also have higher coaches' total pay.

(b)



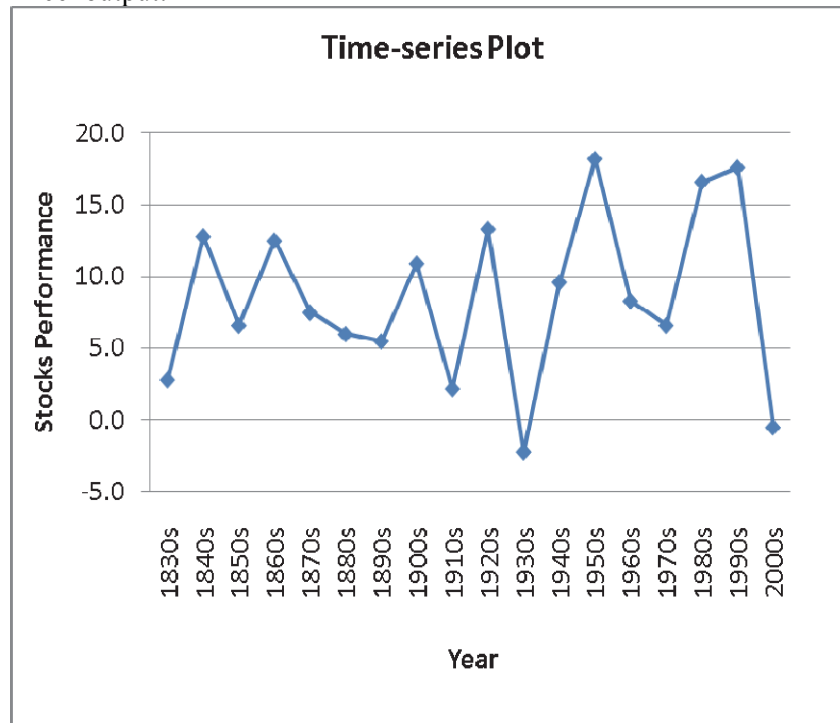
(c) The scatter plot contradicts your answer to (a).

2.53 (a)



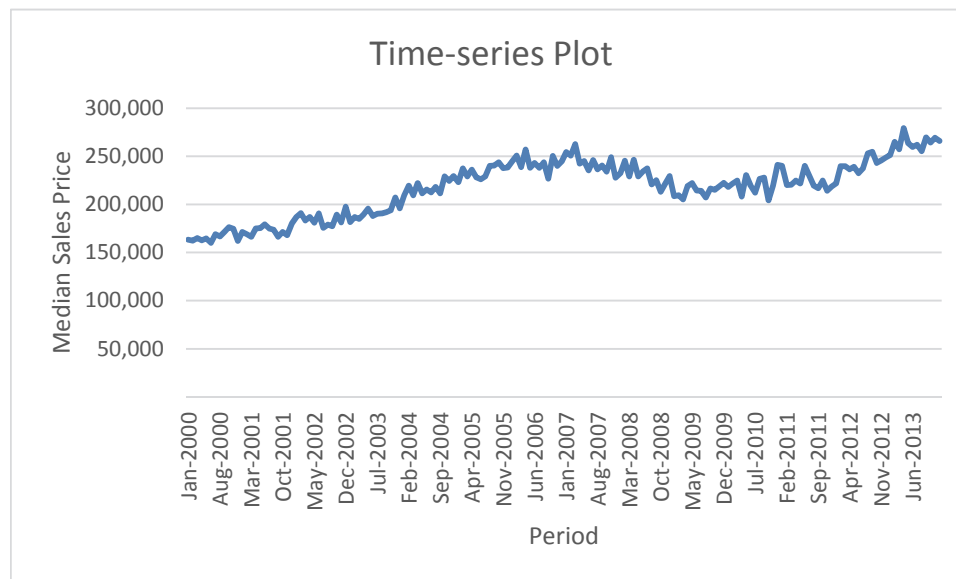
(b) There is a positive relationship between GDP and social media usage.

2.54 (a) Excel output:



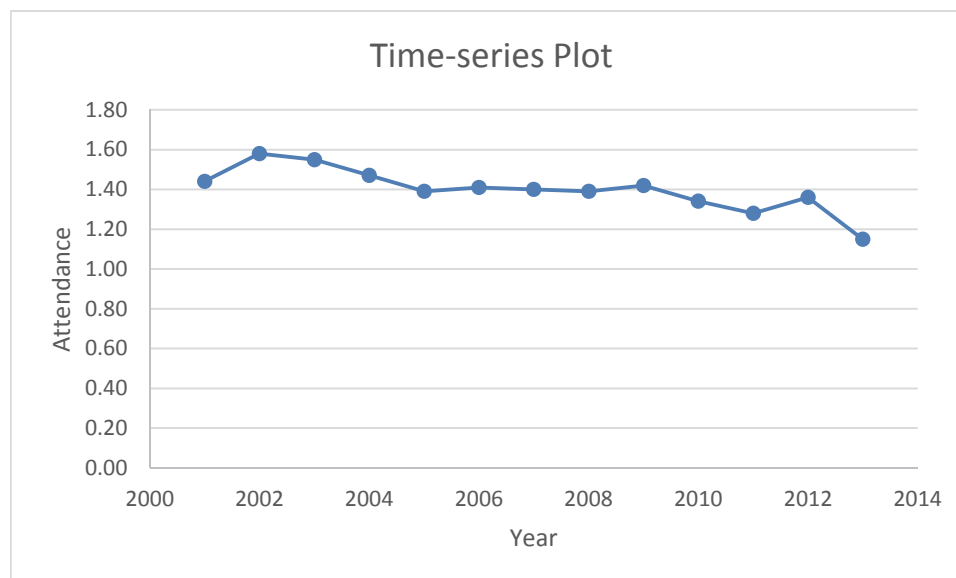
(b) There is a great deal of variation in the returns from decade to decade. Most of the returns are between 5% and 15%. The 1950s, 1980s, and 1990s had exceptionally high returns, and only the 1930s and 2000s had negative returns.

2.55 (a)



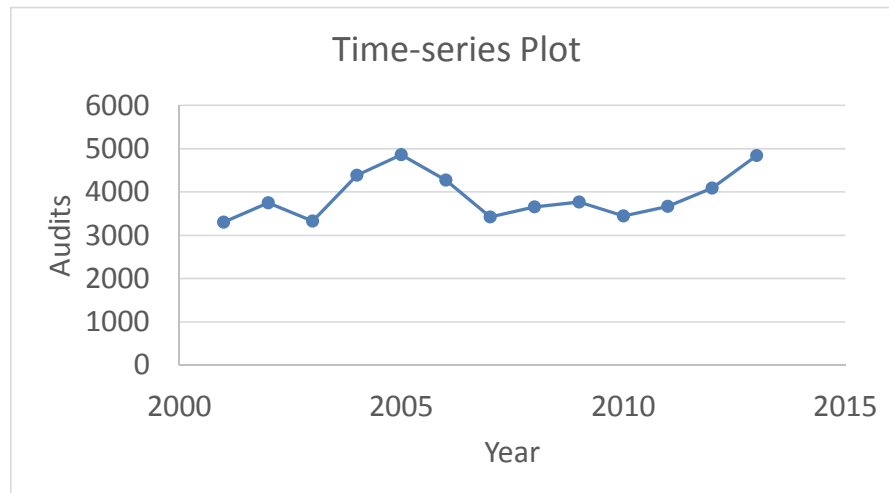
- (b) There is an upward trend on the median home sales price till 2007 and the sales price started a downward trend from then on till 2009 when it started to trend up again.

2.56 (a)



- (b) There was a slight decline in movie attendance between 2001 and 2013. During that time, movie attendance increased from 2001 to 2002 but then after 2004 began decreasing to levels below that in 2001.

2.57 (a)



- (b) The number of audits increased from 2001 to 2005, then declined back to the 2001 level in 2007, hover around the same level from then on until 2010 and increased again after 2010.

2.58 (a) Pivotal table of tallies in terms of counts:

Count of 3YrReturn%		Star Rating					
Type		Five	Four	One	Three	Two	Grand Total
<div><div></div>Growth</div>		18	76	16	74	43	227
Large		9	31	5	37	21	103
Mid-Cap		7	28	4	20	13	72
Small		2	17	7	17	9	52
<div><div></div>Value</div>		5	22	7	36	19	89
Large		2	13	5	21	9	50
Mid-Cap		1	4		9	5	19
Small		2	5	2	6	5	20
Grand Total		23	98	23	110	62	316

Pivotal table of tallies in terms of % of grand total:

Count of 3YrReturn%		Star Rating					
Type		Five	Four	One	Three	Two	Grand Total
<input checked="" type="checkbox"/> Growth		5.70%	24.05%	5.06%	23.42%	13.61%	71.84%
Large		2.85%	9.81%	1.58%	11.71%	6.65%	32.59%
Mid-Cap		2.22%	8.86%	1.27%	6.33%	4.11%	22.78%
Small		0.63%	5.38%	2.22%	5.38%	2.85%	16.46%
<input checked="" type="checkbox"/> Value		1.58%	6.96%	2.22%	11.39%	6.01%	28.16%
Large		0.63%	4.11%	1.58%	6.65%	2.85%	15.82%
Mid-Cap		0.32%	1.27%	0.00%	2.85%	1.58%	6.01%
Small		0.63%	1.58%	0.63%	1.90%	1.58%	6.33%
Grand Total		7.28%	31.01%	7.28%	34.81%	19.62%	100.00%

- 2.58 (b) Patterns of star rating conditioned on market cap:
 cont. For the growth funds as a group, most are rated as four-star, followed by three-star, two-star, five-star and one-star. The pattern of star rating is similar across the different market cap within the growth funds with most of the mid-cap funds receiving a four-star rating, followed by three-star, two-star, five-star and one-star, most of the small-cap funds receiving a four-star or three-star rating, followed by two-star, one-star and five-star while most of large cap funds receiving a three-star rating, followed by two-star, five-star and one-star.

For the value funds as a group, most are rated as three-star, followed by four-star, two-star, one-star and five-star. Within the value funds, the large-cap funds follow the same pattern as the value funds as a group. Most of the mid-cap funds are rated as three-star, followed by two-star, four-star, five-star and one-star while most of the small-cap funds are rated as three-star, followed by either two-star or four-star, and either one-star or five star.

Patterns of market cap conditioned on star rating:

Most of the growth funds are large-cap, followed by mid-cap and small-cap. The pattern is similar among the five-star, four-star, three-star and two-star growth funds but among the one-star growth funds, most are small-cap, followed by large-cap and mid-cap.

The largest share of the value funds is large-cap, followed by small-cap and mid-cap. The pattern is similar among the four-star and one-star value funds. Among the three-star value funds, most are large-cap, followed by mid-cap and then small-cap while most are large-cap, followed by equal portions of mid-cap and small-cap among the two-star value funds and most are either large-cap or small-cap followed by mid-cap among the five-star value funds.

- 2.59 (a) Pivotal table of tallies in terms of counts:

Count of 3YrReturn% Star Rating						
Market Cap	Five	Four	One	Three	Two	Grand Total
<input checked="" type="checkbox"/> Large	11	44	10	58	30	153
Average	1	2	3	4	4	14
High			1		3	4
Low	10	42	6	54	23	135
<input checked="" type="checkbox"/> Mid-Cap	8	32	4	29	18	91
Average	2	7	3	13	14	39
Low	6	25	1	16	4	52
<input checked="" type="checkbox"/> Small	4	22	9	23	14	72
Average	1	6	3	18	10	38
High		1	6	1	1	9
Low	3	15		4	3	25
Grand Total	23	98	23	110	62	316

- 2.59 (a) Pivotal table of tallies in terms of % of grand total:
cont.

Count of 3YrReturn% Star Rating							
Market Cap		Five	Four	One	Three	Two	Grand Total
<input checked="" type="checkbox"/> Large		3.48%	13.92%	3.16%	18.35%	9.49%	48.42%
Average		0.32%	0.63%	0.95%	1.27%	1.27%	4.43%
High		0.00%	0.00%	0.32%	0.00%	0.95%	1.27%
Low		3.16%	13.29%	1.90%	17.09%	7.28%	42.72%
<input checked="" type="checkbox"/> Mid-Cap		2.53%	10.13%	1.27%	9.18%	5.70%	28.80%
Average		0.63%	2.22%	0.95%	4.11%	4.43%	12.34%
Low		1.90%	7.91%	0.32%	5.06%	1.27%	16.46%
<input checked="" type="checkbox"/> Small		1.27%	6.96%	2.85%	7.28%	4.43%	22.78%
Average		0.32%	1.90%	0.95%	5.70%	3.16%	12.03%
High		0.00%	0.32%	1.90%	0.32%	0.32%	2.85%
Low		0.95%	4.75%	0.00%	1.27%	0.95%	7.91%
Grand Total		7.28%	31.01%	7.28%	34.81%	19.62%	100.00%

- (b) Patterns of star rating conditioned on risk:
For the large-cap funds as a group, most are rated as three-star, followed by four-star, two-star, five-star and then one-star. The pattern of star rating is the same among the low-risk large-cap funds. The pattern is different among the high-risk and average-risk large-cap funds. Among the high-risk large-cap funds, most are rated as two-star, followed by one three-star with no three-star, four-star or five-star rating. Among the average-risk large-cap funds, most are two-star and three-star, followed by one-star, four-star and five-star rating.
For the mid-cap funds as a group, most are rated as four-star, followed by three-star, two-star, five-star and then one-star. The pattern of star rating is different among the average-risk mid-cap funds with the largest portion of two-star, followed by three-star, four-star, one-star and five-star. Among the low-risk mid-cap funds, most are rated as four-star, followed by three-star, five-star, two-star and one-star.
For the small-cap funds as a group, most are rated as three-star, followed by four-star, two-star, one-star and then five-star. Among the average-risk small-cap funds, most are three-star, followed by two-star, four-star, one-star and five-star. Among the high-risk small-cap funds, most are rated as one-star, followed by equal portions of two-star, three-star and four-star and no five-star. Among the low-risk small-cap funds, most are four-star, followed by three-star and equal portions of two-star and five-star with none rated as one-star.
Patterns of risk conditioned on star rating:
Among the large-cap funds, most are low-risk, followed by average-risk and finally high-risk. The pattern is the same among the one-star, two-star, three-star, four-star and five-star large-cap funds. Among the mid-cap funds, most are low-risk, followed by average-risk with no high-risk. The pattern is the same among the five-star, four-star and three-star mid-cap funds.
Among the small-cap funds, most are average-risk, followed by low-risk and finally high-risk. The pattern is the same for the two-star and three-star small-cap funds. Among the one-star small-cap funds, most are high-risk, followed by average-risk with no low-risk. Among the four-star and five-star small-cap funds, most are low-risk, followed by average-risk and high-risk.

2.60 (a) Pivotal table of tallies in terms of counts:

Count of 3YrReturn%		Star Rating						
Type		Five	Four	One	Three	Two	Grand Total	
☐ Growth		18	76	16	74	43	227	
	Average	3	15	6	28	22	74	
	High		1	5	1	3	10	
	Low	15	60	5	45	18	143	
☐ Value		5	22	7	36	19	89	
	Average	1		3	7	6	17	
	High			2		1	3	
	Low	4	22	2	29	12	69	
Grand Total		23	98	23	110	62	316	

Pivotal table of tallies in terms of % of grand total:

Count of 3YrReturn% Star Rating							
Type		Five	Four	One	Three	Two	Grand Total
☐ Growth		5.70%	24.05%	5.06%	23.42%	13.61%	71.84%
	Average	0.95%	4.75%	1.90%	8.86%	6.96%	23.42%
	High	0.00%	0.32%	1.58%	0.32%	0.95%	3.16%
	Low	4.75%	18.99%	1.58%	14.24%	5.70%	45.25%
☐ Value		1.58%	6.96%	2.22%	11.39%	6.01%	28.16%
	Average	0.32%	0.00%	0.95%	2.22%	1.90%	5.38%
	High	0.00%	0.00%	0.63%	0.00%	0.32%	0.95%
	Low	1.27%	6.96%	0.63%	9.18%	3.80%	21.84%
Grand Total		7.28%	31.01%	7.28%	34.81%	19.62%	100.00%

(b) Patterns of star rating conditioned on risk:

For the growth funds as a group, most are rated as four-star, followed by three-star, two-star, five-star and one-star. The pattern of star rating is the same among the low-risk growth funds. The pattern is different among the high-risk and average-risk growth funds. Among the high-risk growth funds, most are rated as one-star, followed by two-star, equal portions of three-star and four-star with no five-star. Among the average-risk growth funds, most are rated as three-star, followed by two-star, four-star, one-star and five-star.

For the value funds as a group, most are rated as three-star, followed by four-star, two-star, one-star and five-star. Among the average-risk value funds, most are three-star, followed by two-star, one-star, and five-star with no four-star. Among the high-risk value funds, most are one-star, followed by two-star with no three-star, four-star or five-star. Among the low-risk value funds, most are three-star, followed by four-star, two-star, five-star and one-star.

Patterns of risk conditioned on star rating:

Most of the growth funds are rated as low-risk, followed by average-risk and then high-risk. The pattern is the same among the three-star, four-star and five-star growth funds. Among the one-star growth funds, most are average-risk, followed by equal portions of high-risk and low-risk. Among the two-star growth funds, most are average-risk, followed by low-risk and high-risk.

Most of the value funds are rated as low-risk, followed by average-risk and then high-risk. The pattern is the same among the two-star, three-star and five-star value funds. Among the one-star value funds, most are average-risk, followed by equal portions of high-risk and low-risk. Among the four-star value funds, all are low-risk with no average-risk or high-risk.

- 2.61 (a) Presented below are just one of the $4 \cdot 3 \cdot 2 \cdot 1 = 24$ possible pivotal tables of tallies in terms of counts:

Count of 3YrReturn%		Star ra																											
Type	Five		Five Total			Four			Four Total			One			One Total			Three			Three Total			Two			Two Total		Grand Total
	Average	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low			
Growth	3	15	18	15	1	60	76	6	5	5	16	28	1	45	74	22	3	18	43	227									
Large	1	8	9	2	29	31	1	4	5	2	35	37	4	3	14	21				103									
Mid-Cap	2	5	7	7	21	28	3	1	4	12	8	20	11	2	13				72										
Small	2	2	6	1	10	17	2	5	7	14	1	2	17	7	2	9				52									
Value	1	4	5		22	22	3	2	2	7	7	29	36	6	1	12	19			89									
Large	2	2			13	13	2	1	2	5	2	19	21		9		9			50									
Mid-Cap	1	1			4	4				1	8	9	3	2	5					19									
Small	1	1	2		5	5	1	1	2	4	2	6	3	1	1	5				20									
Grand Total	4	19	23	15	1	82	98	9	7	7	23	35	1	74	110	28	4	30	62	316									

Presented below are just one of the $4 \cdot 3 \cdot 2 \cdot 1 = 24$ possible pivotal tables of tallies in terms of % of grand total:

Count of 3YrReturn%		Star ra																											
Type	Five		Five Total			Four			Four Total			One			One Total			Three			Three Total			Two			Two Total		Grand Total
	Average	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low	Average	High	Low			
Growth	0.95%	4.75%	5.70%	4.75%	0.32%	18.99%	24.05%	1.90%	1.58%	1.58%	5.06%	8.86%	0.32%	14.24%	23.42%	6.96%	0.95%	5.70%	13.61%	71.84%									
Large	0.32%	2.53%	2.85%	0.63%	0.00%	9.18%	9.81%	0.32%	0.00%	1.27%	1.58%	0.63%	0.00%	11.08%	11.71%	1.27%	0.95%	4.43%	6.65%	32.59%									
Mid-Cap	0.63%	1.58%	2.22%	2.22%	0.00%	6.65%	8.86%	0.95%	0.00%	0.32%	1.27%	3.80%	0.00%	2.53%	6.33%	3.48%	0.00%	0.63%	4.11%	22.78%									
Small	0.00%	0.63%	0.63%	1.90%	0.32%	3.16%	5.38%	0.63%	1.58%	0.00%	2.22%	4.43%	0.32%	0.63%	5.38%	2.22%	0.00%	0.63%	2.85%	16.46%									
Value	0.32%	1.27%	1.58%	0.00%	0.00%	6.96%	6.96%	0.95%	0.63%	0.63%	2.22%	2.22%	0.00%	9.18%	11.39%	1.90%	0.32%	3.80%	6.01%	28.16%									
Large	0.00%	0.63%	0.63%	0.00%	0.00%	4.11%	4.11%	0.63%	0.32%	0.63%	1.58%	0.63%	0.00%	6.01%	6.65%	0.00%	0.00%	2.85%	2.85%	15.82%									
Mid-Cap	0.00%	0.32%	0.32%	0.00%	0.00%	1.27%	1.27%	0.00%	0.00%	0.00%	0.00%	0.32%	0.00%	2.53%	2.85%	0.95%	0.00%	0.63%	1.58%	6.01%									
Small	0.32%	0.32%	0.63%	0.00%	0.00%	1.58%	1.58%	0.32%	0.32%	0.00%	0.63%	1.27%	0.00%	0.63%	1.90%	0.95%	0.32%	0.32%	1.58%	6.33%									
Grand Total	1.27%	6.01%	7.28%	4.75%	0.32%	25.95%	31.01%	2.85%	2.22%	2.22%	7.28%	11.08%	0.32%	23.42%	34.81%	8.86%	1.27%	9.49%	19.62%	100.00%									

- (b) Patterns of star rating conditioned on type, market cap and risk:

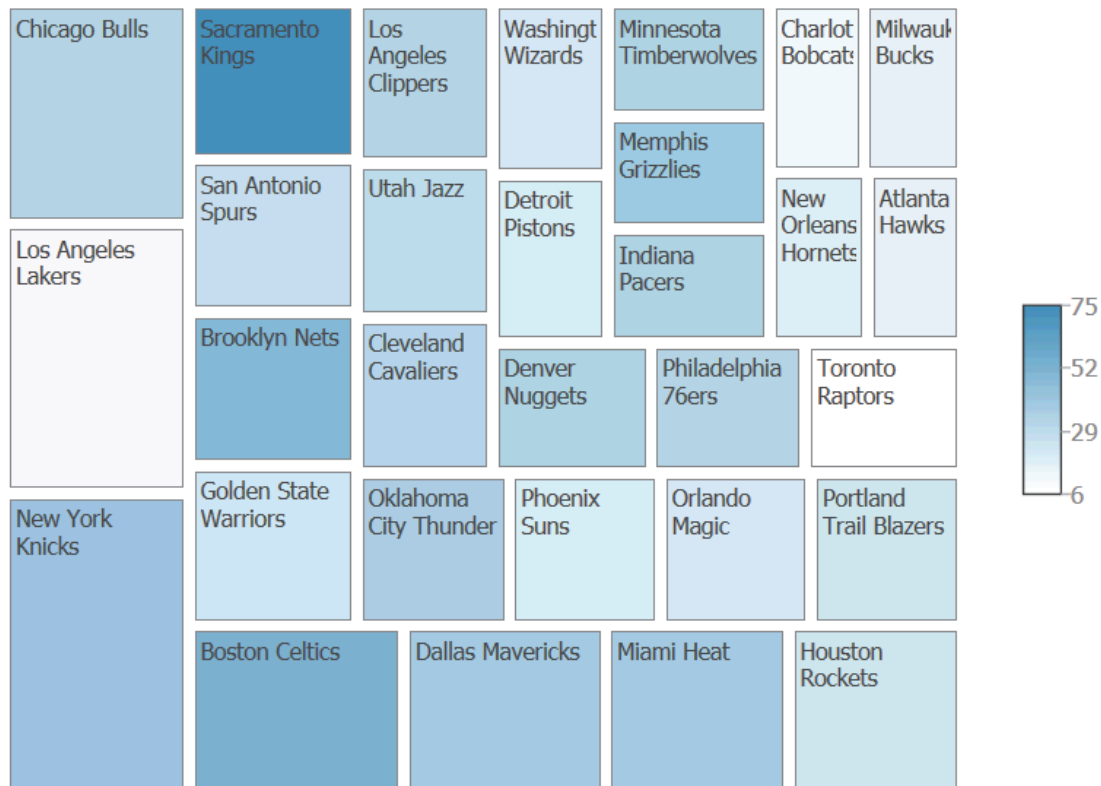
From Problem 2.58 (b), we know that the growth funds as a group, most are rated as four-star, followed by three-star, two-star, five-star and one-star. The pattern of star rating is the same across the different market cap within the growth funds with most of the funds receiving a four-star rating, followed by three-star, two-star, five-star and one-star with the exception of small-cap funds with most of the funds receiving a four-star or three-star rating, followed by two-star, one-star and five-star. If we want to bore further down into the subsets of star-rating among the large-cap growth funds, we see that similar pattern does not hold for the various risk ratings. For example, among the large-cap growth funds with an average-risk rating, most are rated as two-star, followed by equal shares of three-star and four-star, and then equal shares of one-star and five-star. Among the large-cap growth funds with a low-risk rating, most are rated as three-star, followed by four-star, two-star, five-star and one-star.

For the value funds as a group, most are rated as three-star, followed by four-star, two-star, one-star and five-star. Within the value funds, the large-cap funds follow the same pattern as the value funds as a group. If we want to bore further down into the subsets of star-rating among the large-cap value funds, we see that similar pattern does not hold through for the various risk ratings. For example, among the large-cap value funds with an average-risk rating, most are rated as one-star or three-star with none rated as two-star, four-star or five-star.

Patterns of market cap conditioned on type, risk and star-rating:

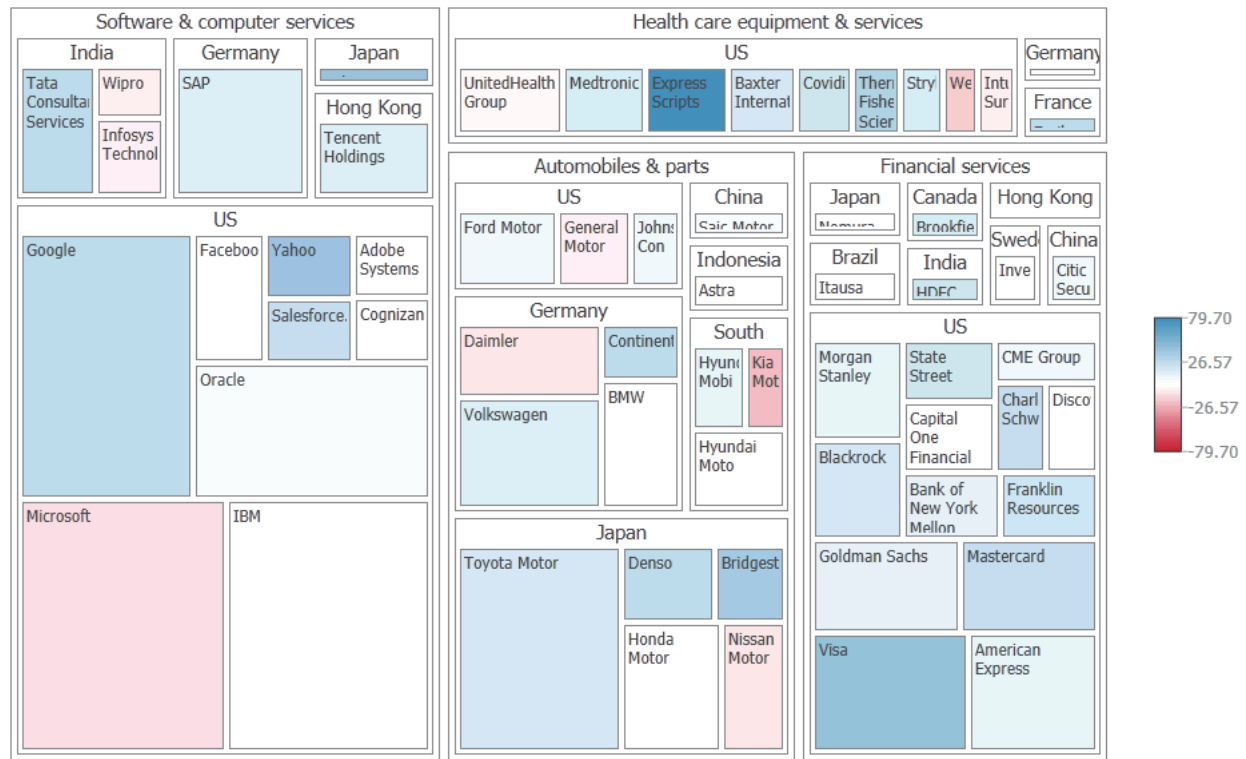
Again, from Problem 2.58 (b), we know that most of the growth funds are large-cap, followed by mid-cap and small-cap. The pattern is similar among the five-star, four-star, three-star and two-star growth funds but among the one-star growth funds, most are small-cap, followed by large-cap and mid-cap. If we bore further down into the subsets of risk-rating, we see that for all the star-ratings, the low-risk growth funds have the most large-cap, followed by mid-cap and then small-cap. However, similar pattern does not hold through among the average-risk and high-risk funds with one-star through five-star ratings.

2.62 (a)



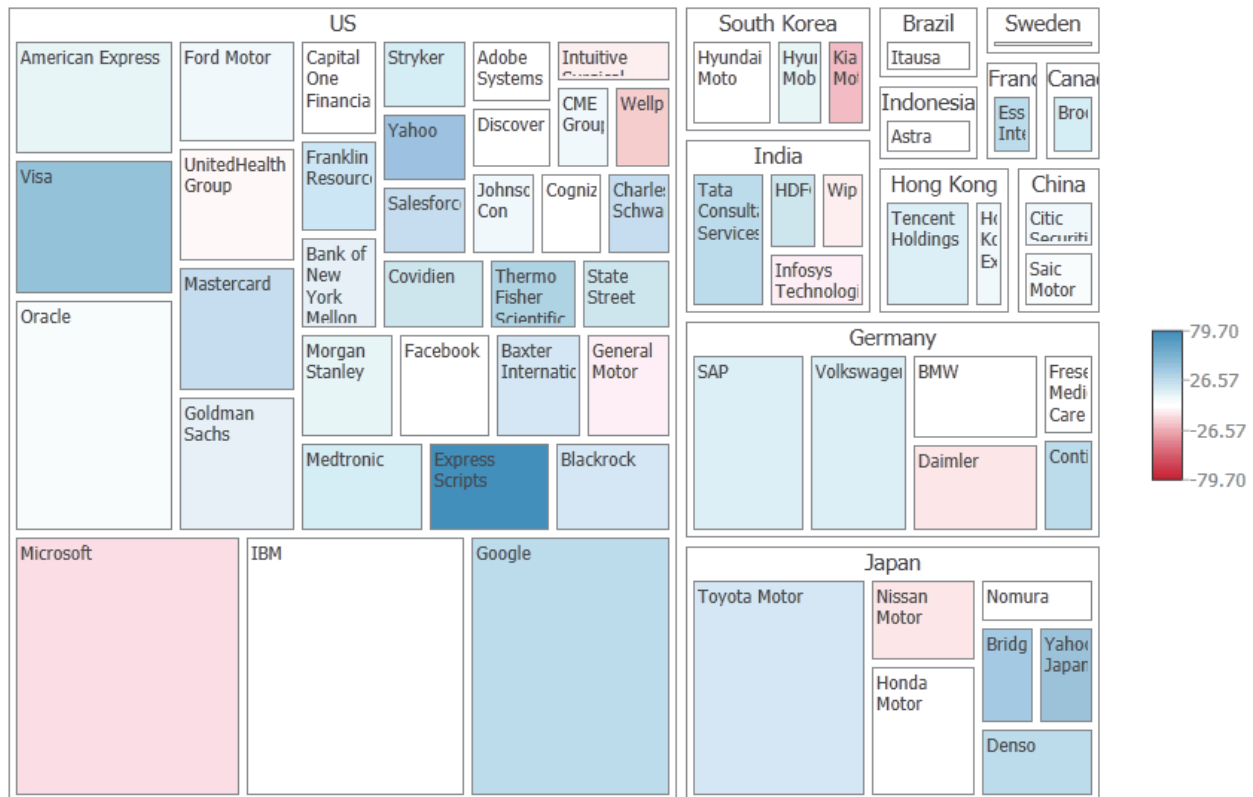
(b) The values of the teams varied from \$312 million for the Milwaukee Bucks to \$1,100 million for the New York Knicks. The change in values was not consistent across the teams. The two most valuable teams, the Los Angeles Lakers, and the New York Knicks had very different increases in value (11% and 41% respectively.)

2.63 (a)



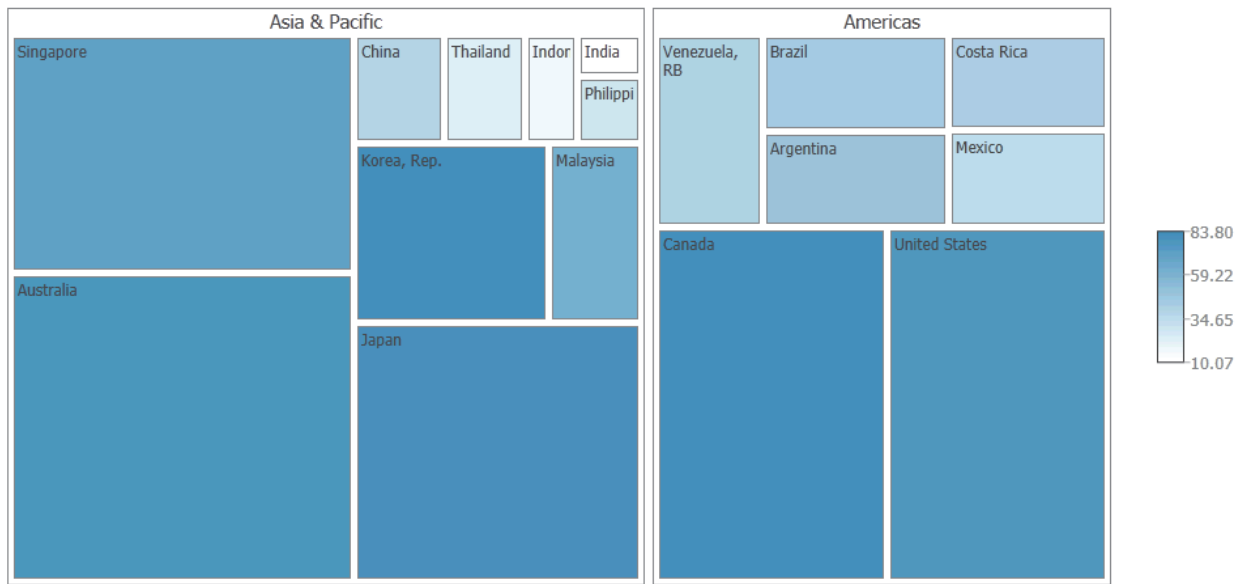
(b) The “financial services” sector seems to have the best gains while the “automobiles & parts” and “software & computer services” sectors seem to have the worst gain.

2.63 (c)
cont.



- (d) The treemap constructed in (c) allows easier recognition of the following: (i) the variation in market capitalization among companies in a particular country and (ii) in which countries are the larger market cap companies found; (iii) which country seems to have the most/more of the largest market cap companies. The treemap constructed in that in (a) allows easier recognition of the following: (i) the variation in market capitalization among companies in a particular country and a particular sector and (ii) in which countries and sectors are the larger market cap companies found; (iii) which countries and sectors seem to have the most/more of the largest market cap companies.

2.64 (a)



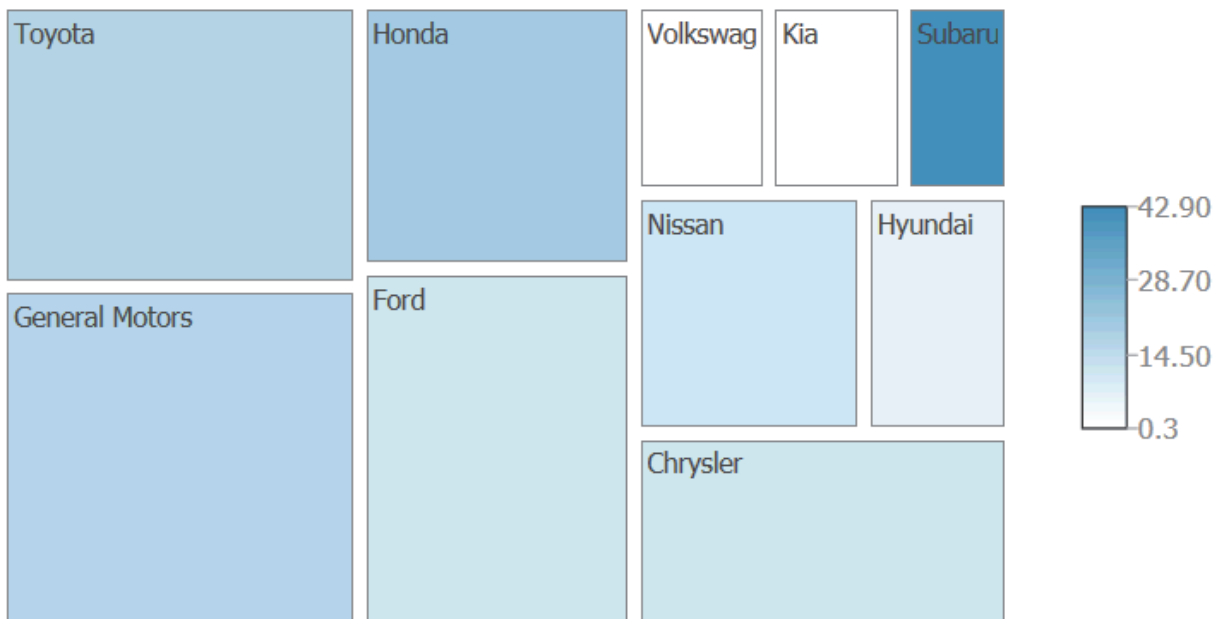
(b)



- (c) Almost all the countries that had lower GDP had lower Internet use except for the Republic of Korea. The pattern of mobile cellular subscriptions does not seem to depend on the GDP of the country.

78 Chapter 2: Organizing and Visualizing Variables

2.65 (a)



(b) General Motors has the highest sales at 234,071 while Subaru has the lowest sales at 35,994. The change in sales is the highest for Subaru at 42.9% and lowest for Volkswagen at 0.3%. The change in sales for General Motors is 16.3%.

2.66 (a)

Count of Market Cap		Column Labels					
Row Labels		Five	Four	One	Three	Two	Grand Total
Growth		18	76	16	74	43	227
Large		9	31	5	37	21	103
Mid-Cap		7	28	4	20	13	72
Small		2	17	7	17	9	52
Value		5	22	7	36	19	89
Large		2	13	5	21	9	50
Mid-Cap		1	4		9	5	19
Small		2	5	2	6	5	20
Grand Total		23	98	23	110	62	316

2.66 (b) There are 37 of such funds.

PHStat output of the summary statistics of the variables:

	Assets	Turnover Ratio	Beta	SD	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%	Expense Ratio
Mean	649.0489189	57.6627027	1.071081081	16.97513514	14.87432432	9.263243243	0.802972973	6.178648649	1.418378378
Median	123.8	38	1.1	17.32	14.65	9.25	0.84	6.4	1.21
Mode	#N/A	13	1.1	17.93	11.8	9.46	1.22	6.81	1.07
Minimum	0.1	0	0.85	13.29	9.16	6.62	-3.27	0.9	0.58
Maximum	6190	413	1.25	19.9	20.85	12.59	3.36	9.69	6.97
Range	6189.9	413	0.4	6.61	11.69	5.97	6.63	8.79	6.39
Variance	1650071.1335	5340.2103	0.0069	2.1347	7.3247	1.5361	1.7426	2.8816	1.0450
Standard Deviation	1284.5509	73.0767	0.0833	1.4610	2.7064	1.2394	1.3201	1.6975	1.0223
Coeff. of Variation	197.91%	126.73%	7.77%	8.61%	18.20%	13.38%	164.40%	27.47%	72.07%
Skewness	3.0908	3.4983	-1.0282	-0.7817	0.1305	0.2500	-0.8352	-1.1914	4.6897
Kurtosis	10.4021	15.5456	1.1615	0.5130	-0.3423	0.2923	1.8633	2.6974	25.3885
Count	37	37	37	37	37	37	37	37	37
Standard Error	211.1789	12.0137	0.0137	0.2402	0.4449	0.2038	0.2170	0.2791	0.1681

2.67 (a)

Count of Market Cap		Column Labels				
Row Labels		Five				
		Four	One	Three	Two	Grand Total
<input checked="" type="checkbox"/> Large		11	44	10	58	30
Average		1	2	3	4	4
High				1	3	4
Low		10	42	6	54	23
<input checked="" type="checkbox"/> Mid-Cap		8	32	4	29	18
Average		2	7	3	13	14
Low		6	25	1	16	4
<input checked="" type="checkbox"/> Small		4	22	9	23	14
Average		1	6	3	18	10
High			1	6	1	1
Low		3	15		4	3
Grand Total		23	98	23	110	62
						316

(b) There is none of such fund.

80 Chapter 2: Organizing and Visualizing Variables

2.68 (a)

Count of Market Cap		Column Labels				
Row Labels		Five				
		Four	One	Three	Two	Grand Total
Growth		18	76	16	74	43
Average		3	15	6	28	22
High			1	5	1	3
Low		15	60	5	45	18
Value		5	22	7	36	19
Average		1		3	7	6
High				2		1
Low		4	22	2	29	12
Grand Total		23	98	23	110	62
						316

(b) There is only one such fund.

2.69 (a)

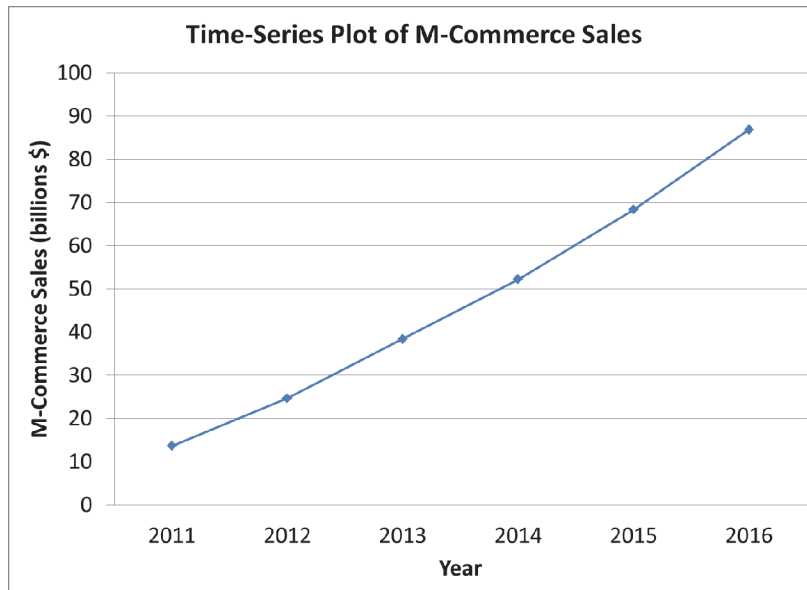
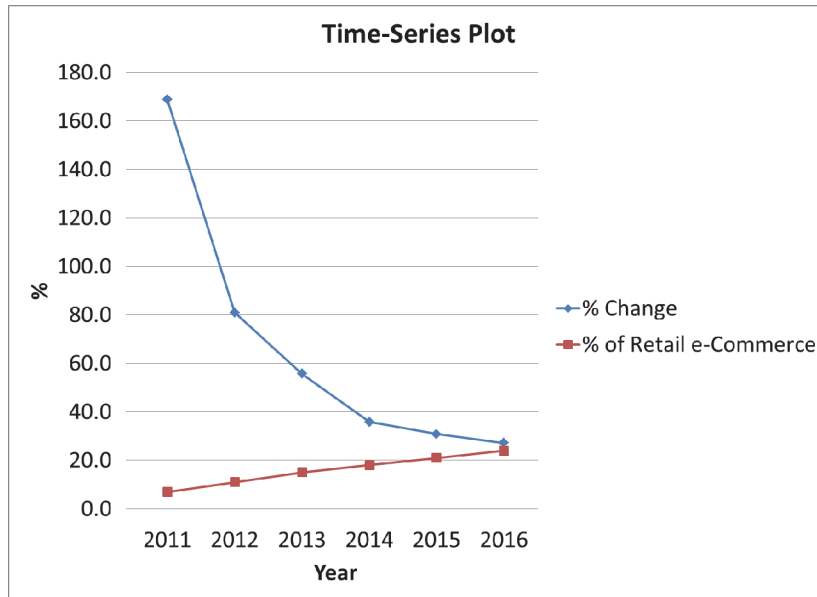
Count of Market Cap		Column Labels		
Row Labels		Average		
		High	Low	Grand Total
Growth		74	10	143
Large		10	3	90
Mid-Cap		35		37
Small		29	7	16
Value		17	3	69
Large		4	1	45
Mid-Cap		4		15
Small		9	2	9
Grand Total		91	13	212
				316

(b) There are 3 of such funds.

PHStat output of the summary statistics of the variables:

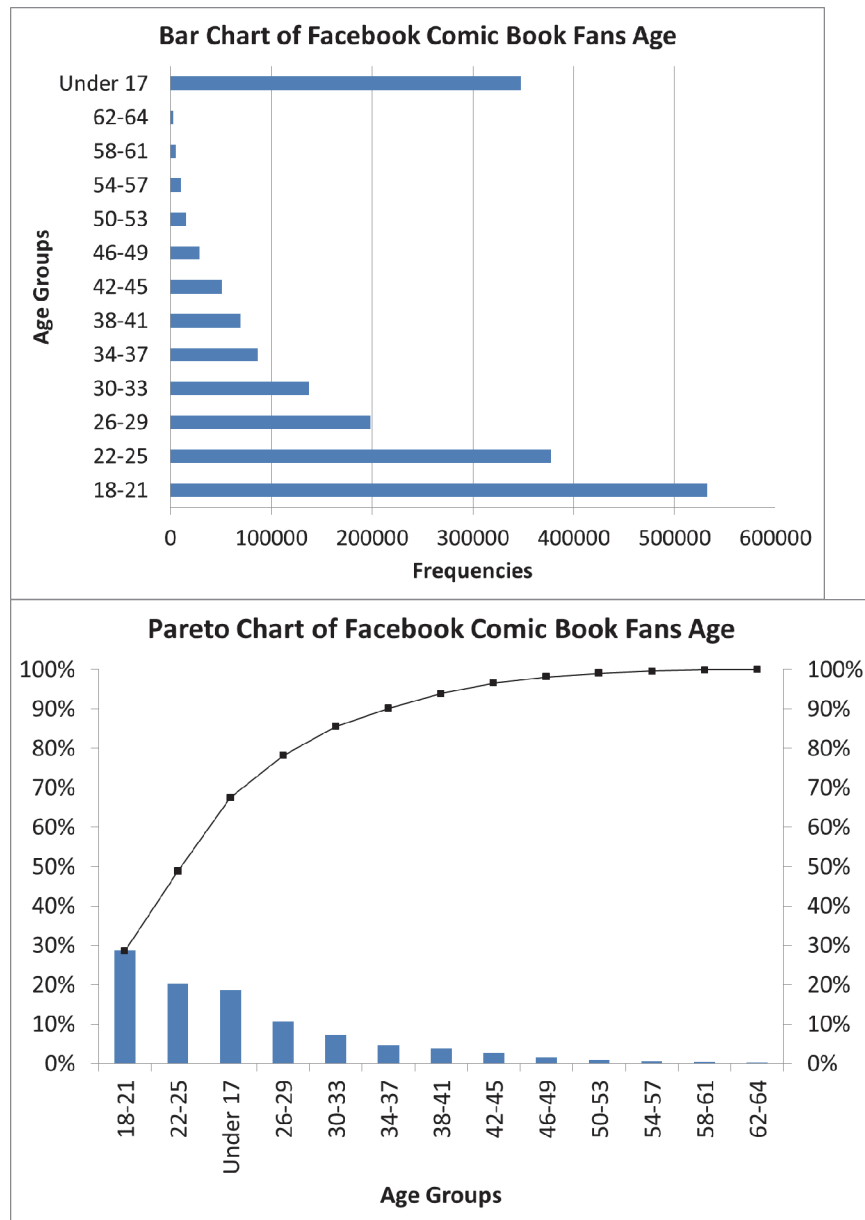
	Assets	Turnover Ratio	Beta	SD	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%	Expense Ratio
Mean	94.43333333	50.66666667	2.223333333	36.34666667	33.72	22.17333333	-0.596666667	8.42	1.85
Median	118.5	5	2.24	36.63	33.78	21.91	1.53	12.55	1.9
Mode	#N/A	#N/A	2.24	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Minimum	23.3	0	2.19	35.72	33.4	21.89	-4.89	0.02	1.73
Maximum	141.5	147	2.24	36.69	33.98	22.72	1.57	12.69	1.92
Range	118.2	147	0.05	0.97	0.58	0.83	6.46	12.67	0.19
Variance	3927.2133	6966.3333	0.0008	0.2954	0.0868	0.2242	13.8249	52.9249	0.0109
Standard Deviation	62.6675	83.4646	0.0289	0.5435	0.2946	0.4735	3.7182	7.2750	0.1044
Coeff. of Variation	66.36%	164.73%	1.30%	1.50%	0.87%	2.14%	-623.16%	86.40%	5.64%
Skewness	-1.4733	1.7251	-1.7321	-1.7083	-0.8784	1.7286	-1.7318	-1.7313	-1.6608
Kurtosis	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Count	3	3	3	3	3	3	3	3	3
Standard Error	36.1811	48.1883	0.0167	0.3138	0.1701	0.2734	2.1467	4.2002	0.0603

- 2.73 (a) There is a title.
 (b) None of the axes are labeled.
 (c)

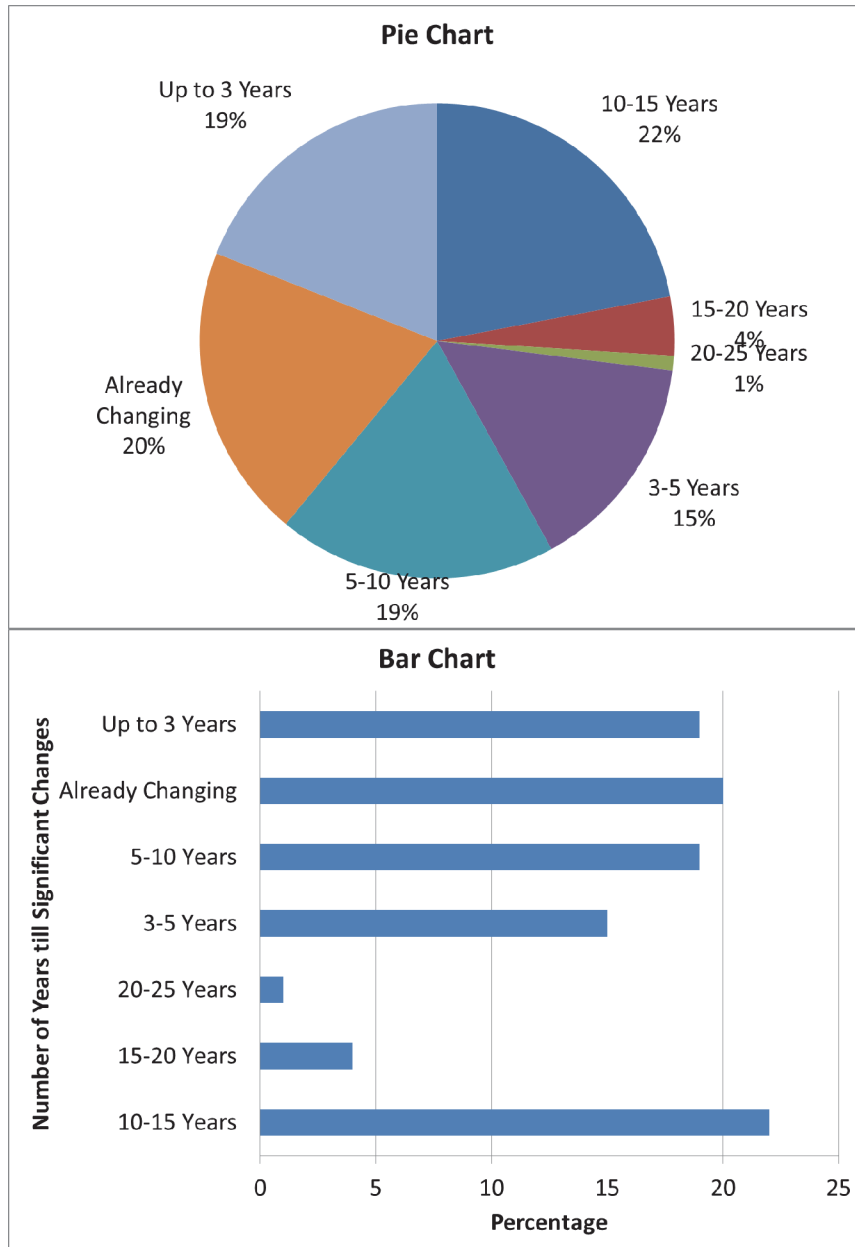


82 Chapter 2: Organizing and Visualizing Variables

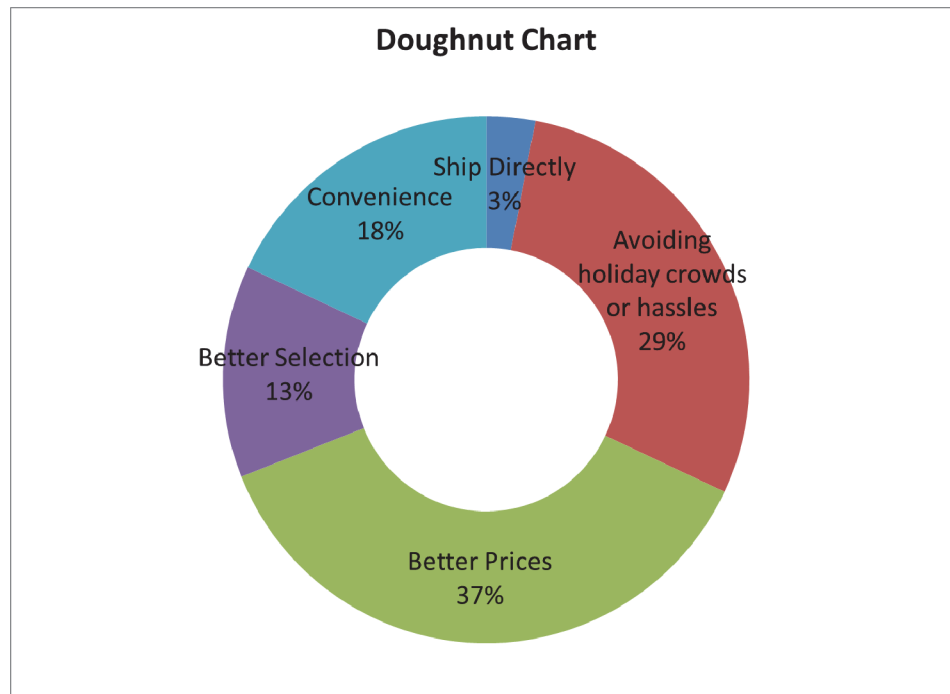
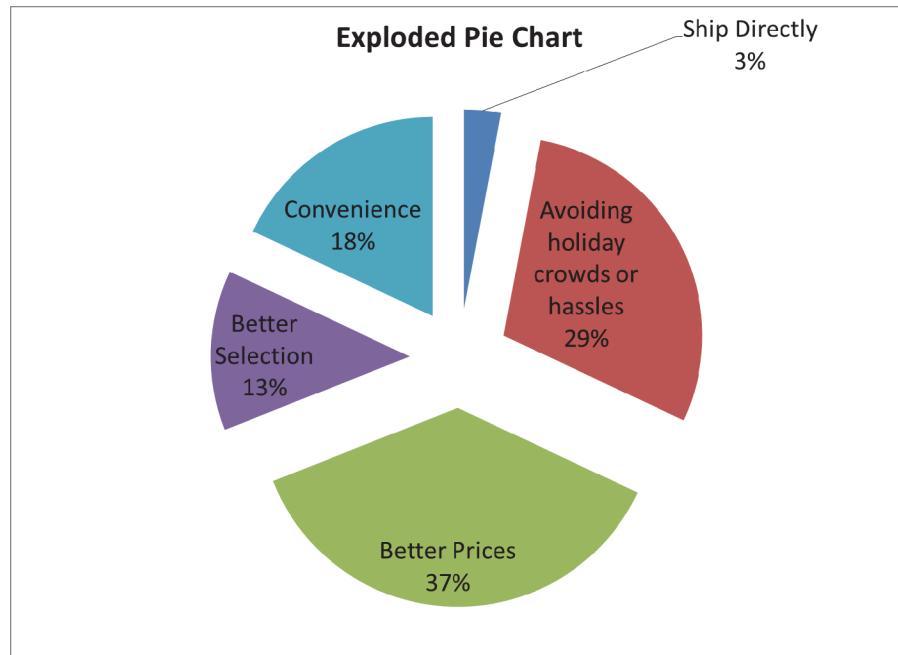
- 2.74 (a) There is a title.
(b) The simplest possible visualization is not used.
(c)



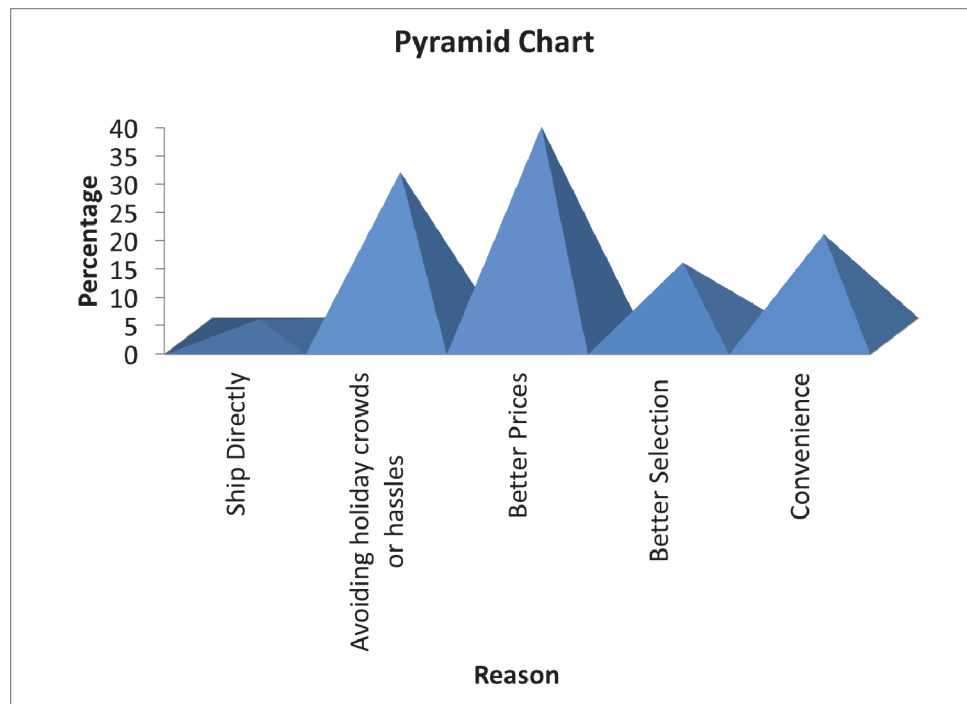
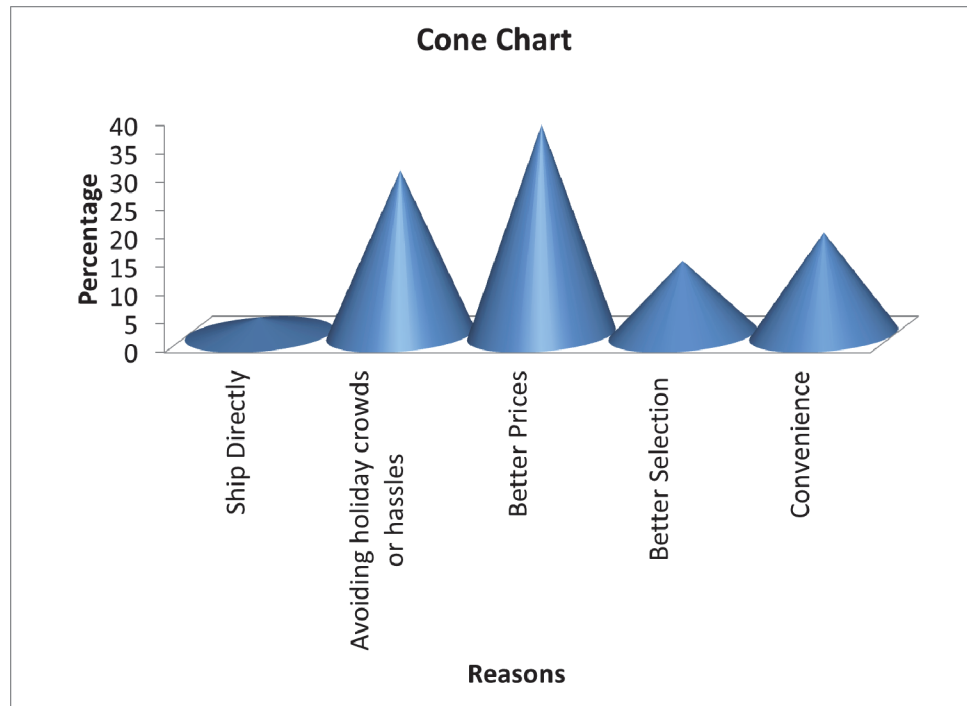
- 2.75 (a) None.
 (b) The use of chartjunk.
 (c)



2.77 (a)

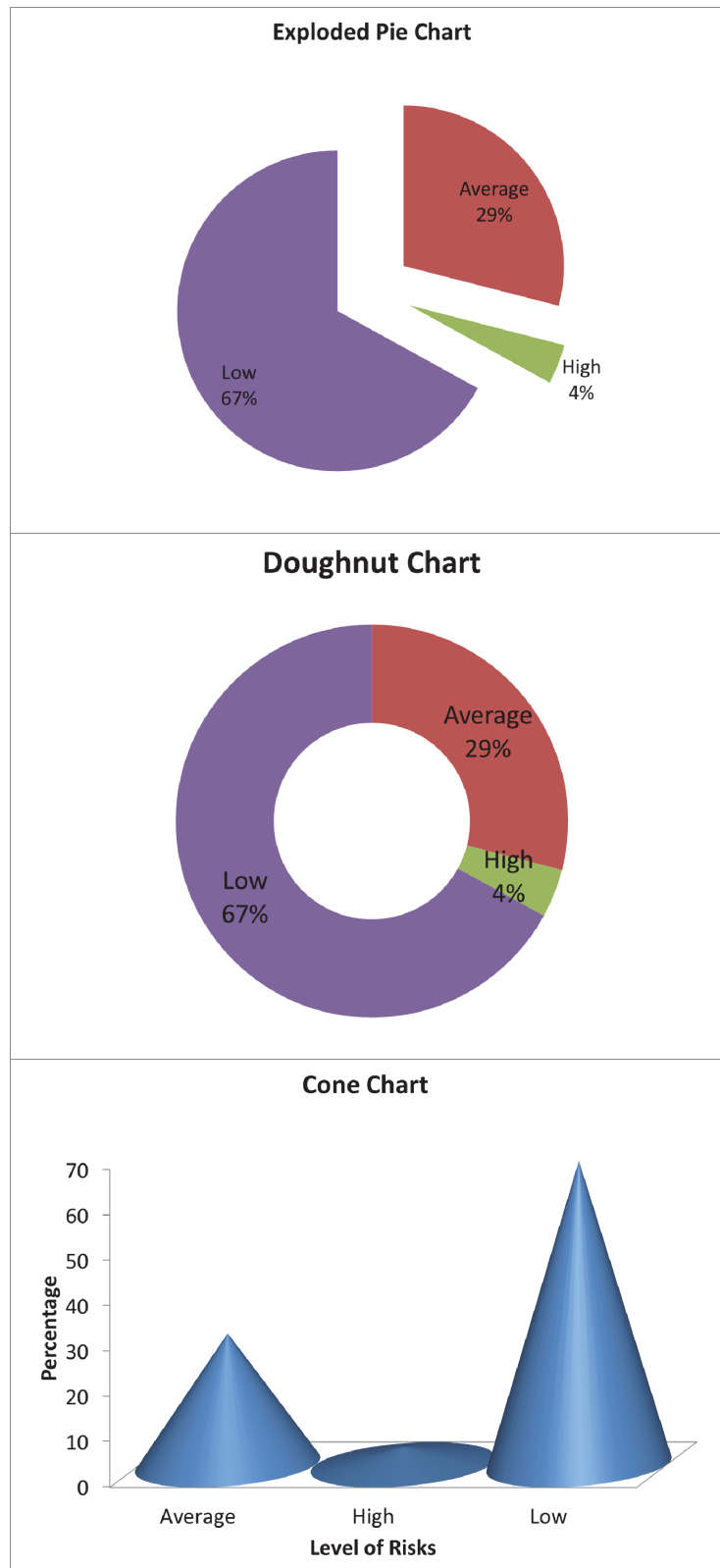


2.77 (a)
cont.

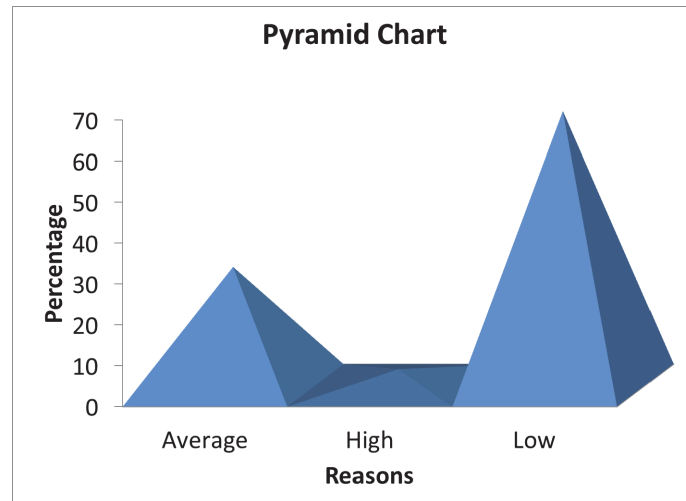


- (b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.78 (a)



2.78 (a)
cont.

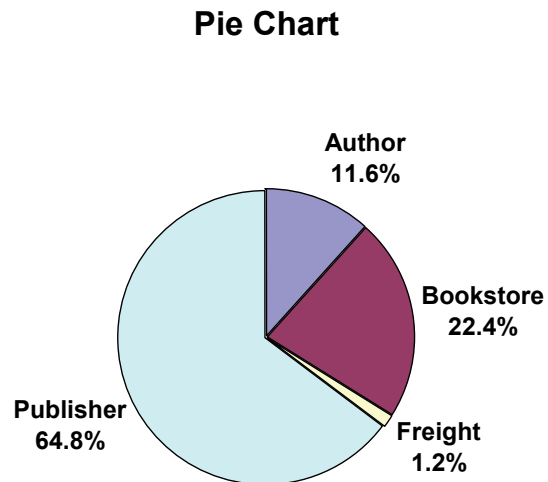
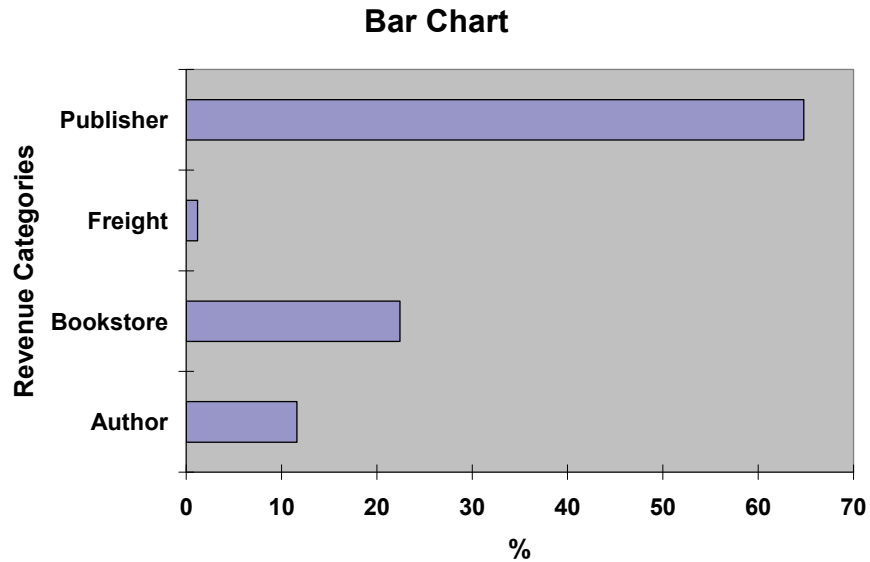


- (b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.
- 2.79 A histogram uses bars to represent each class while a polygon uses a single point. The histogram should be used for only one group, while several polygons can be plotted on a single graph.
- 2.80 A summary table allows one to determine the frequency or percentage of occurrences in each category.
- 2.81 A bar chart is useful for comparing categories. A pie chart is useful when examining the portion of the whole that is in each category. A Pareto diagram is useful in focusing on the categories that make up most of the frequencies or percentages.
- 2.82 The bar chart for categorical data is plotted with the categories on the vertical axis and the frequencies or percentages on the horizontal axis. In addition, there is a separation between categories. The histogram is plotted with the class grouping on the horizontal axis and the frequencies or percentages on the vertical axis. This allows one to more easily determine the distribution of the data. In addition, there are no gaps between classes in the histogram.
- 2.83 A time-series plot is a type of scatter diagram with time on the x-axis.
- 2.84 Because the categories are arranged according to frequency or importance, it allows the user to focus attention on the categories that have the greatest frequency or importance.
- 2.85 Percentage breakdowns according to the total percentage, the row percentage, and/or the column percentage allow the interpretation of data in a two-way contingency table from several different perspectives.
- 2.86 A contingency table contains information on two categorical variables whereas a multidimensional table can display information on more than two categorical variables.

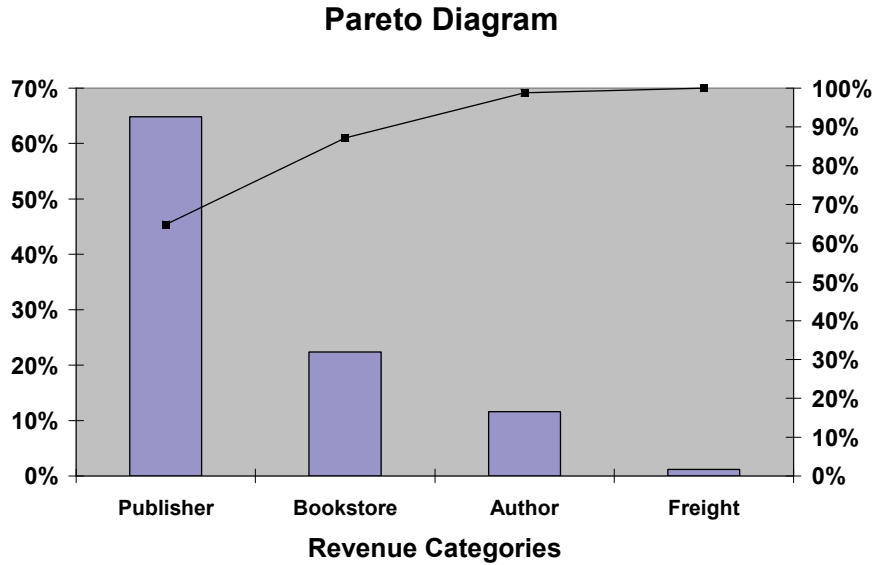
88 Chapter 2: Organizing and Visualizing Variables

2.87 The multidimensional PivotTable can reveal additional patterns that cannot be seen in the contingency table. One can also change the statistic displayed and compute descriptive statistics which can add insight into the data.

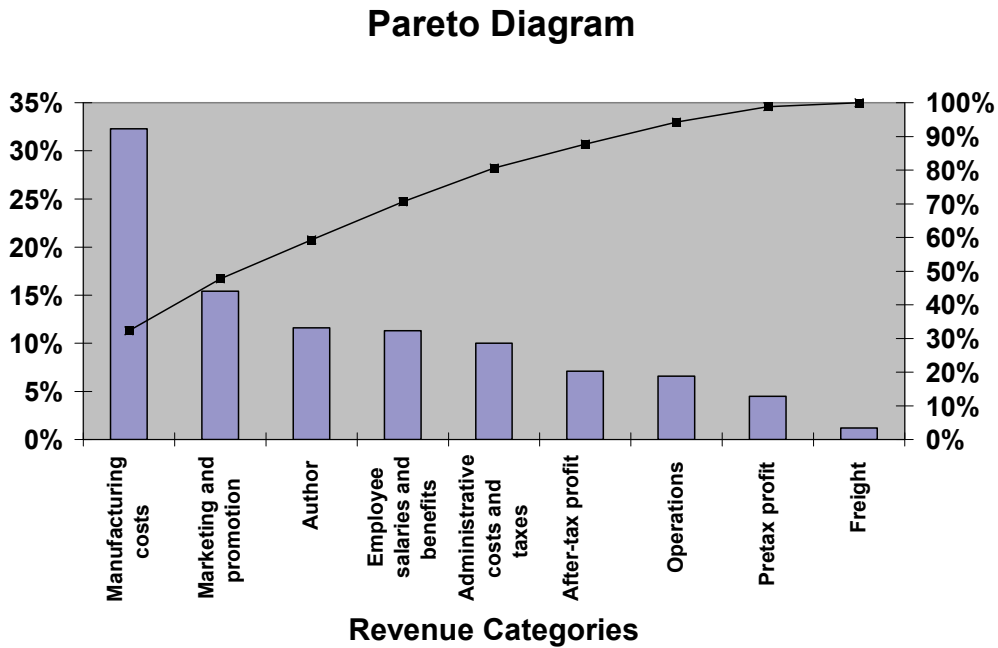
2.88 (a)



2.88 (a)
cont.

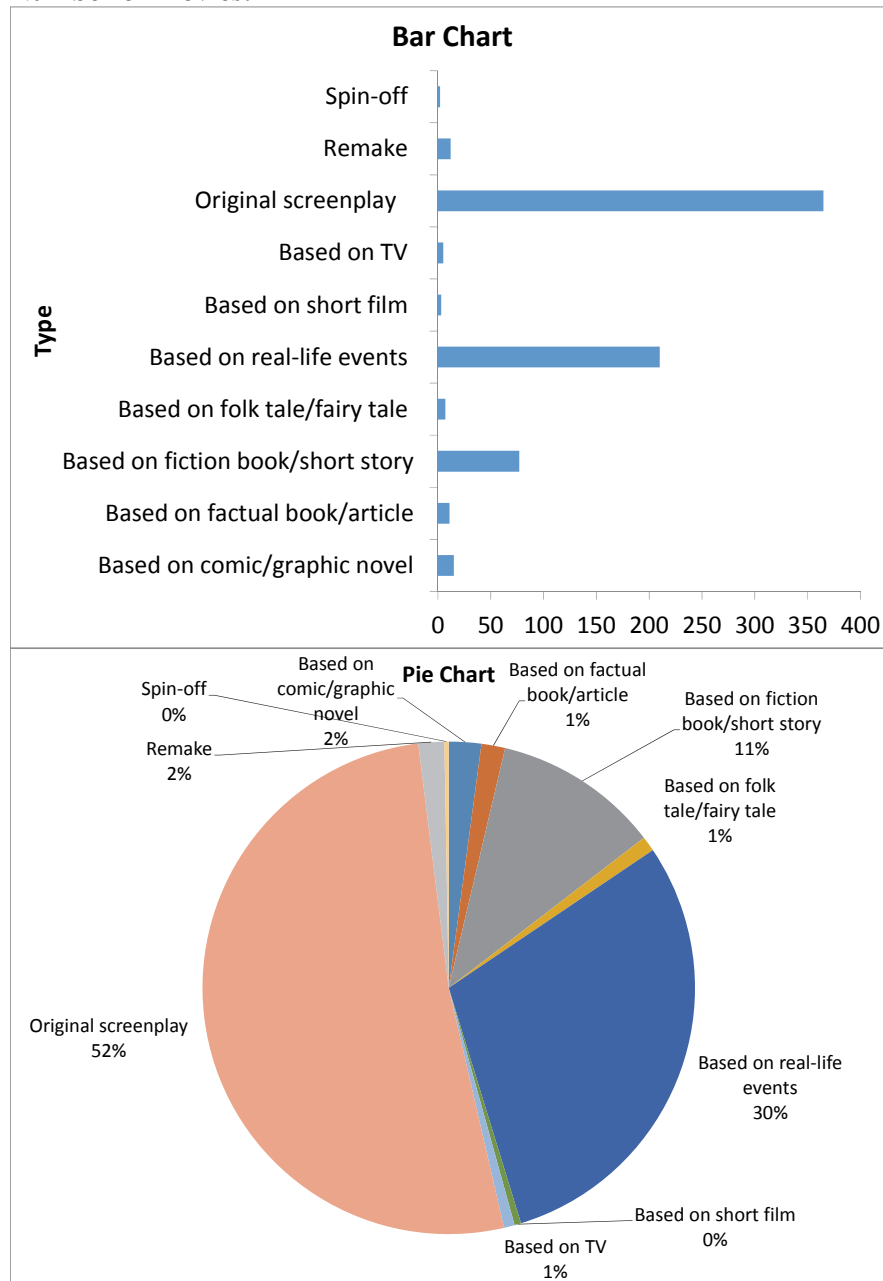


(b)

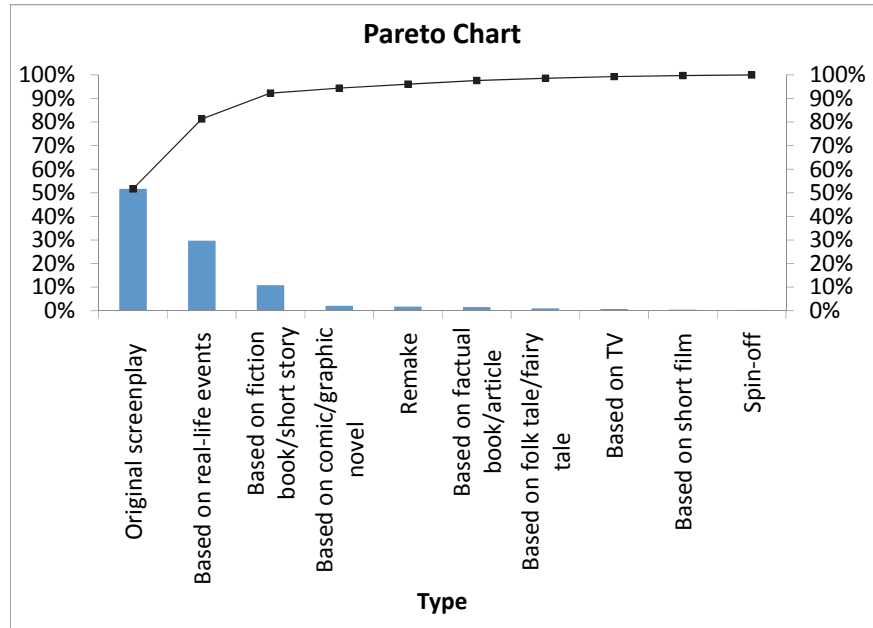


- (c) The publisher gets the largest portion (64.8%) of the revenue. About half (32.3%) of the revenue received by the publisher covers manufacturing costs. The publisher's marketing and promotion account for the next largest share of the revenue, at 15.4%. Author, bookstore employee salaries and benefits, and publisher administrative costs and taxes each account for around 10% of the revenue, whereas the publisher after-tax profit, bookstore operations, bookstore pretax profit, and freight constitute the "trivial few" allocations of the revenue. Yes, the bookstore gets twice the revenue of the authors.

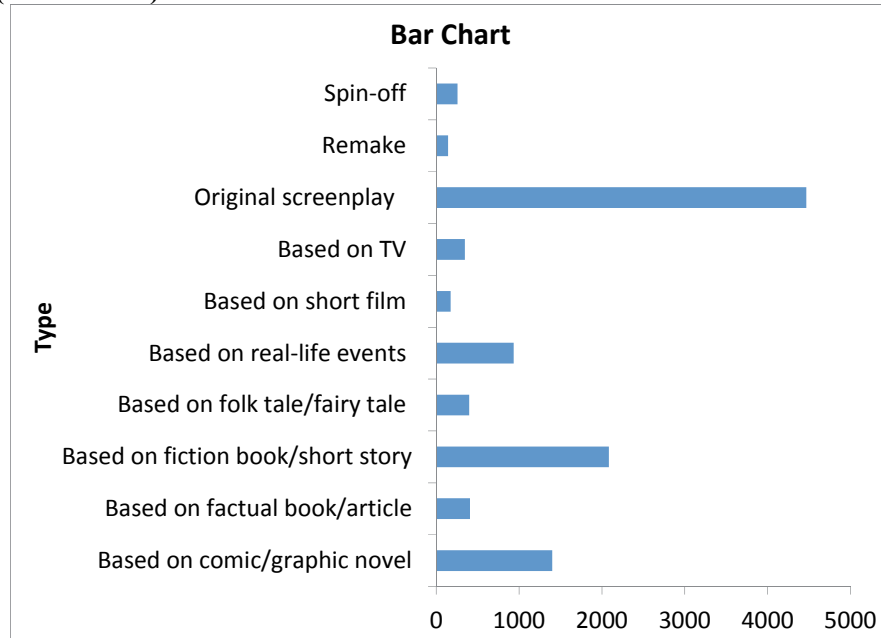
2.89 (a) Number of Movies:



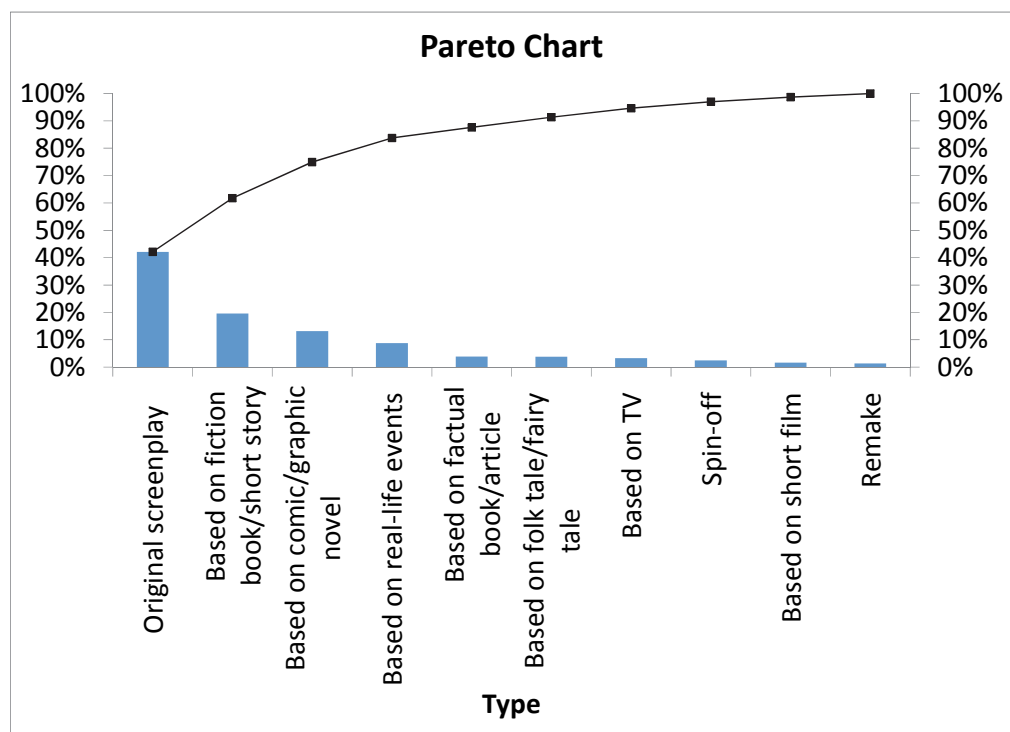
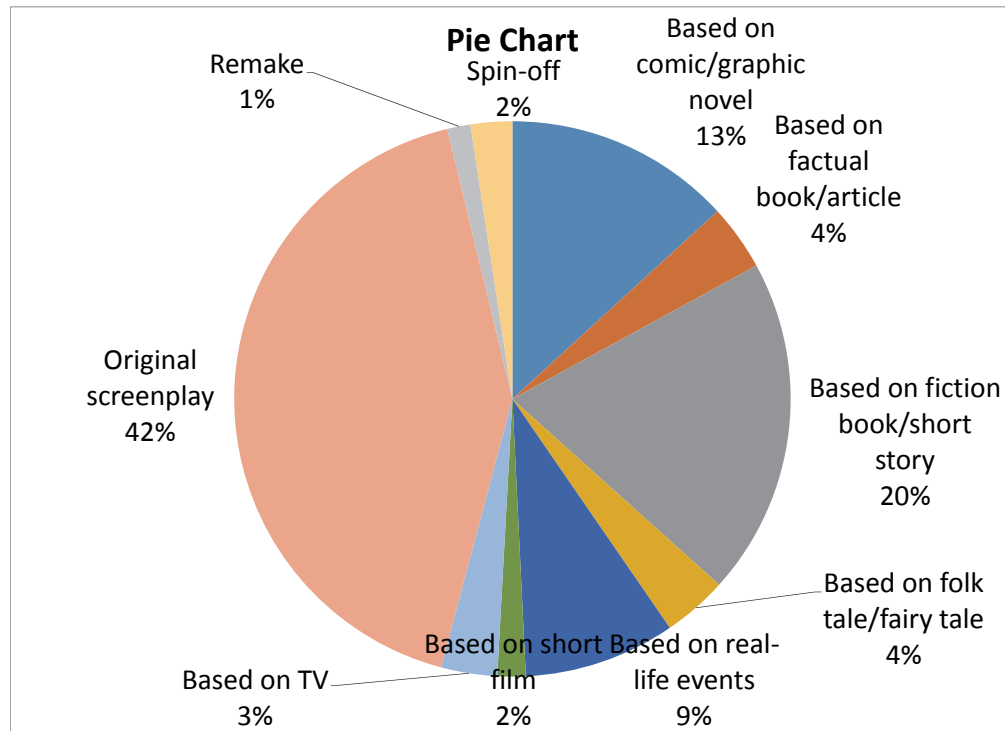
2.89 (a)
cont.



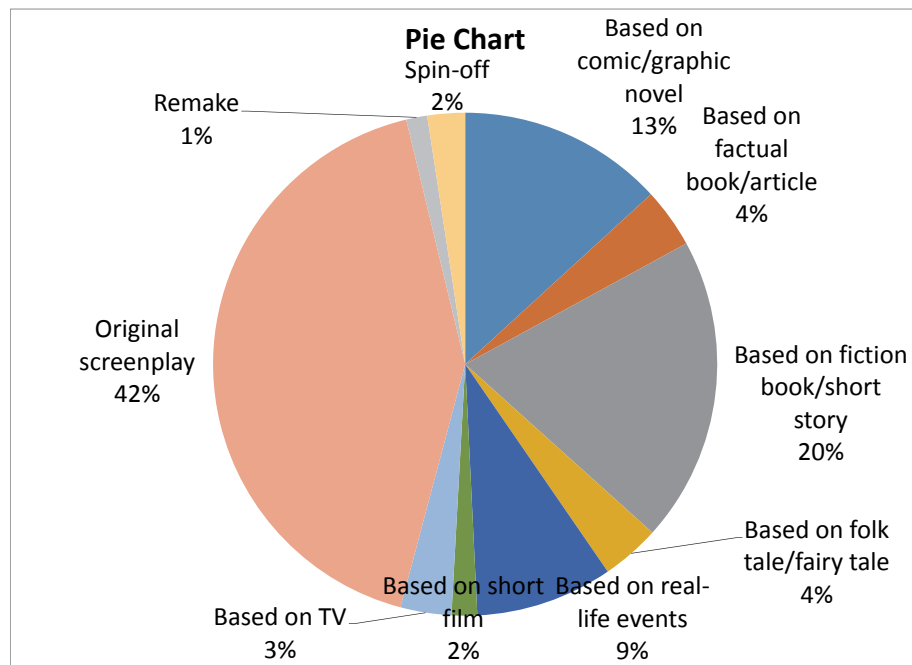
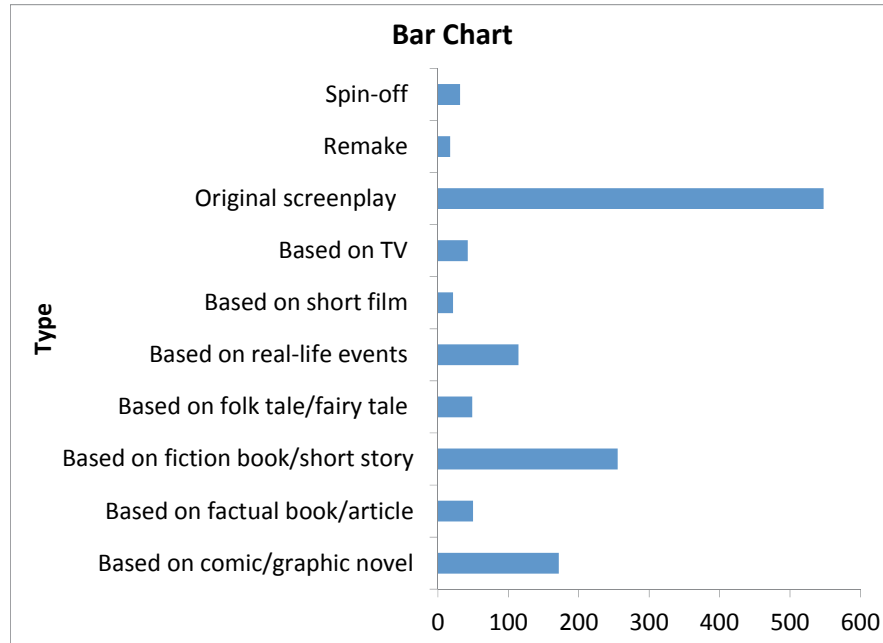
Gross (in \$millions):



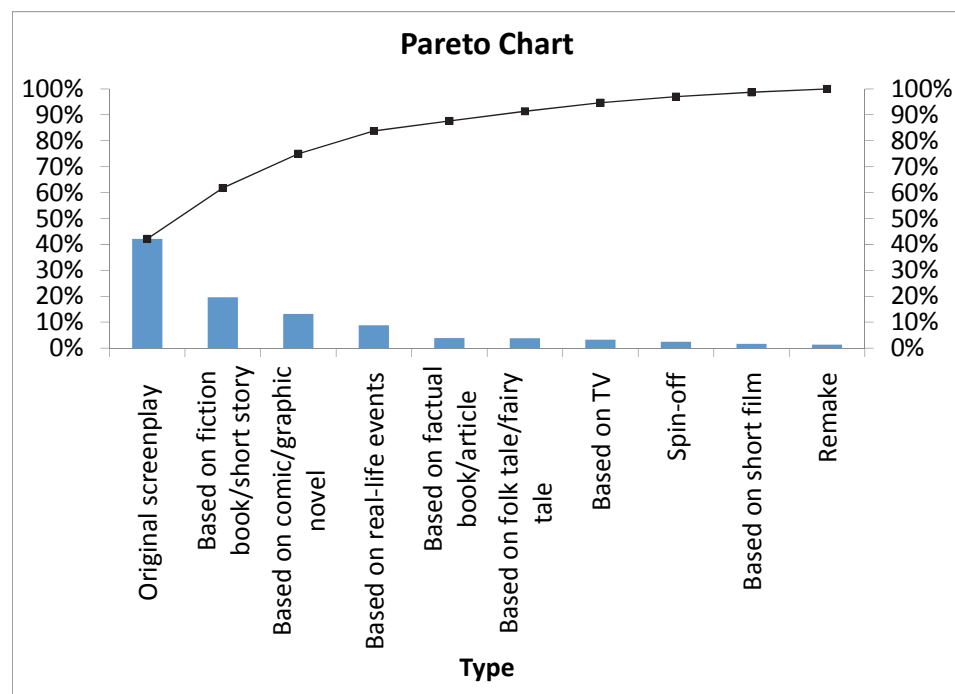
2.89 (a)
cont.



2.89 (a) Number of Tickets Sold (millions):
cont.

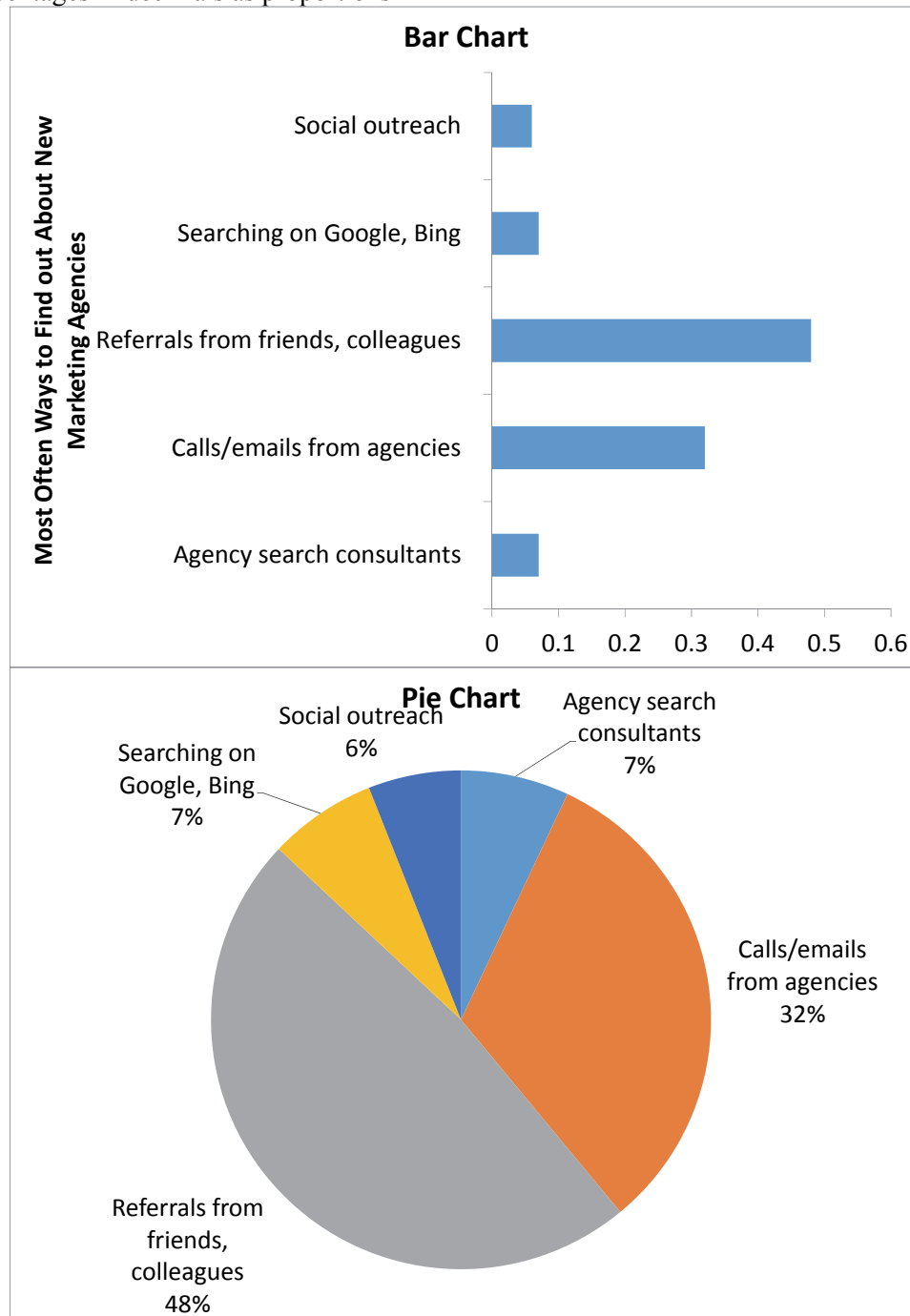


2.89 (a)
cont.

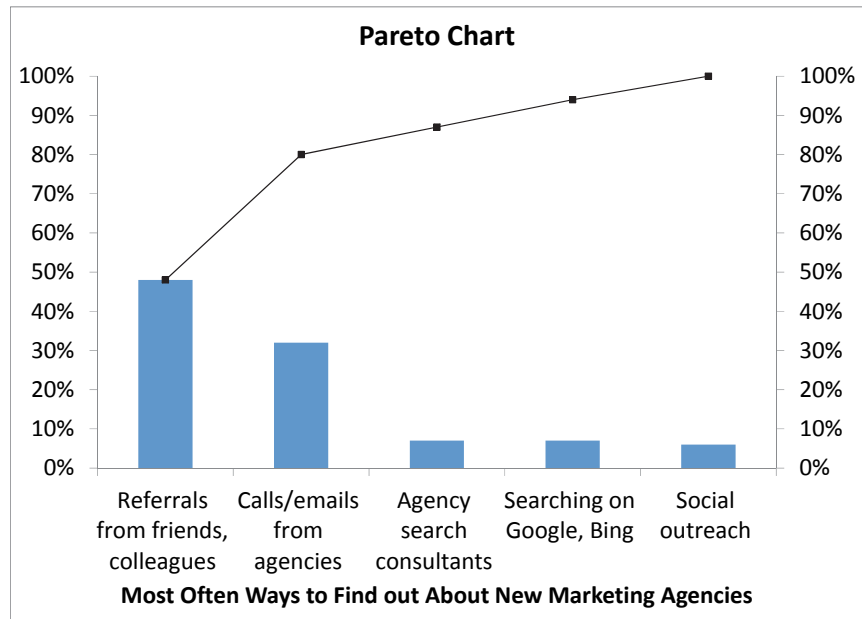


- (b) Based on the Pareto chart for the number of movies, “Original screenplay”, “Based on real life events” and “Based on fiction/short story” are the “vital few” and capture about 92% of the market share. According to the Pareto chart for gross (in \$millions), “Original screenplay”, “Based on fiction book/short story”, “Based on comic/graphic novel” and “Based on real-life events” are the “vital few” and capture about 84% of the market share. According to the Pareto chart for number of tickets sold (in millions), “Original screenplay”, “Based on fiction book/short story”, “Based on comic/graphic novel” and “Based on real-life events” are the “vital few” and capture about 84% of the market share.

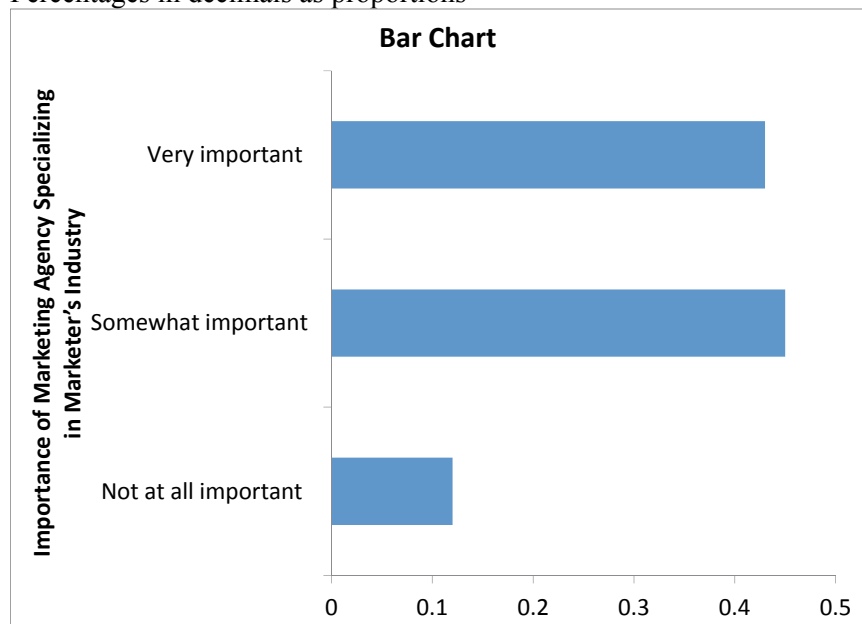
2.90 (a) Percentages in decimals as proportions



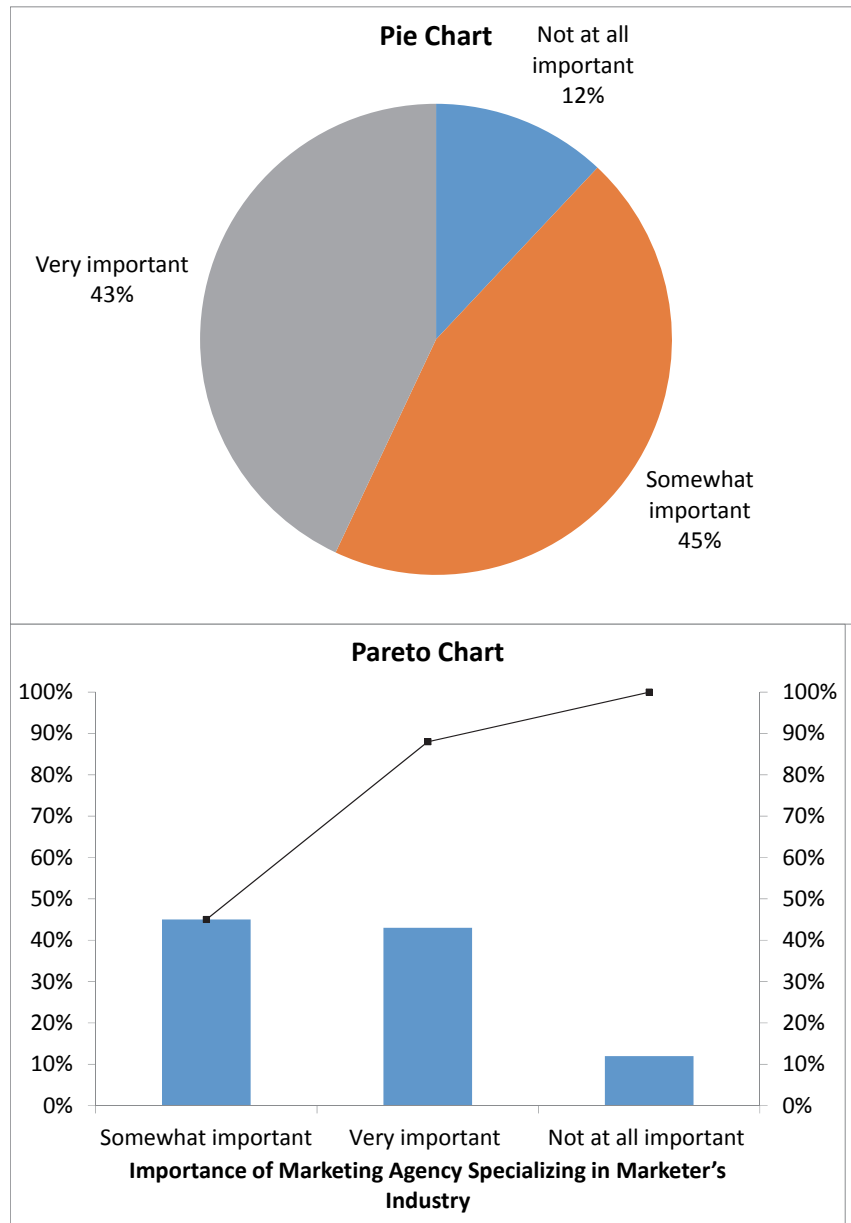
2.90 (a)
cont.



- (b) The pie chart may be best since with only five categories, it enables you to see the portion of the whole in each category.
- (c) Percentages in decimals as proportions



2.90 (c)
cont.



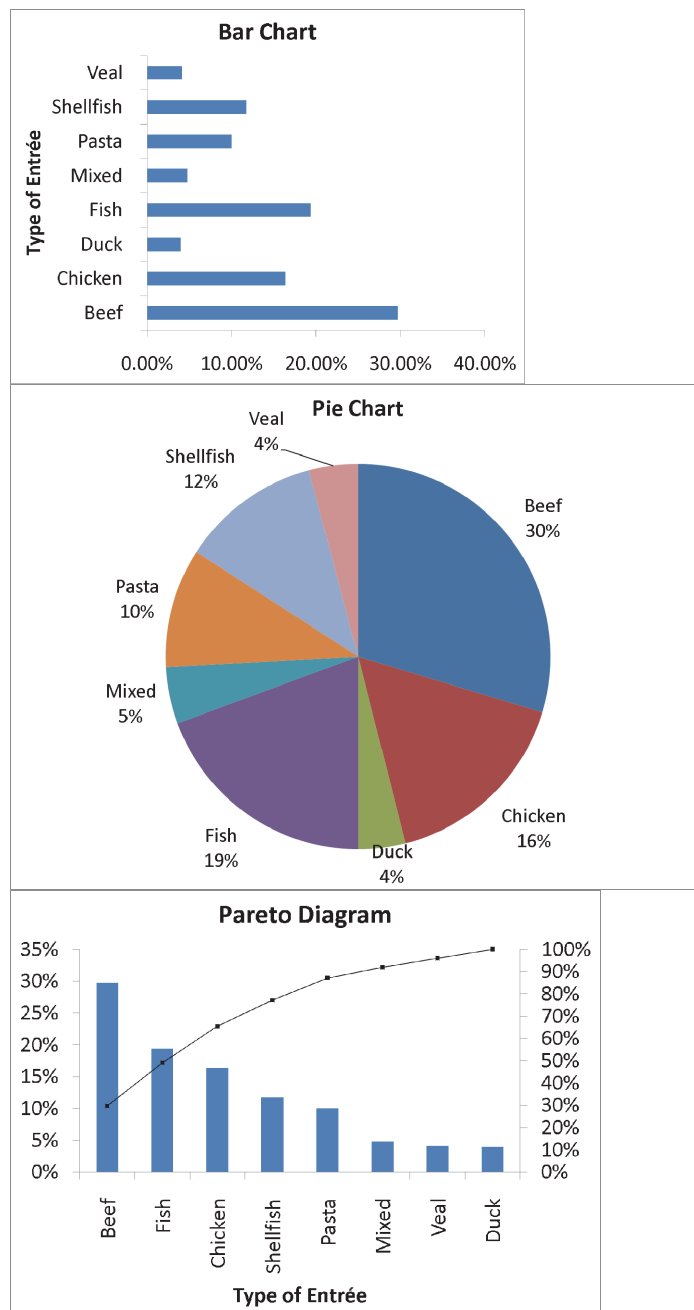
- (d) The pie chart may be best since, with only four categories it enables you to see the portion of the whole in each category.
- (e) Based on the Pareto chart for “Most Often Ways to Find out About New Marketing Agencies”, about 80% of the marketers use “referrals from friends/colleagues” and “calls/emails from agencies” to find out about new marketing agencies for hire. Based on the Pareto chart for “Importance of Marketing Agency Specializing in Marketer’s Industry”, about 88% of the marketers value the marketing agencies that specialize in their industry as “somewhat important” or “very important”.

98 Chapter 2: Organizing and Visualizing Variables

2.91 (a)

Type of Entrée	%	Number S
Beef	29.68%	187
Chicken	16.35%	103
Mixed	4.76%	30
Duck	3.97%	25
Fish	19.37%	122
Pasta	10.00%	63
Shellfish	11.75%	74
Veal	4.13%	26
Total	100.00%	630

(b)



- 2.91 (c) The Pareto diagram has the advantage of offering the cumulative percentage view of the categories and, hence, enables the viewer to separate the "vital few" from the "trivial many".
- (d) Beef and fish account for nearly 50% of all entrees ordered by weekend patrons of a continental restaurant. When chicken is included, nearly two-thirds of the entrees are accounted for.

2.92 (a)

Count of Dessert Ordered		Gender	
Desserts Ordered		Male	Female
Yes		34.25%	65.75%
No		51.65%	48.35%
Grand Total		47.62%	52.38%

Count of Dessert Ordered		Gender	
Desserts Ordered		Male	Female
Yes		16.67%	29.09%
No		83.33%	70.91%
Grand Total		100.00%	100.00%

Count of Dessert Ordered		Gender	
Desserts Ordered		Male	Female
Yes		7.94%	15.24%
No		39.68%	37.14%
Grand Total		47.62%	52.38%

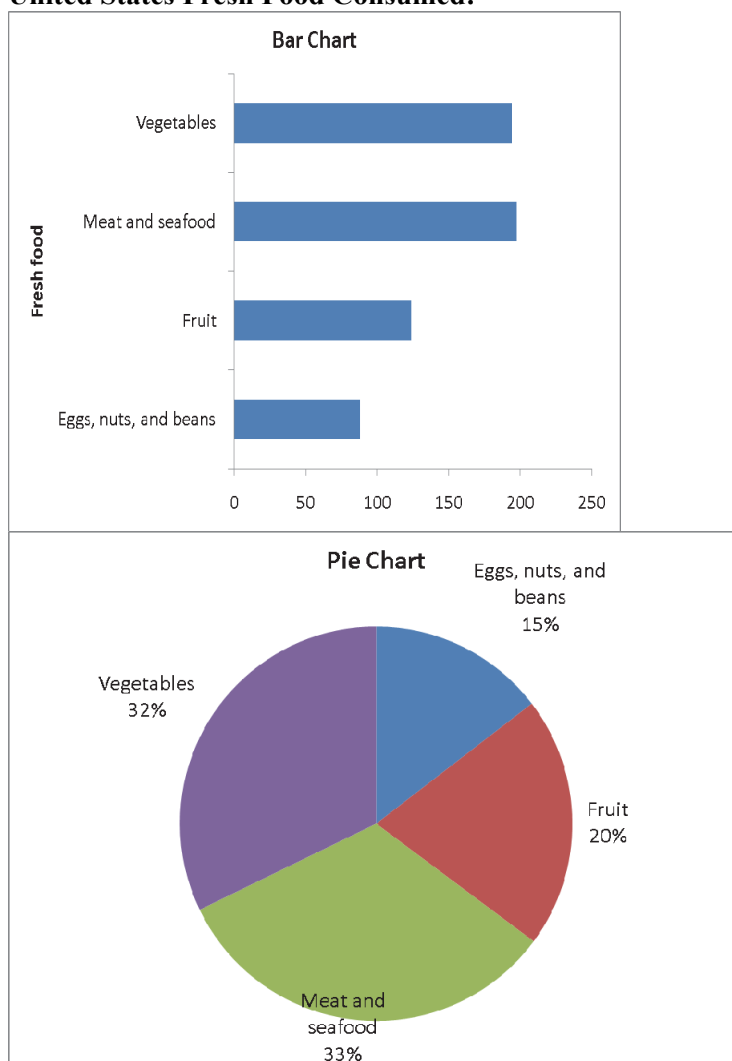
Count of Dessert Ordered		Beef Entrée	
Dessert Ordered		Yes	No
Yes		52.11%	47.89%
No		25.20%	74.80%
Grand Total		31.27%	68.73%

Count of Dessert Ordered		Beef Entrée	
Dessert Ordered		Yes	No
Yes		37.56%	15.70%
No		62.44%	84.30%
Grand Total		100.00%	100.00%

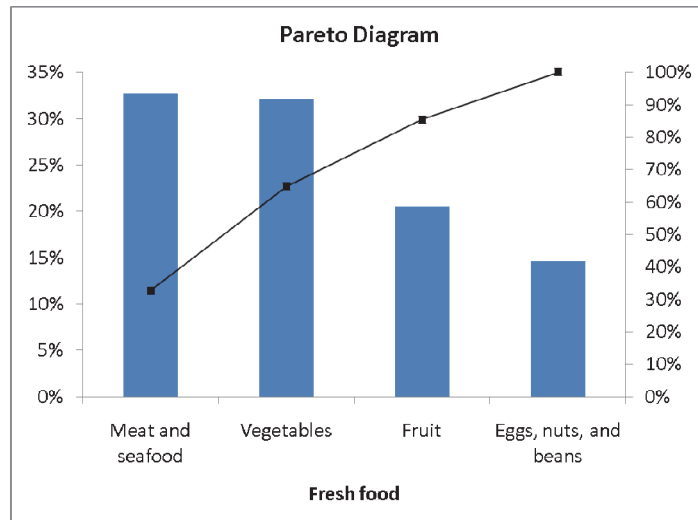
Count of Dessert Ordered		Beef Entrée	
Dessert Ordered		Yes	No
Yes		11.75%	10.79%
No		19.52%	57.94%
Grand Total		31.27%	68.73%

- 2.92 (b) If the owner is interested in finding out the percentage of joint occurrence of gender and ordering of dessert or the percentage of joint occurrence of ordering a beef entrée and a dessert among all patrons, the table of total percentages is most informative. If the owner is interested in the effect of gender on ordering of dessert or the effect of ordering a beef entrée on the ordering of dessert, the table of column percentages will be most informative. Since dessert will usually be ordered after the main entree and the owner has no direct control over the gender of patrons, the table of row percentages is not very useful here.
- (c) 16.67% of the men sampled ordered desserts compared to 29.09% of the women. Women are almost twice as likely to order desserts as men. 37.56% of the patrons ordering a beef entree ordered dessert compared to less than 15.7% of patrons ordering all other entrees. Patrons ordering beef are better than 2.3 times as likely to order dessert as patrons ordering any other entree.

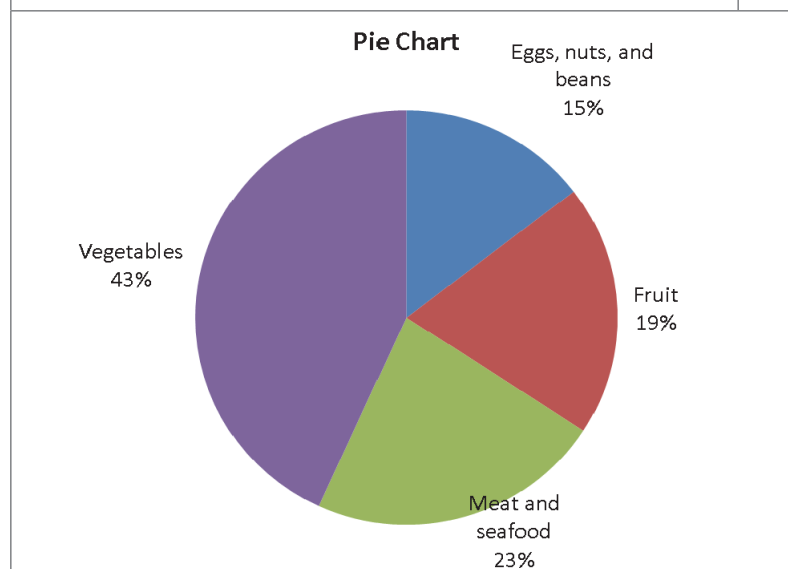
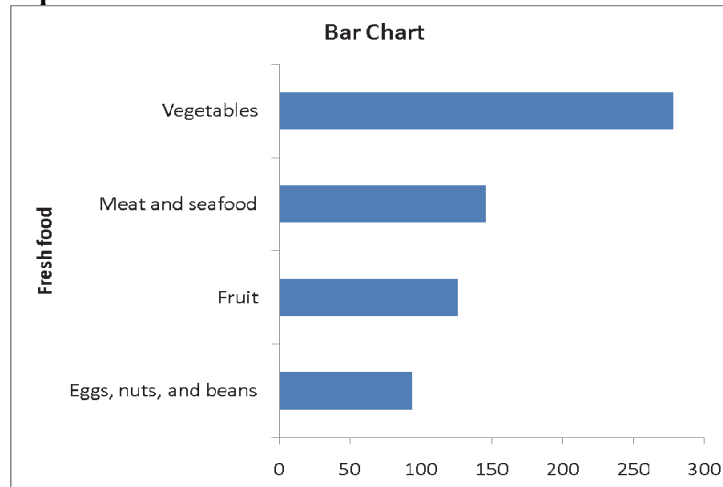
- 2.93 (a) **United States Fresh Food Consumed:**



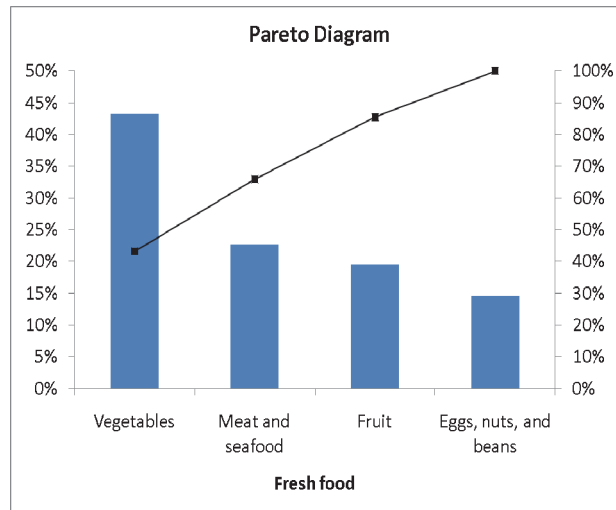
2.93 (a)
cont.



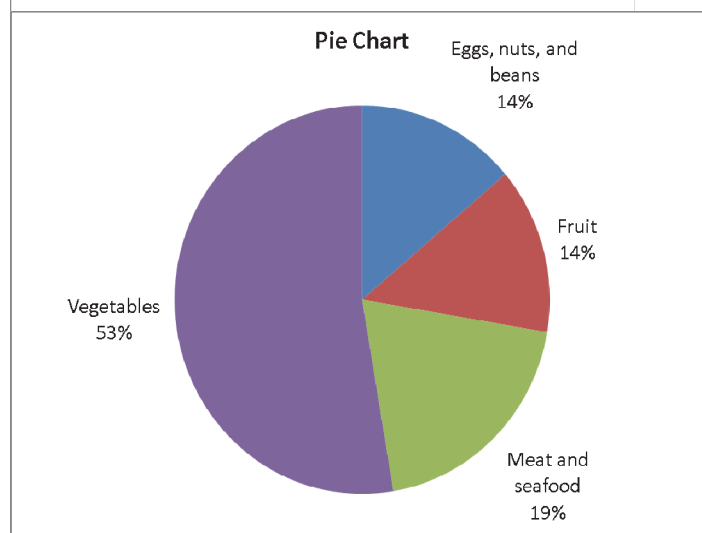
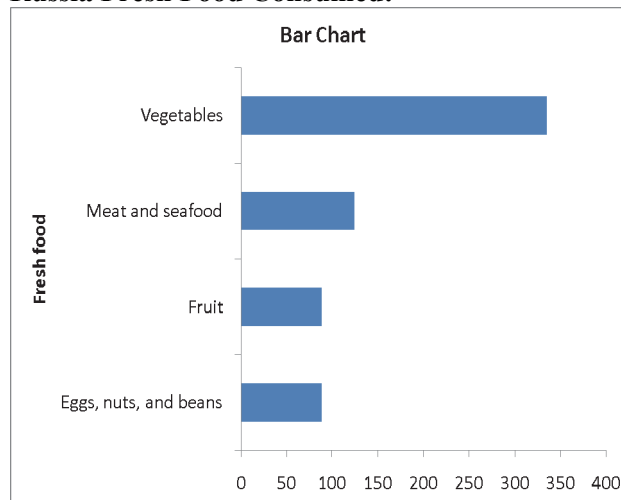
Japan Fresh Food Consumed:



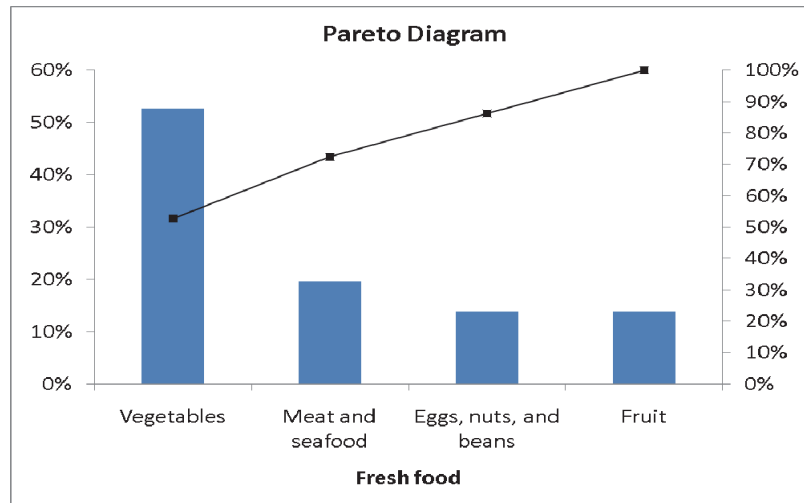
2.93 (a)
cont.



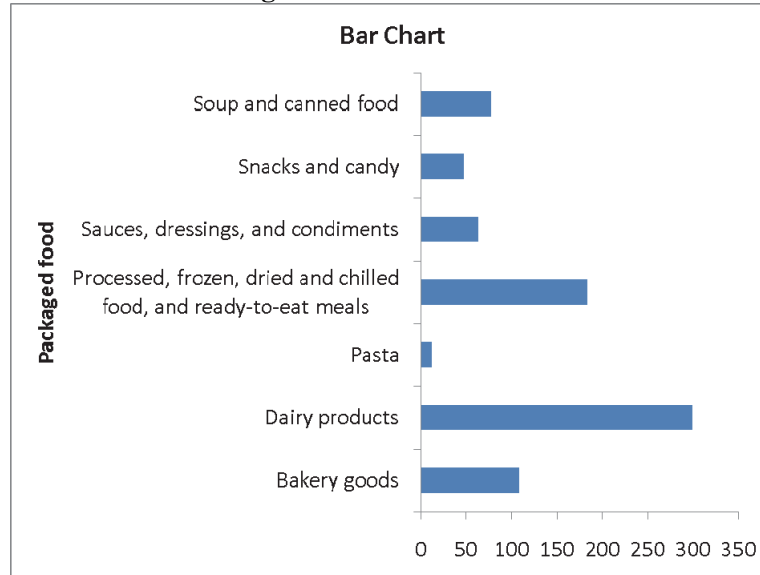
Russia Fresh Food Consumed:



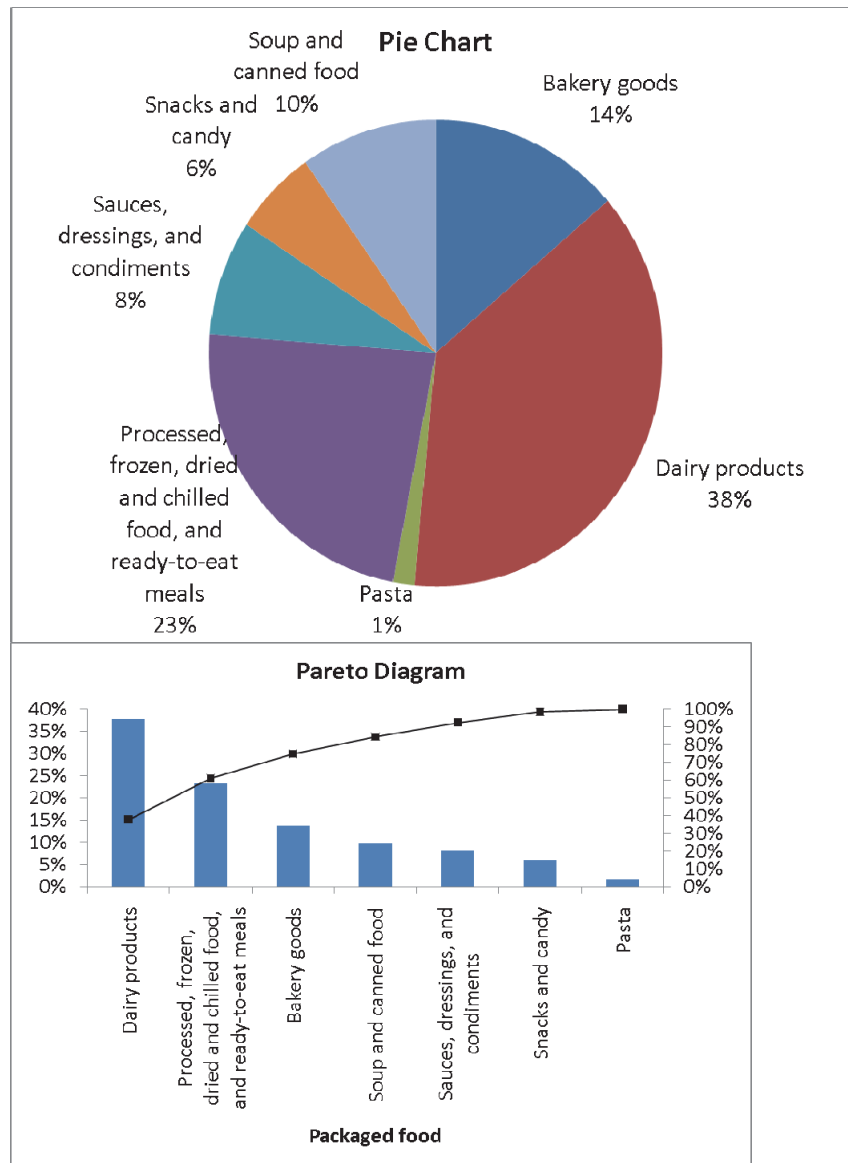
2.93 (a)
cont.



(b) **United States Packaged Food Consumed:**

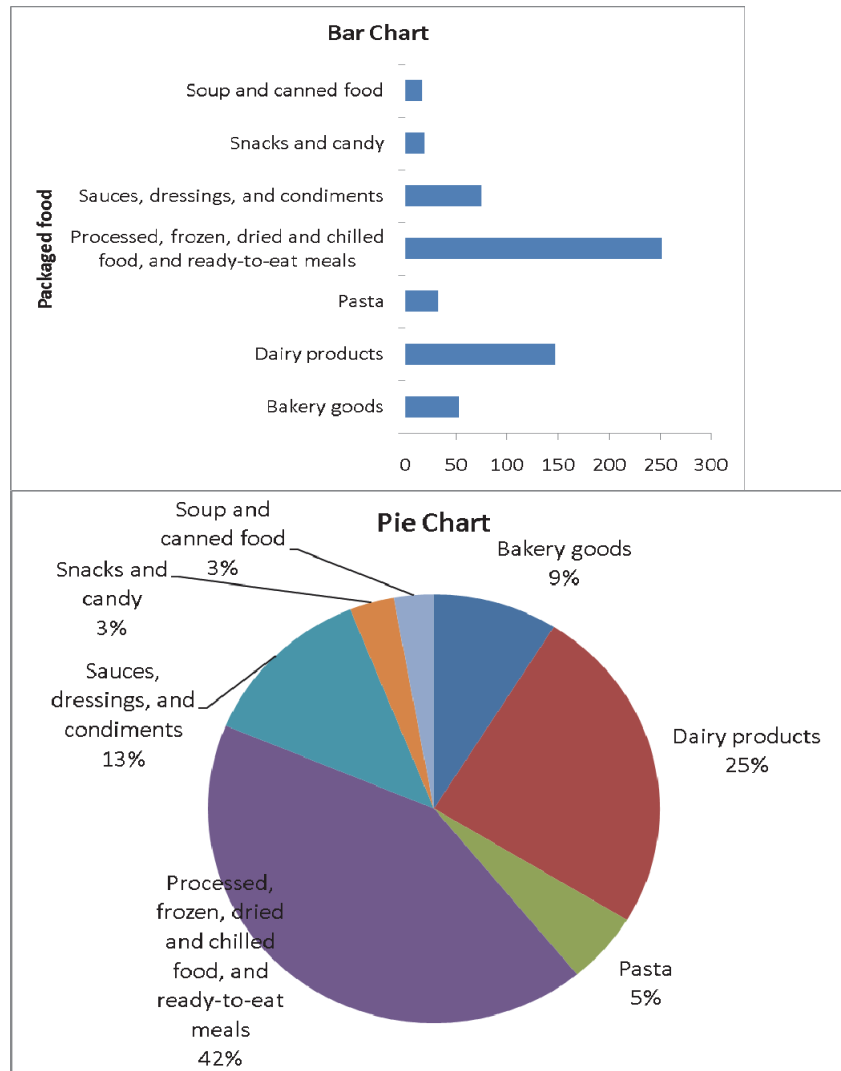


2.93 (b)
cont.

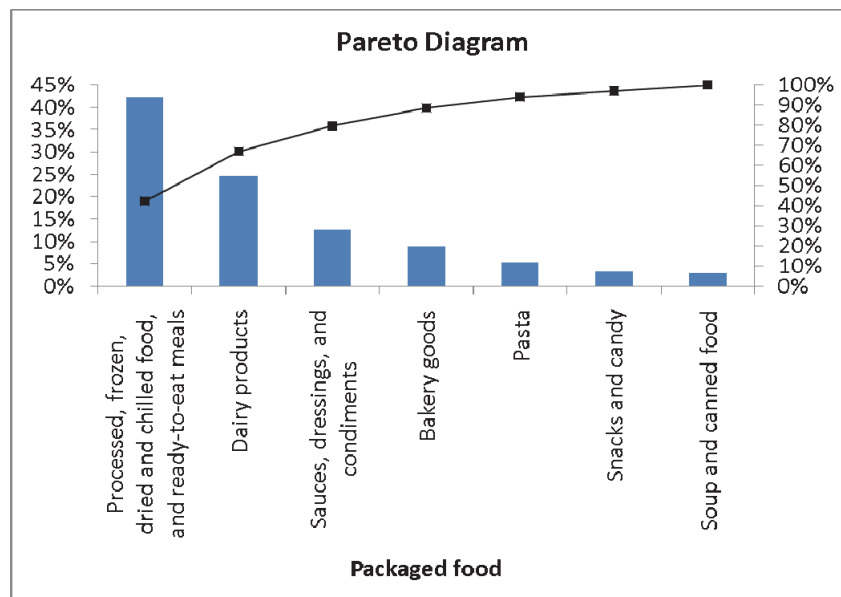


2.93
cont.

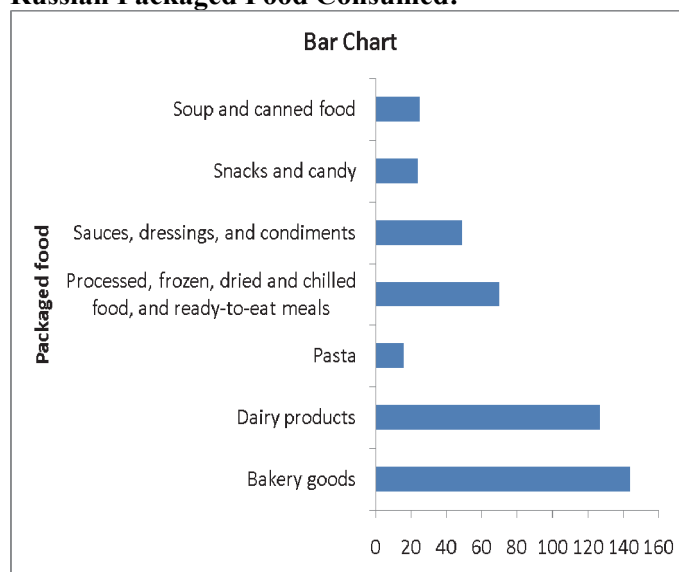
(b) **Japan Packaged Food Consumed:**



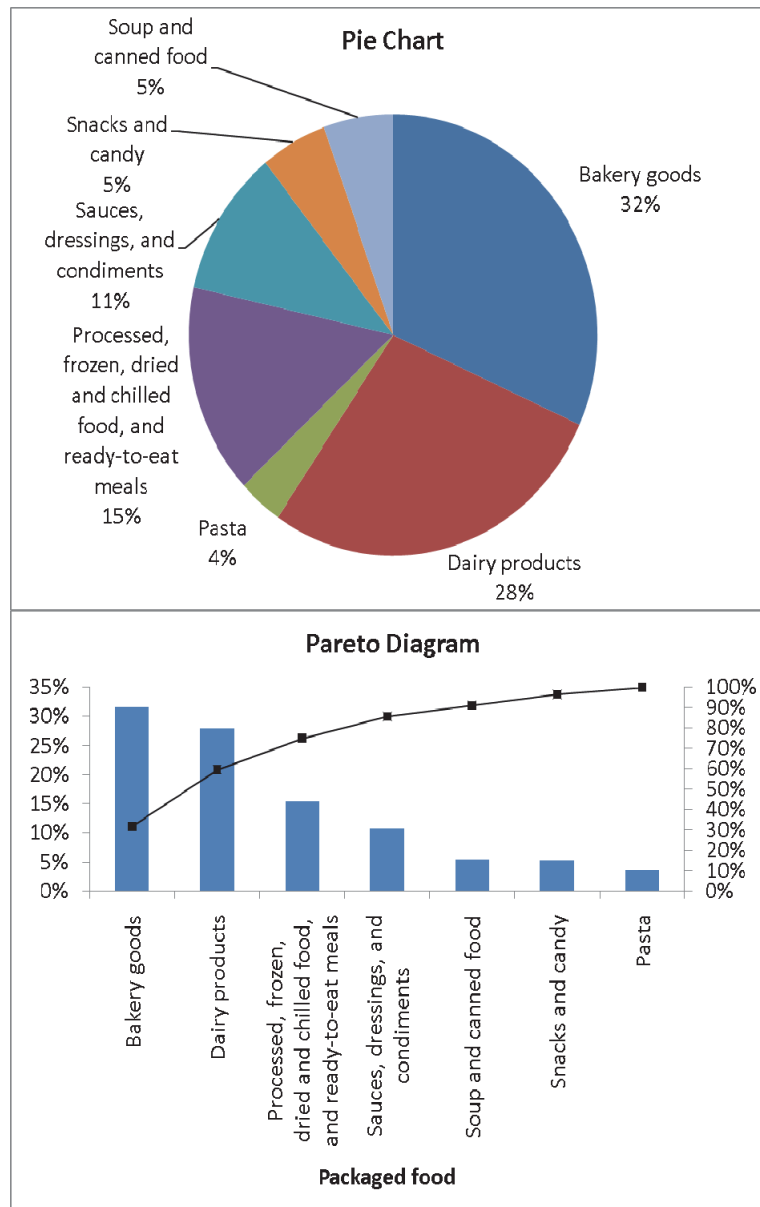
2.93 (b)
cont.



Russian Packaged Food Consumed:



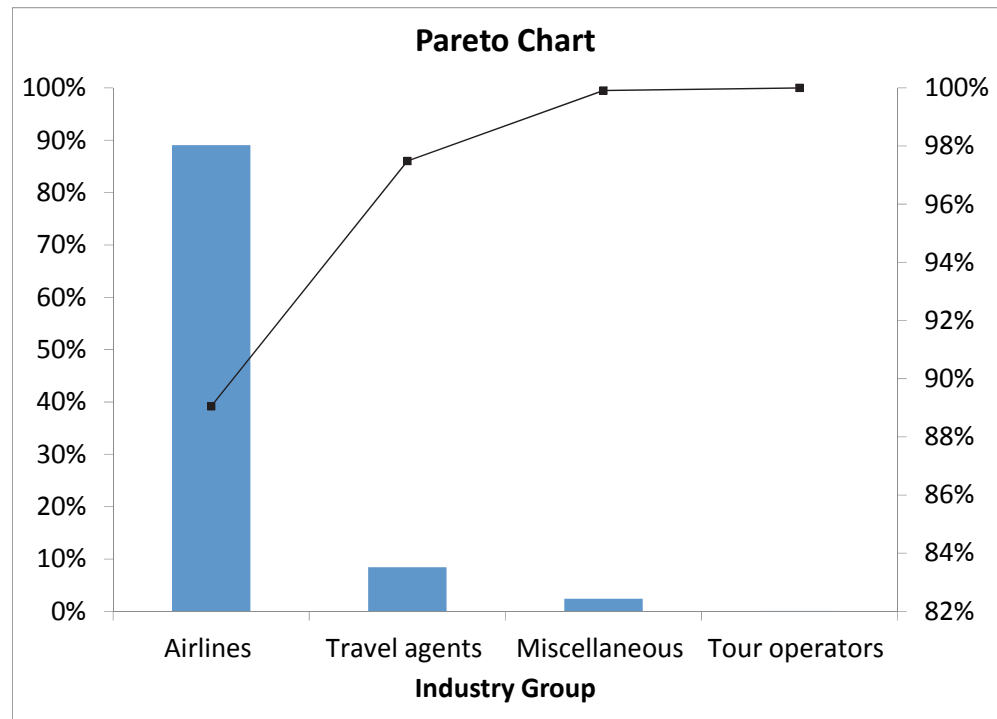
2.93 (b)
cont.



- (c) The fresh food consumption patterns between Japanese and Russians are quite similar with vegetables taking up the largest share followed by meats and seafood while Americans consume about the same amount of meats and seafood, and vegetables. Among the three countries, vegetables, and meats and seafood constitute more than 60% of the fresh food consumption.

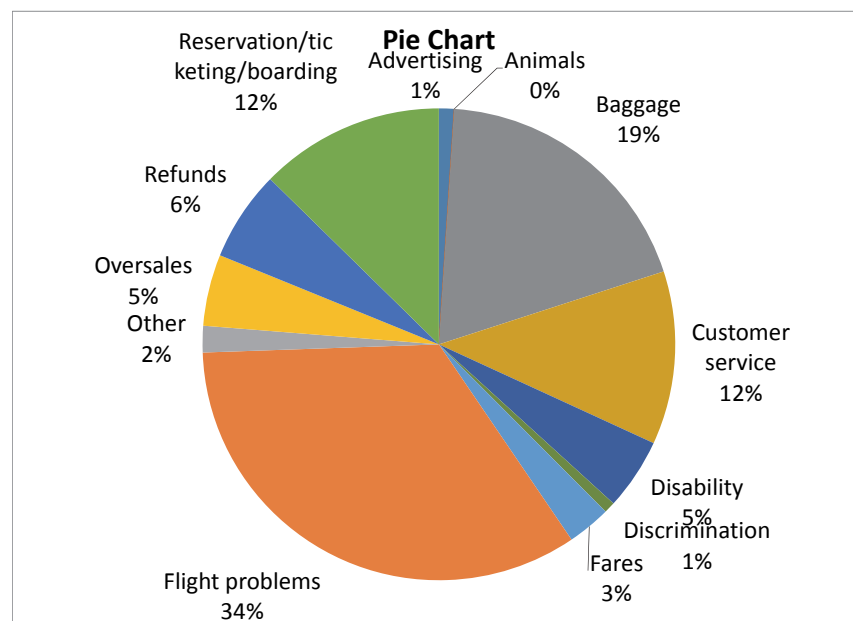
For Americans, dairy products, and processed, frozen, dried and chilled food and ready-to-eat meals make up slightly more than 60% of the packaged food consumption. For Japanese, processed, frozen, dried and chilled food, and ready-to-eat meals, and dairy products constitute more than 60% of their packaged food consumption. For the Russians, bakery goods and dairy products take up 60% of the share of their package food consumption.

2.94 (a)

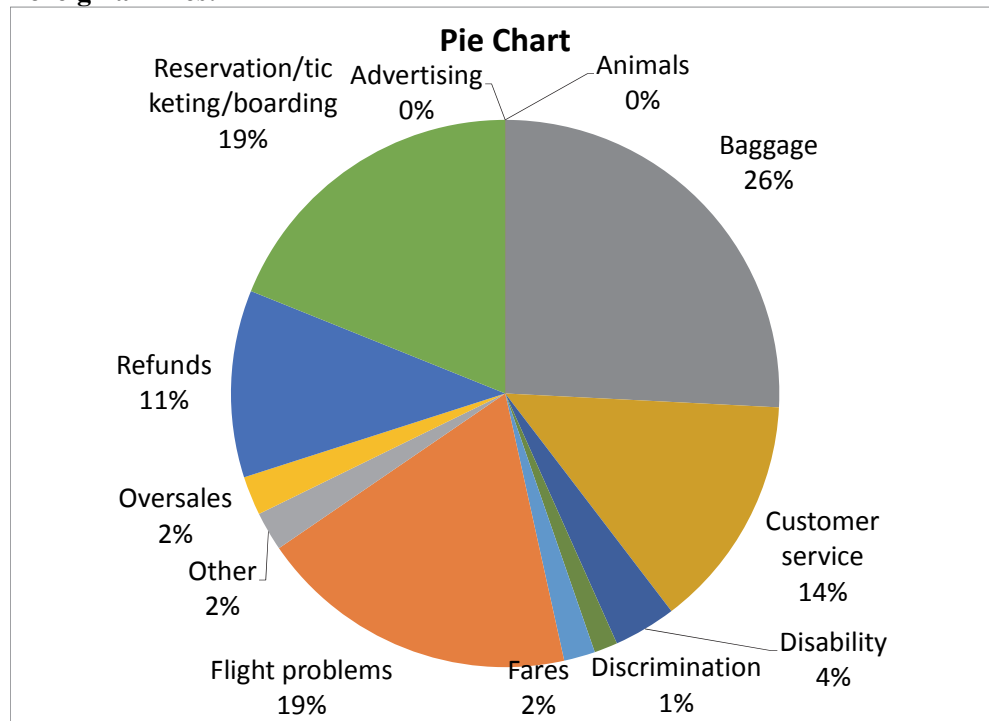


The airline industry accounts for most of the complaints.

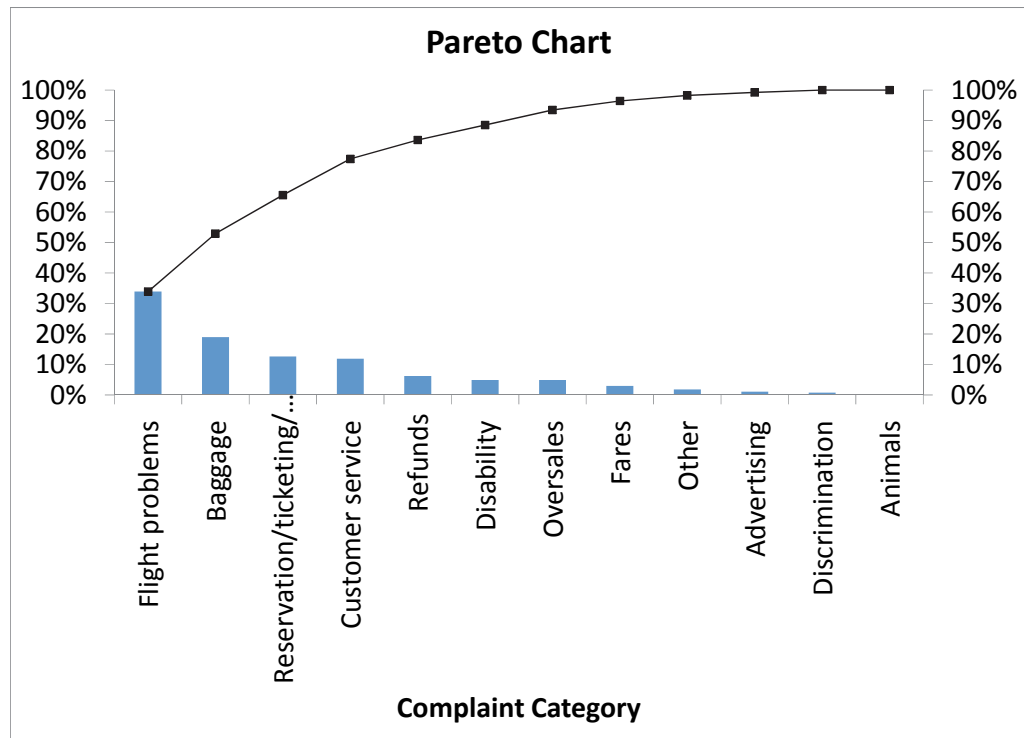
(b) **U.S. airlines:**



2.94 (b) Foreign airlines:

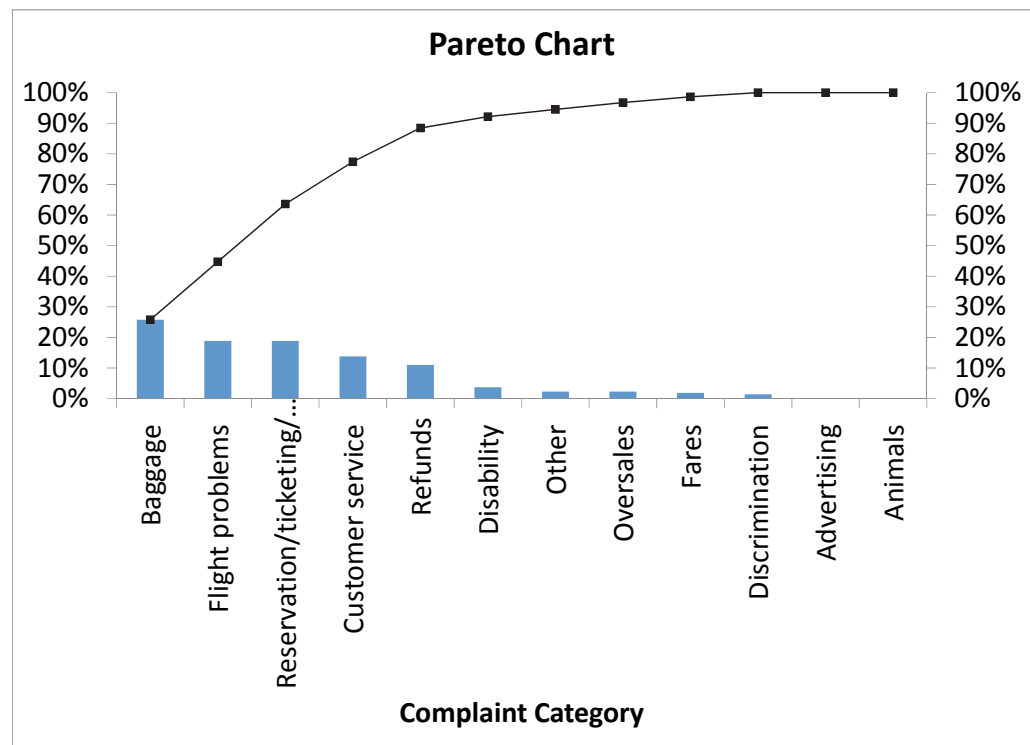


(c)



Flight problems account for most of the complaints against U.S. airlines.

2.94 (d)
cont.

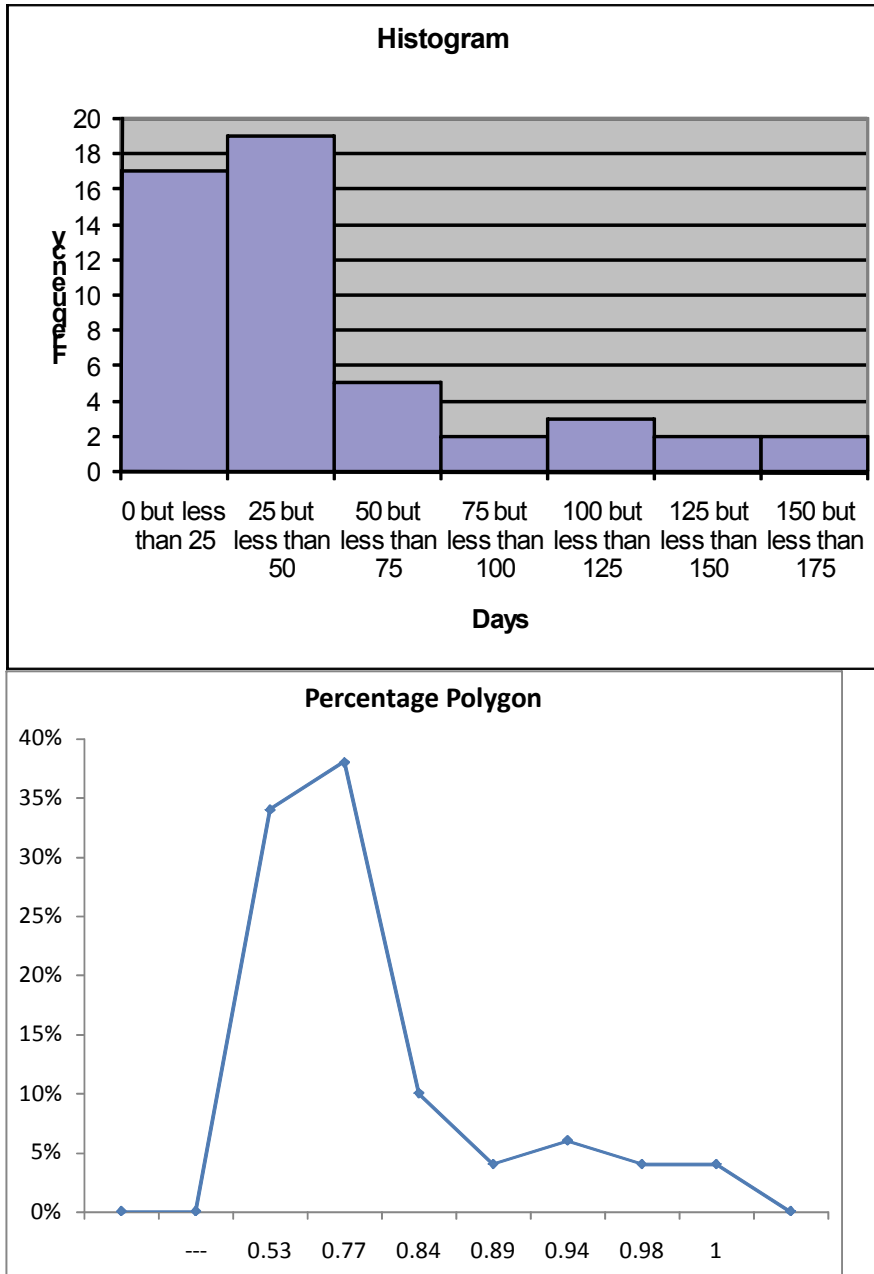


Baggage accounts for most of the complaints against foreign airlines.

2.95 (a)

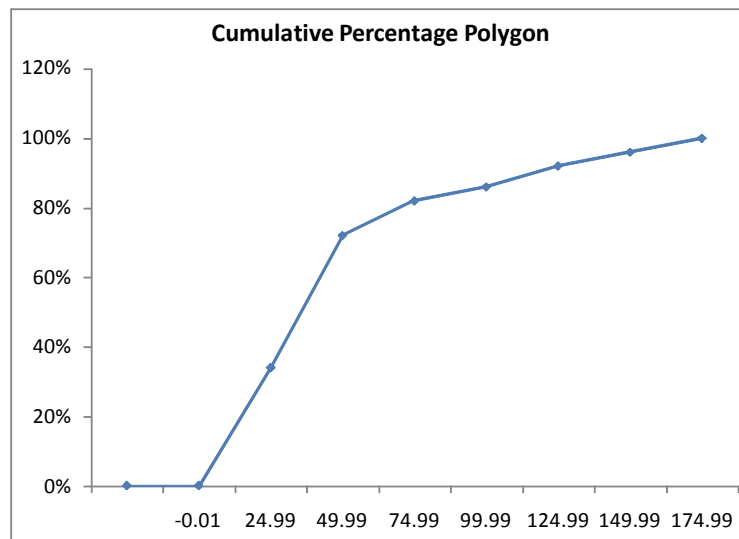
<i>Range</i>	<i>Frequency Percentage</i>	
0 but less than 25	17	34%
25 but less than 50	19	38%
50 but less than 75	5	10%
75 but less than 100	2	4%
100 but less than 125	3	6%
125 but less than 150	2	4%
150 but less than 175	2	4%

2.95 (b)
cont.



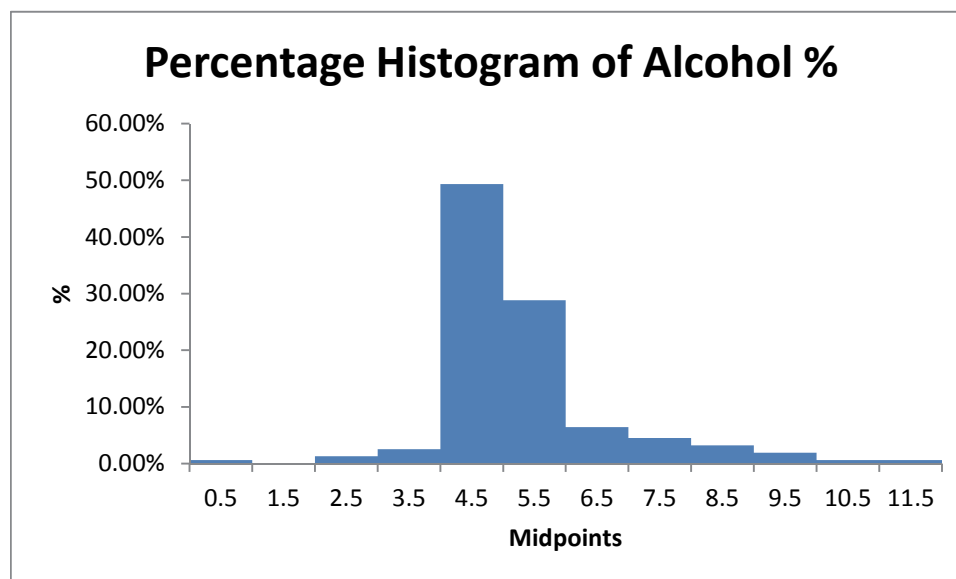
2.95 (c)
cont.

<i>Range</i>	<i>Cumulative %</i>
0 but less than 25	34%
25 but less than 50	72%
50 but less than 75	82%
75 but less than 100	86%
100 but less than 125	92%
125 but less than 150	96%
150 but less than 175	100%

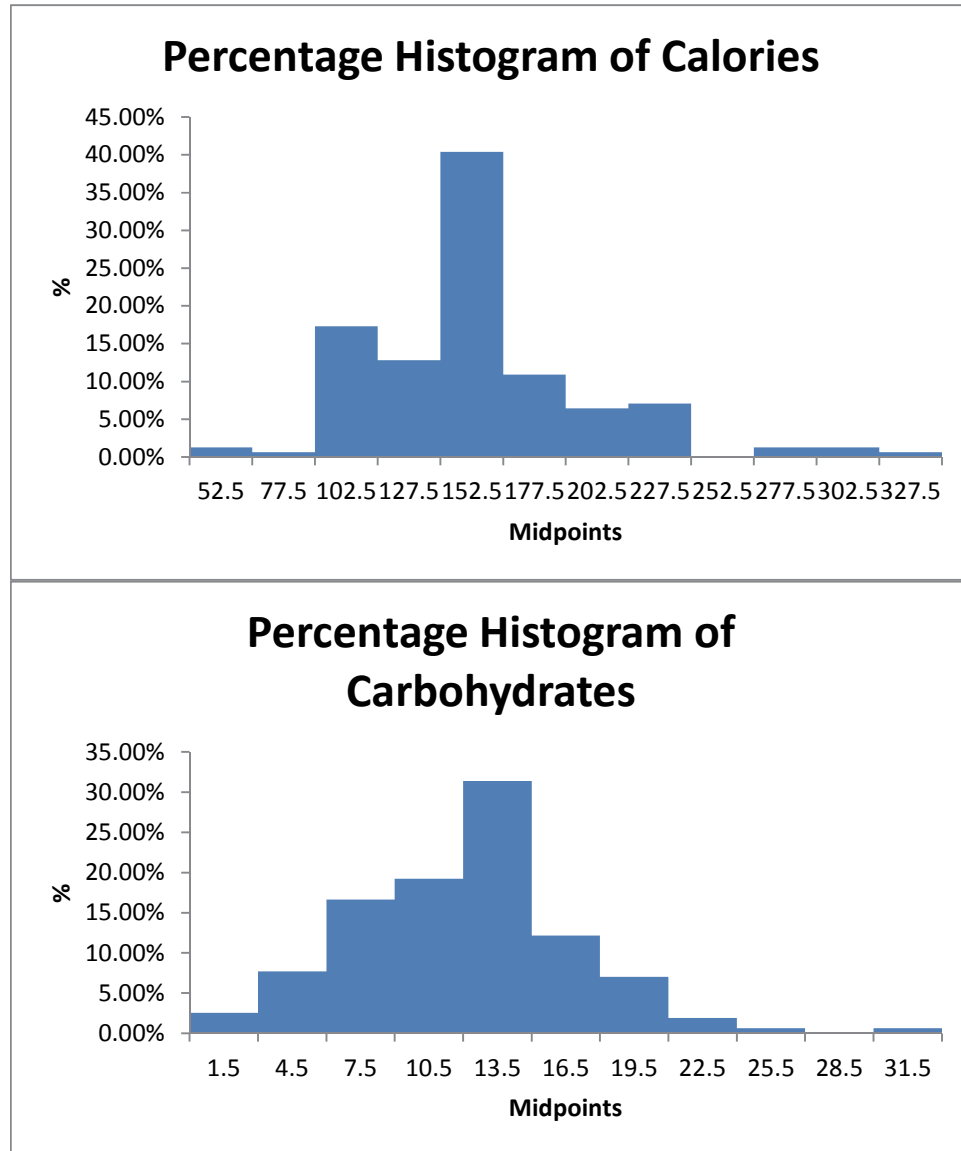


- (d) You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

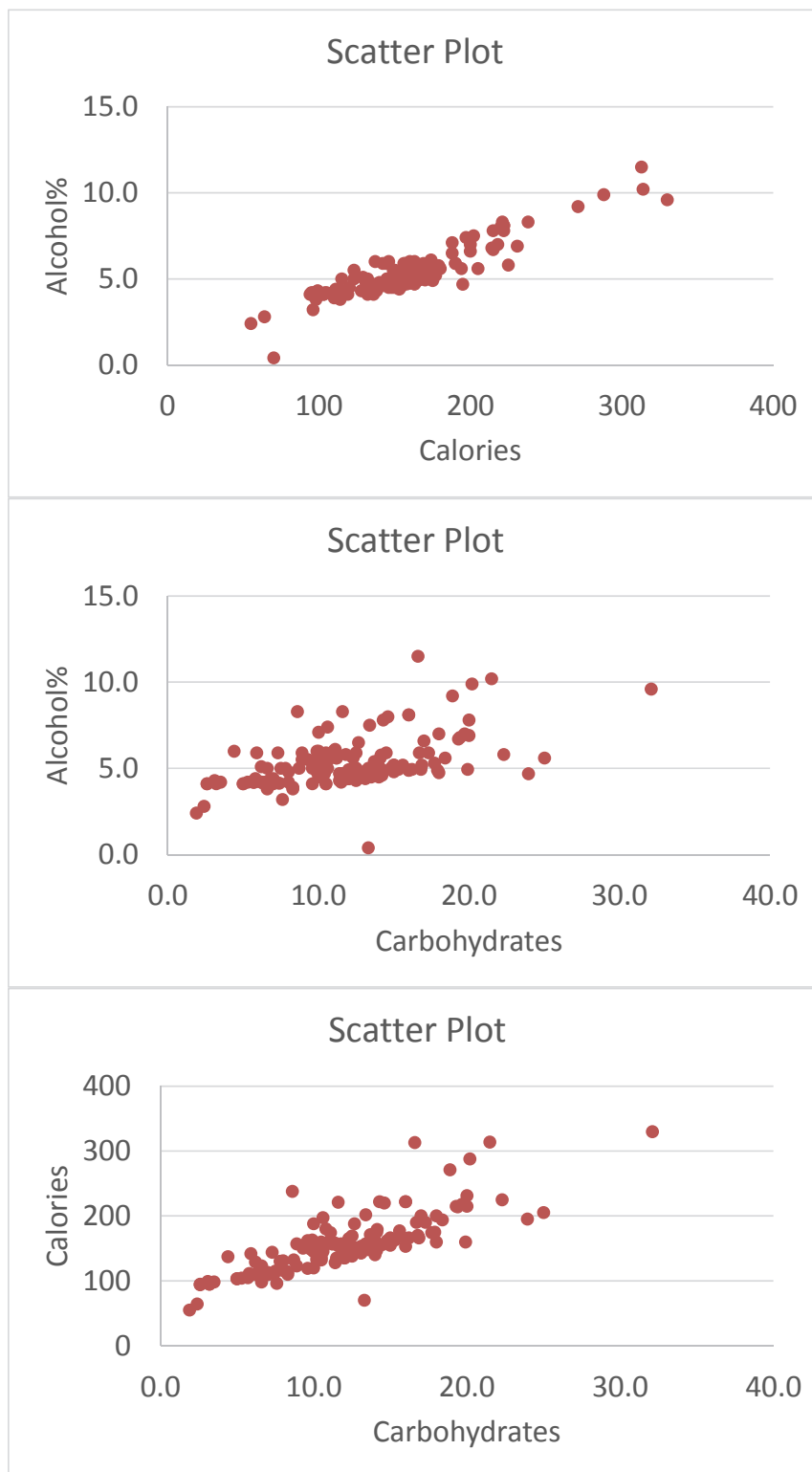
2.96 (a)



2.96 (a)
cont.



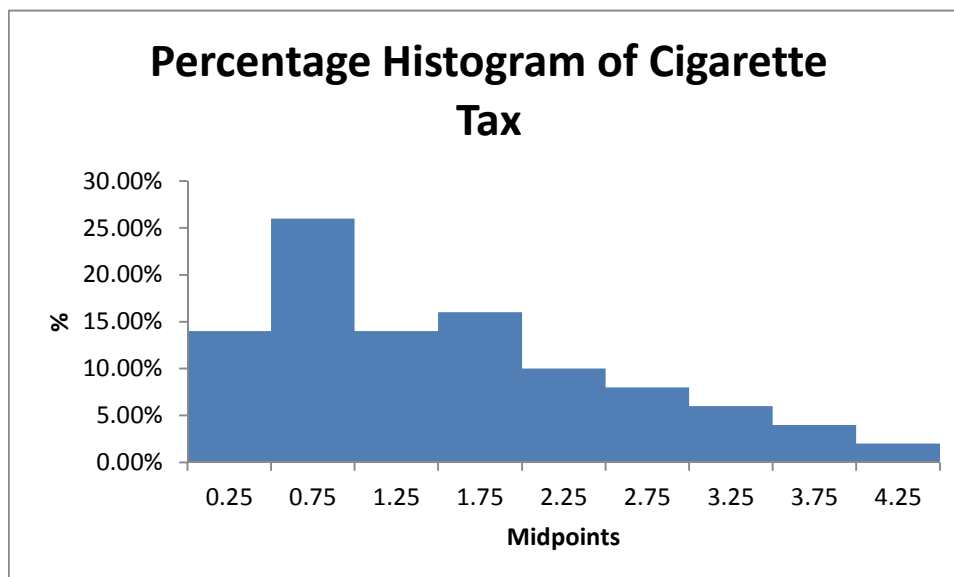
2.96 (b)
cont.



- 2.96 (c) The alcohol % is concentrated between 4 and 6, with more between 4 and 5. The calories are concentrated between 140 and 160. The carbohydrates are concentrated between 12 and 15. There are outliers in the percentage of alcohol in both tails. The outlier in the lower tail is due to the non-alcoholic beer O'Doul's with only a 0.4% alcohol content. There are a few beers with alcohol content as high as around 11.5%. There are a few beers with calories content as high as around 330 and carbohydrates as high as around 32.1.
- There is a strong positive relationship between percentage alcohol and calories, and calories and carbohydrates and a moderately positive relationship between percentage alcohol and carbohydrates.

- 2.97 (a) Ordered array:
- | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.170 | 0.300 | 0.360 | 0.370 | 0.425 | 0.440 | 0.450 | 0.550 | 0.570 | |
| 0.570 | 0.600 | 0.600 | 0.620 | 0.640 | 0.680 | 0.790 | 0.800 | 0.840 | 0.870 |
| 0.995 | 1.030 | 1.150 | 1.250 | 1.310 | 1.339 | 1.360 | 1.410 | 1.530 | 1.600 |
| 1.600 | 1.660 | 1.700 | 1.700 | 1.780 | 1.980 | 2.000 | 2.000 | 2.000 | 2.000 |
| 2.000 | 2.520 | 2.620 | 2.700 | 2.830 | 3.025 | 3.200 | 3.400 | 3.500 | 3.510 |
| 4.350 | | | | | | | | | |

(b)



- (c) There is a \$4.18 difference in the state cigarette tax between the lowest and highest. The distribution of the cigarette tax is somewhat right-skewed with one state having a cigarette tax higher than \$4.0. Majority of the states though have cigarette tax concentrated around \$0.75.

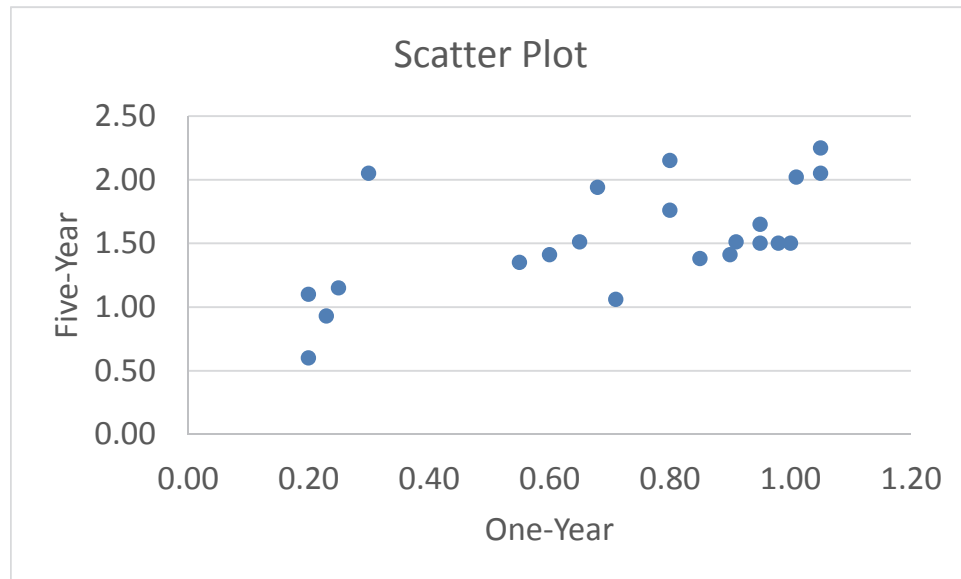
2.98 (a) One-year CD:

		Stem-and-Leaf Display	
		Stem unit: 0.1	
Statistics		2	0 0 3 5
Sample Size	22	3	0
Mean	0.7100	4	
Median	0.8000	5	5
Std. Deviation	0.2990	6	0 5 8
Minimum	0.2000	7	1
Maximum	1.0500	8	0 0 5
		9	0 1 5 5 8
		10	0 1 5 5

5-year CD

		Stem-and-Leaf Display	
		Stem unit: 0.1	
Statistics		6	0
Sample Size	22	7	
Mean	1.5355	8	
Median	1.5000	9	3
Std. Deviation	0.4243	10	6
Minimum	0.6000	11	0 5
Maximum	2.2500	12	
		13	5 8
		14	1 1
		15	0 0 0 1 1
		16	5
		17	6
		18	
		19	4
		20	2 5 5
		21	5
		22	5

2.98 (b)
cont.

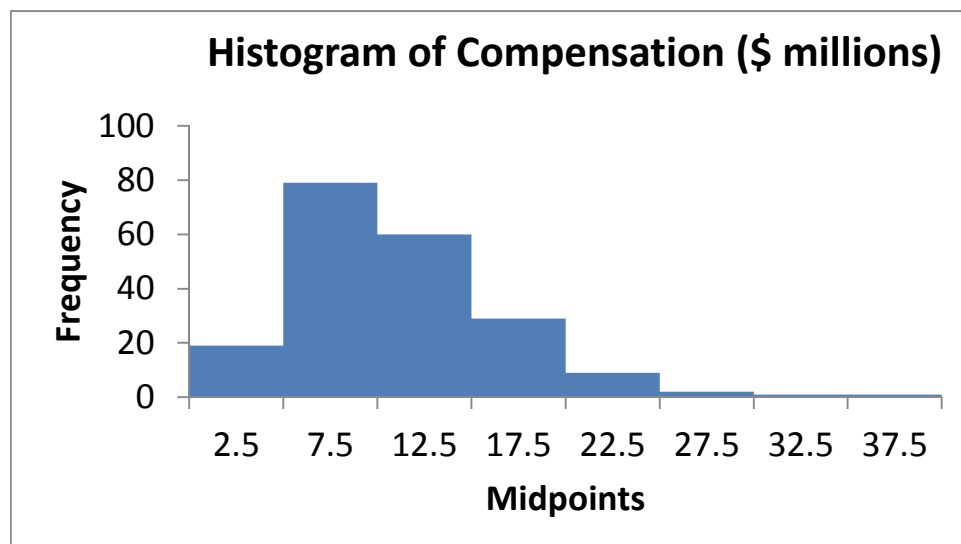


(c) There appears to be a positive relationship between the yield of the one-year CD and the five-year CD.

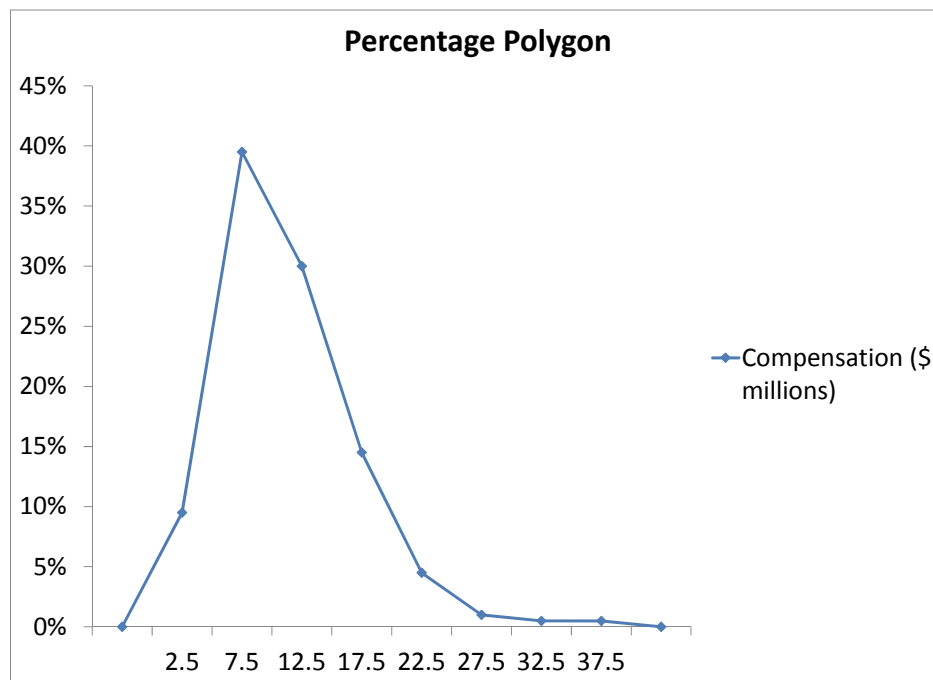
2.99 (a),(c)

bin	Frequency	Percentage
0 but less than 5	19	9.50%
5 but less than 10	79	39.50%
10 but less than 15	60	30.00%
15 but less than 20	29	14.50%
20 but less than 25	9	4.50%
25 but less than 30	2	1.00%
30 but less than 35	1	0.50%
35 but less than 40	1	0.50%

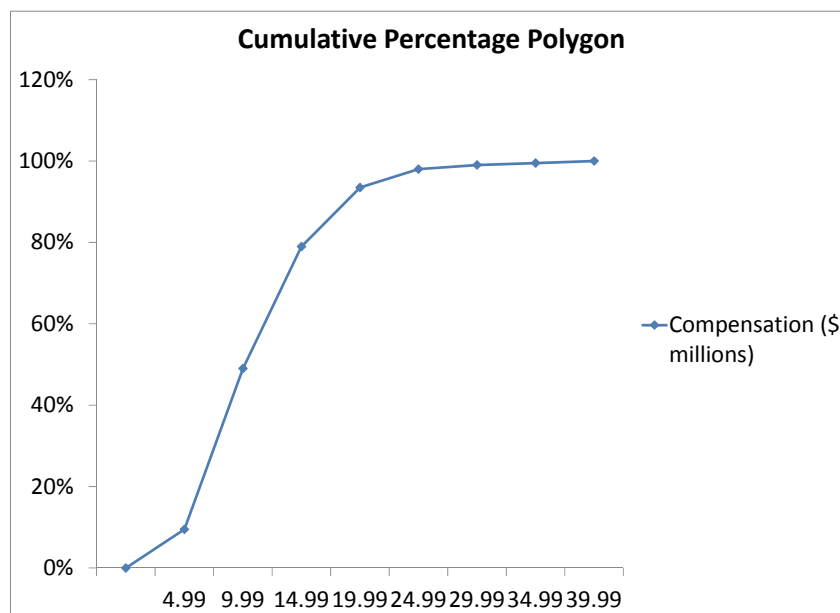
(b)



2.99 (b)
cont.

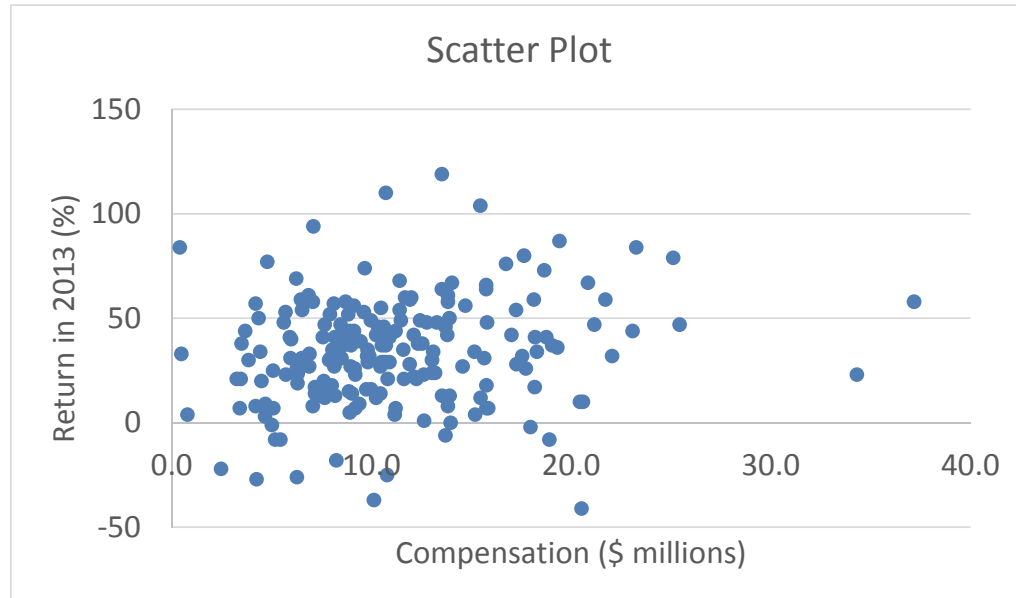


(c)



(d) CEO compensation in 2013 is right skewed. Slightly higher than 80% of the CEOs have compensation lower than \$15,000,000

2.99 (e)
cont.



- (f) There is not any obvious relationship between the total compensation and investment return in 2013.

2.100 (a)

Frequencies (Boston)

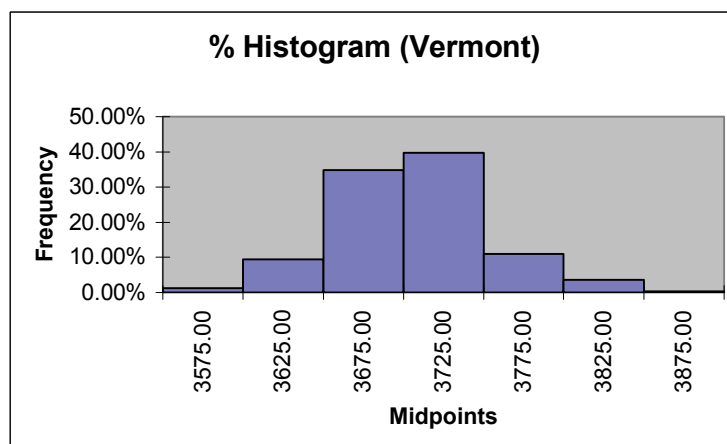
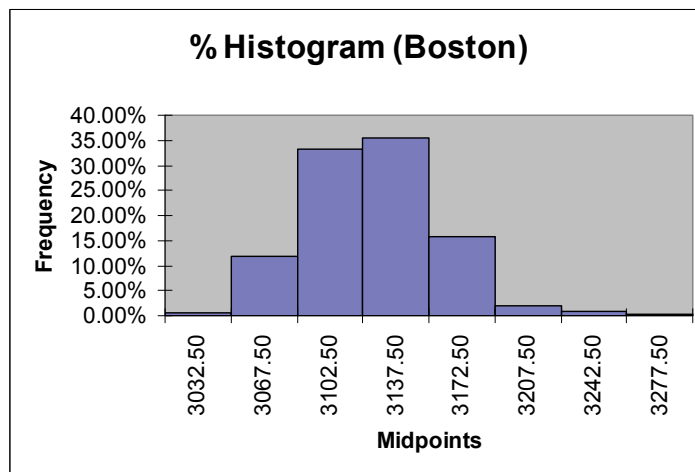
<i>Weight (Boston)</i>	<i>Frequency</i>	<i>Percentage</i>
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1	0.27%

(b)

Frequencies (Vermont)

<i>Weight (Vermont)</i>	<i>Frequency</i>	<i>Percentage</i>
3550 but less than 3600	4	1.21%
3600 but less than 3650	31	9.39%
3650 but less than 3700	115	34.85%
3700 but less than 3750	131	39.70%
3750 but less than 3800	36	10.91%
3800 but less than 3850	12	3.64%
3850 but less than 3900	1	0.30%

2.100 (c)
cont.



- (d) 0.54% of the “Boston” shingles pallets are underweight while 0.27% are overweight.
1.21% of the “Vermont” shingles pallets are underweight while 3.94% are overweight.

2.101 (a),(c) **Two-star:**

bin	Frequency	Percentage	Cumulative Pctage.	Midpts.
20 but less than 30	2	0.0526316	5.26%	25
30 but less than 40	2	0.0526316	10.53%	35
40 but less than 50	8	0.2105263	31.58%	45
50 but less than 60	2	0.0526316	36.84%	55
60 but less than 70	5	0.1315789	50.00%	65
70 but less than 80	10	0.2631579	76.32%	75
80 but less than 90	5	0.1315789	89.47%	85
90 but less than 100	2	0.0526316	94.74%	95
100 but less than 110	1	0.0263158	97.37%	105
110 but less than 120	1	0.0263158	100.00%	115

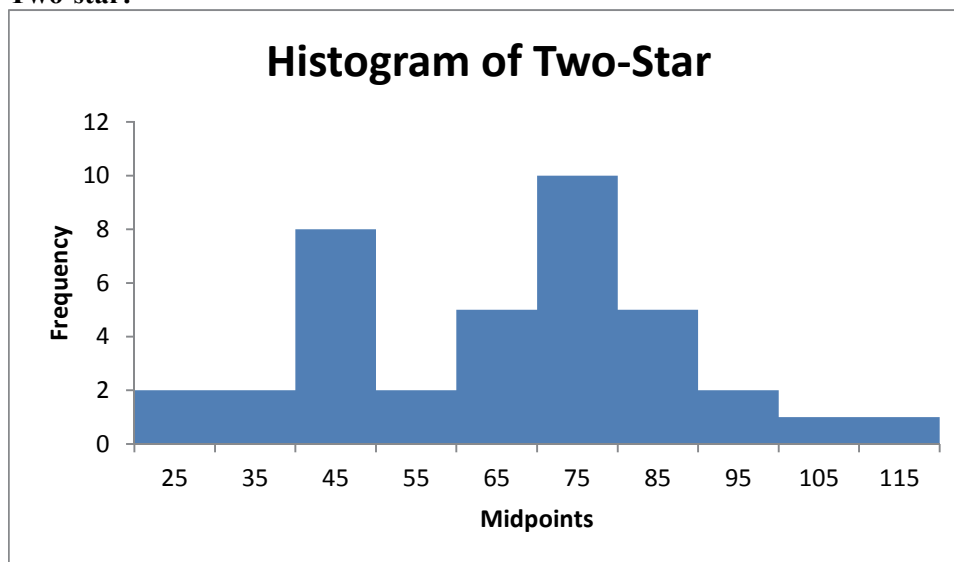
2.101 (a),(c) **Three-star:**
cont.

bin	Frequency	Percentage	Cumulative Pctage.	Midpts.
30 but less than 45	1	0.0263158	2.63%	37.5
45 but less than 60	5	0.1315789	15.79%	52.5
60 but less than 75	4	0.1052632	26.32%	67.5
75 but less than 90	11	0.2894737	55.26%	82.5
90 but less than 105	7	0.1842105	73.68%	97.5
105 but less than 120	4	0.1052632	84.21%	112.5
120 but less than 135	4	0.1052632	94.74%	127.5
135 but less than 150	1	0.0263158	97.37%	142.5
150 but less than 165	1	0.0263158	100.00%	157.5

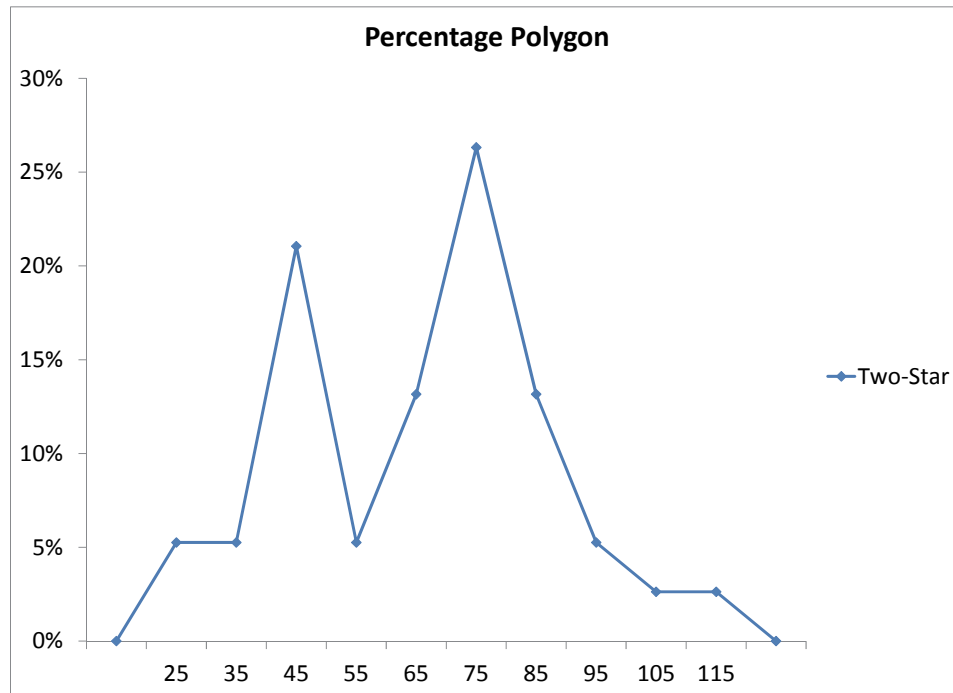
Four-star:

bin	Frequency	Percentage	Cumulative Pctage.	Midpts.
65 but less than 80	6	0.1578947	15.79%	72.5
80 but less than 95	4	0.1052632	26.32%	87.5
95 but less than 110	5	0.1315789	39.47%	102.5
110 but less than 125	4	0.1052632	50.00%	117.5
125 but less than 140	4	0.1052632	60.53%	132.5
140 but less than 155	7	0.1842105	78.95%	147.5
155 but less than 170	2	0.0526316	84.21%	162.5
170 but less than 185	2	0.0526316	89.47%	177.5
185 but less than 200	0	0	89.47%	192.5
200 but less than 215	4	0.1052632	100.00%	207.5

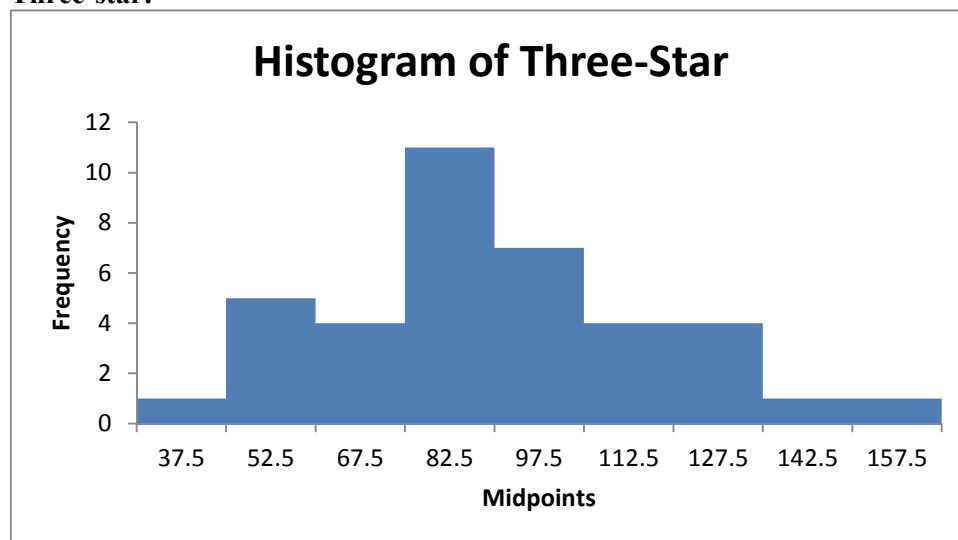
(b) **Two-star:**



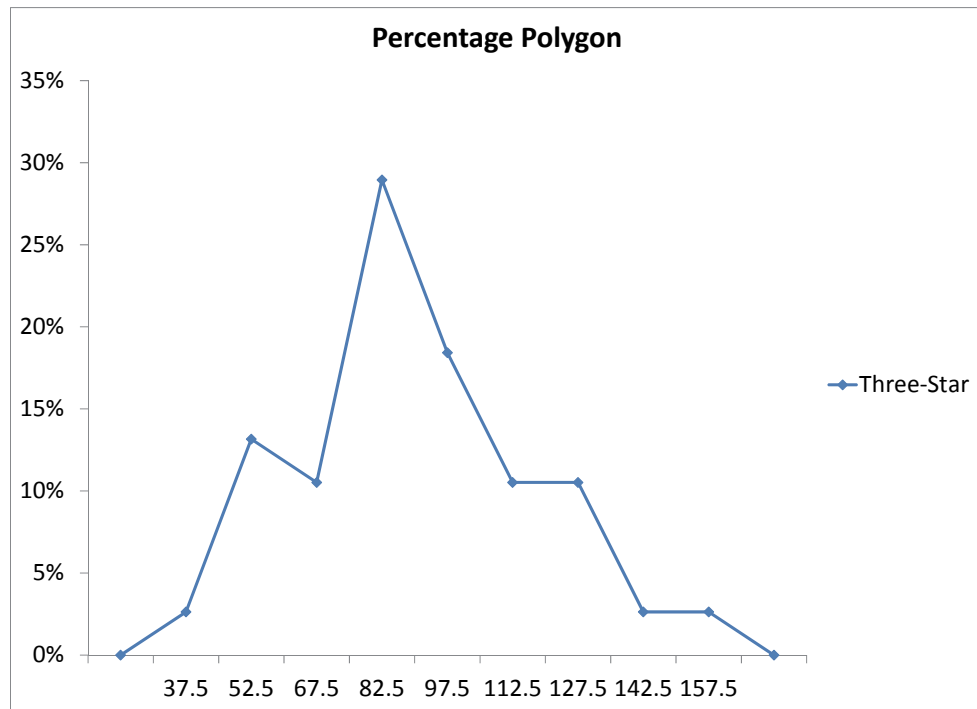
2.101 (b)
cont.



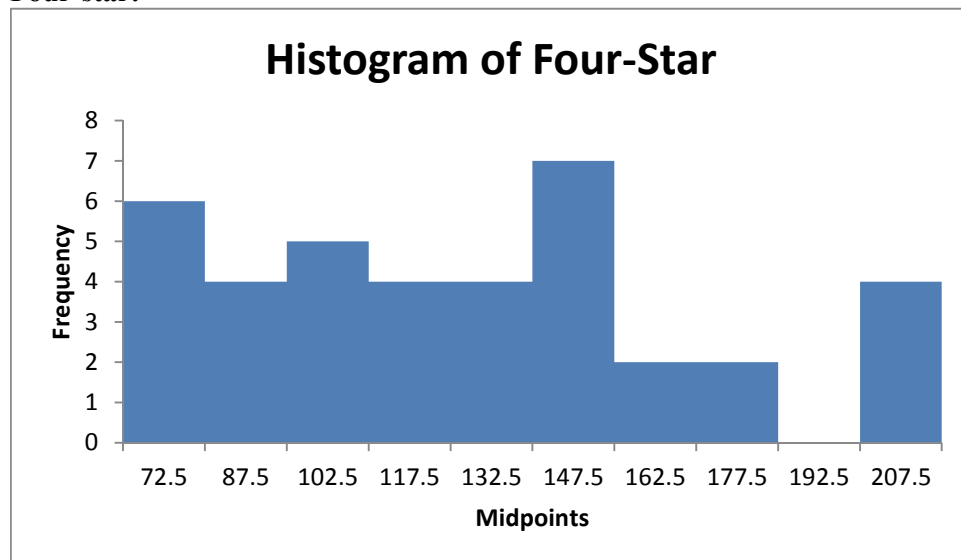
Three-star:



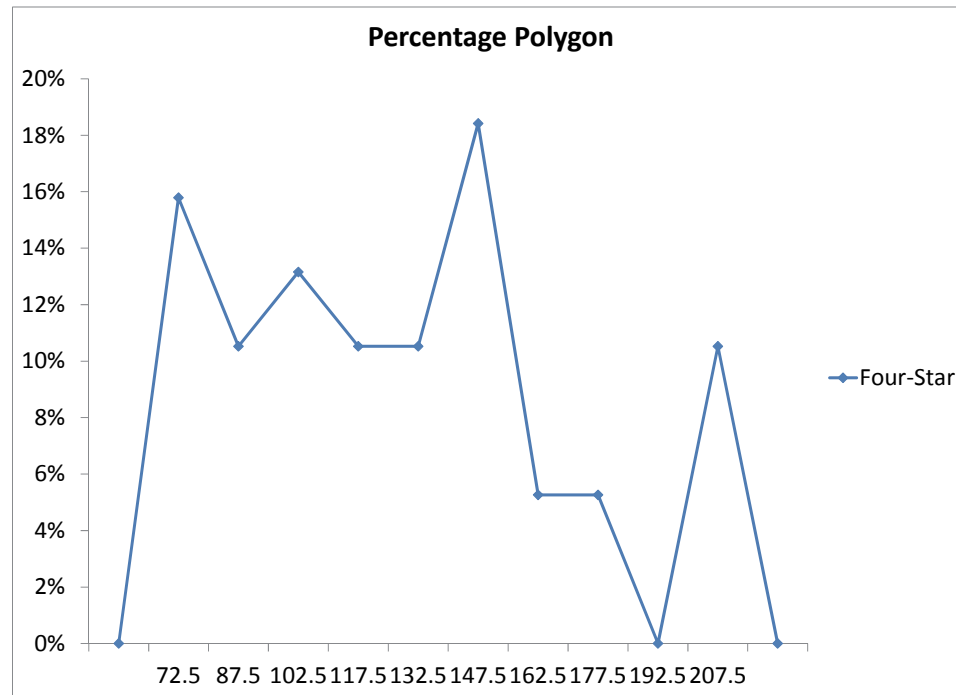
2.101 (b)
cont.



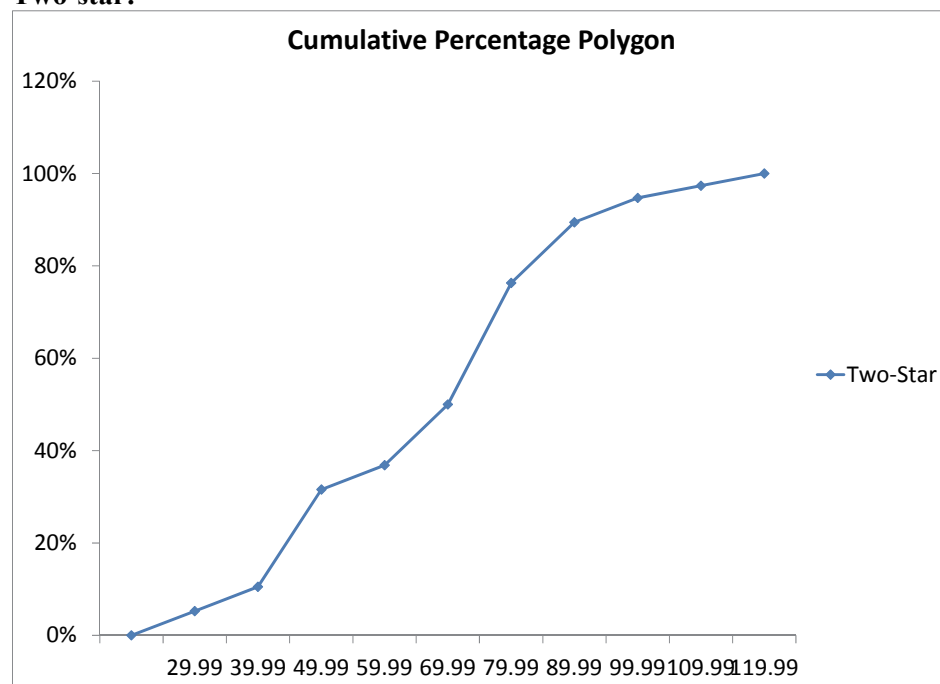
Four-star:



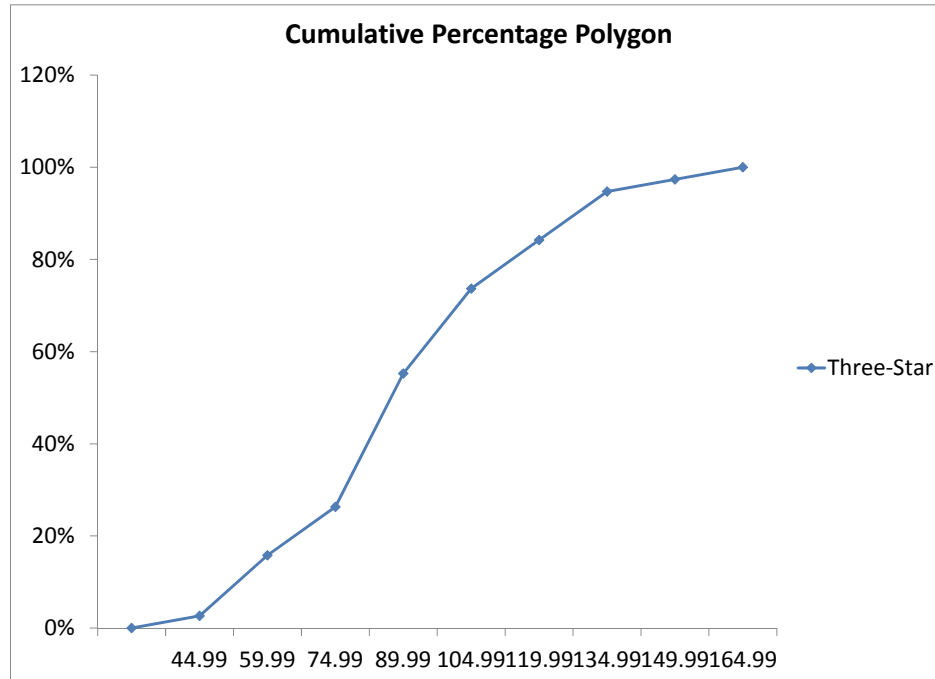
2.101 (b)
cont.



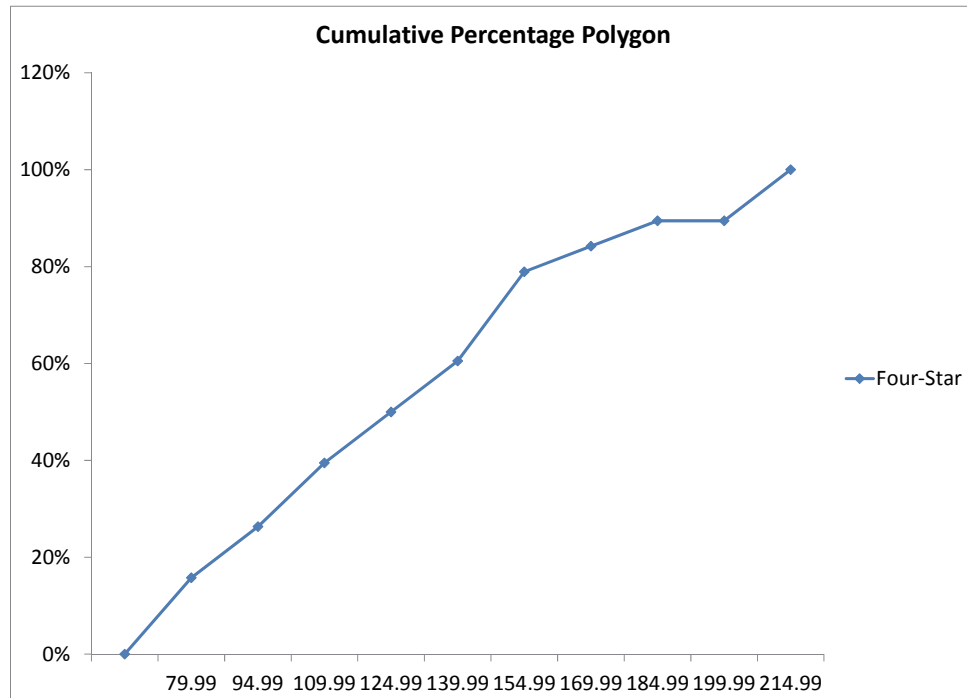
(c) **Two-star:**



2.101 (c) **Three-star:**
cont.

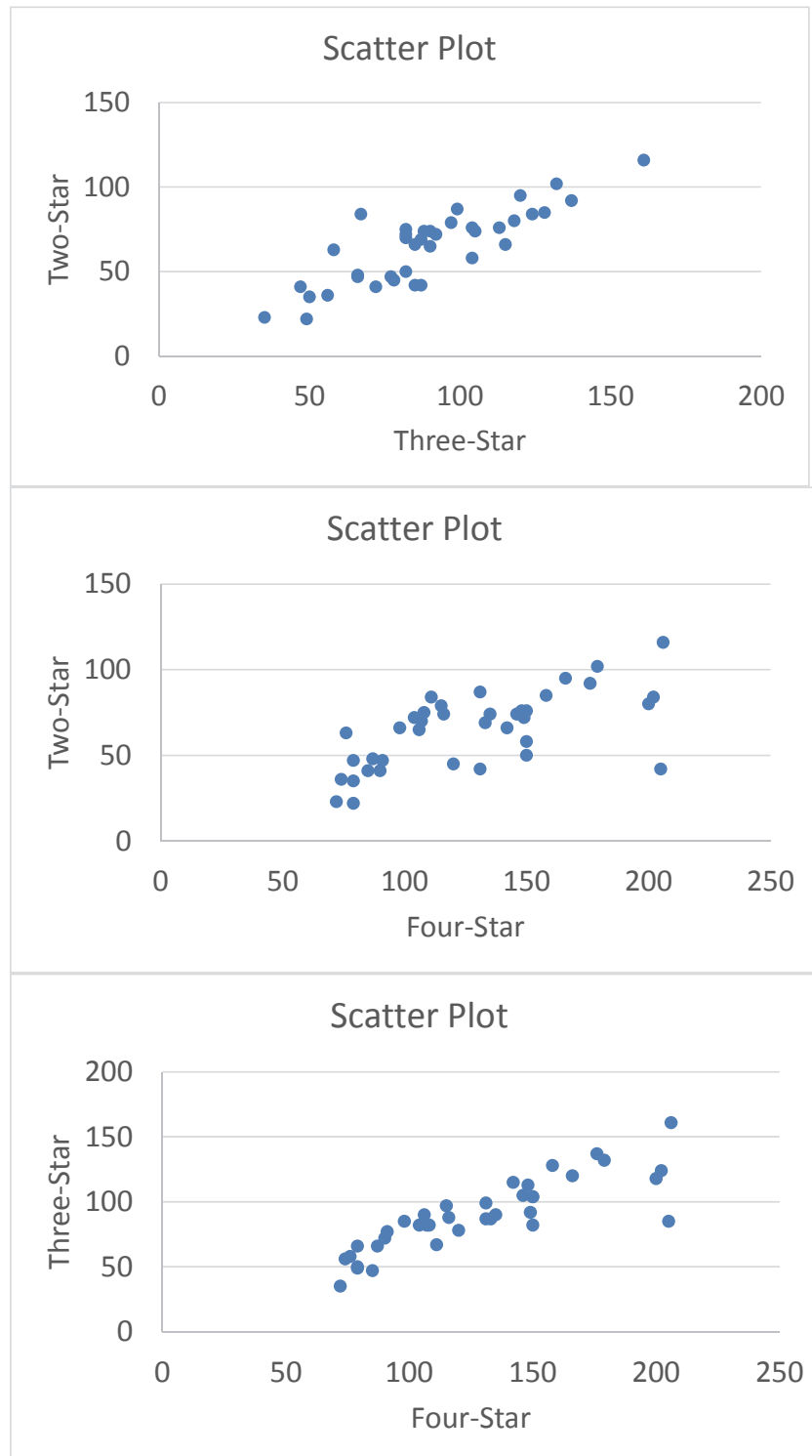


Four-star:



- (d) The price of two-star, three-star and four-star hotels are all right-skewed. The median price of two-star, three-star and four-star hotels is around 70, 82.5, and 125 English pounds, respectively.

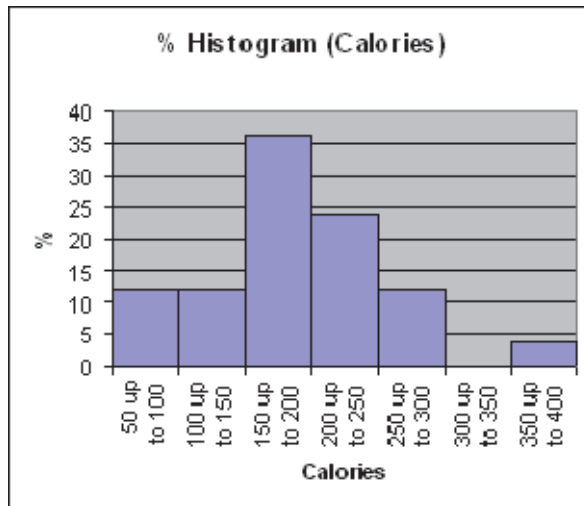
2.101 (e)
cont.



(f) The relationship of the price between two-star and three-star, three-star and four-star, and two-star and four-star hotels are all positive.

2.102 (a)

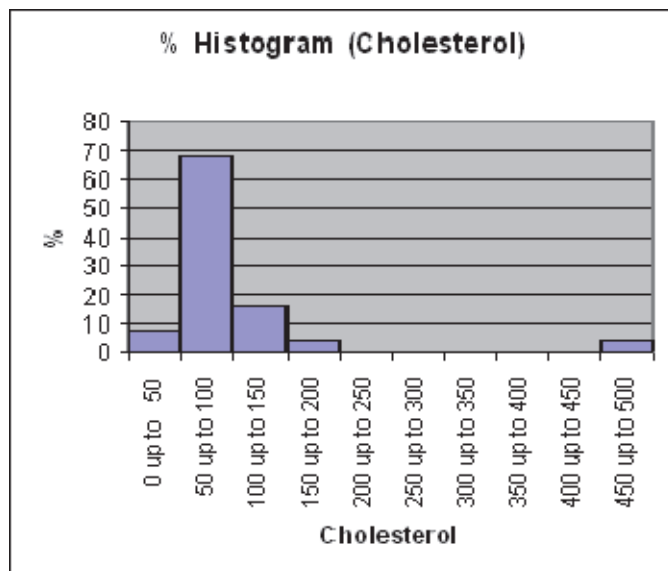
Calories	Frequency	Percentage	Percentage Less Than
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100



(b)

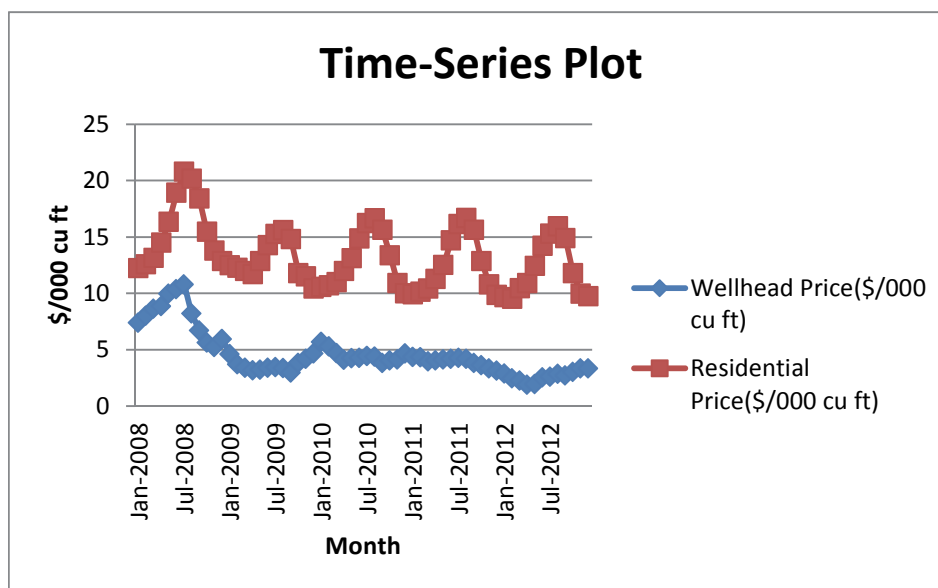
Cholesterol	Frequency	Percentage	Percentage Less Than
0 up to 50	2	8	8%
50 up to 100	17	68	76
100 up to 150	4	16	92
150 up to 200	1	4	96
200 up to 250	0	0	96
250 up to 300	0	0	96
300 up to 350	0	0	96
350 up to 400	0	0	96
400 up to 450	0	0	96
450 up to 500	1	4	100

2.102 (b)
cont.



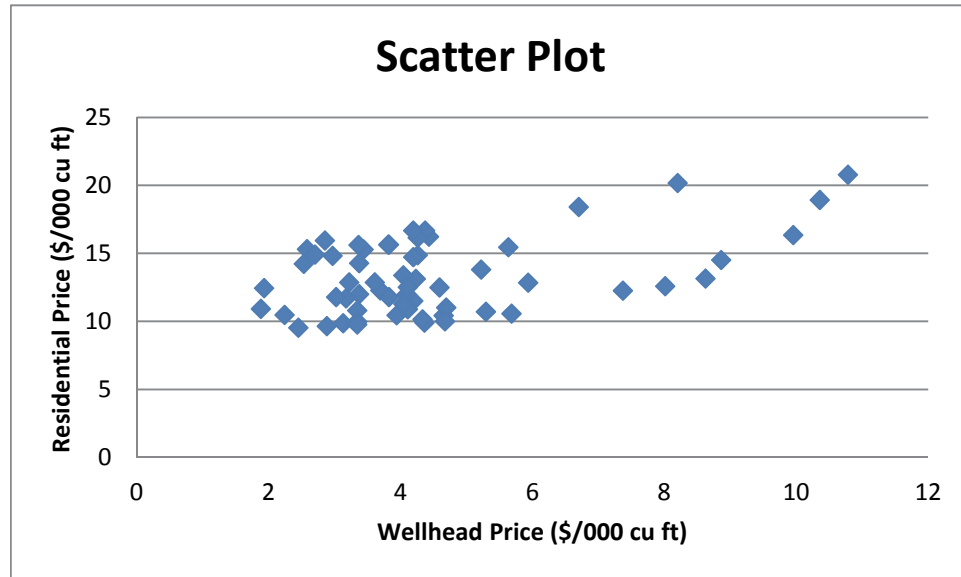
- (c) The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled foods varies from 16 to 33 grams, with 68% of the data values falling between 24 and 32 grams. Spareribs and fried liver are both very different from other foods sampled—the former on calories and the latter on cholesterol content.

2.103 (a)



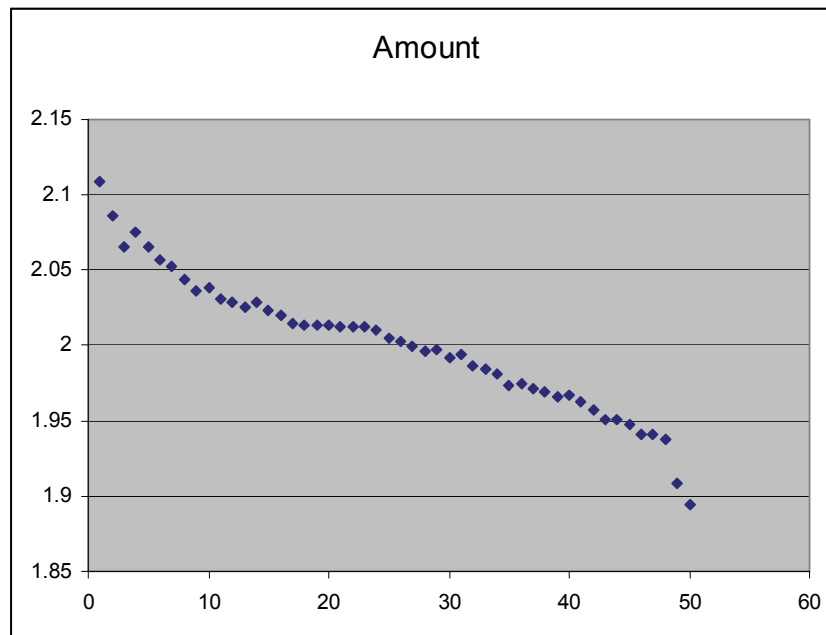
- (b) The wellhead average price was highest in the summer of 2008 and had since declined. The residential average price of gasoline in the United States is higher in the summer in general and seems to peak in June.

2.103 (c)
cont.



- (d) There appears to be a slight positive relationship between the wellhead price and residential price.

2.104 (a)



- (b) There is a downward trend in the amount filled.
 (c) The amount filled in the next bottle will most likely be below 1.894 liter.
 (d) The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution of the data.