

CHAPTER 1

- 1.1 (a) The type of beverage sold yields categorical or “qualitative” responses.
 (b) 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 (a) 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.
 (b) The download time is a ratio scaled variable because the true zero point in the measurement is zero units of time.
- 1.4 (a) The number of cellphones is a numerical variable that is discrete because the outcome is a count. It is ratio scaled because it has a true zero point.
 (b) Monthly data usage is a numerical variable that is continuous because any value within a range of values can occur. It is ratio scaled because it has a true zero point.
 (c) Number of text messages exchanged per month is a numerical variable that is discrete because the outcome is a count. It is ratio scaled because it has a true zero point.
 (d) Voice usage per month is a numerical variable that is continuous because any value within a range of values can occur. It is ratio scaled because it has a true zero point.
 (e) Whether a cellphone is used for email is a categorical variable because the answer can be only yes or no. This also makes it a nominal-scaled variable.
- 1.5 (a) numerical, continuous, ratio scale
 (b) numerical, discrete, ratio scale
 (c) categorical, nominal scale
 (d) categorical, nominal scale
- 1.6 (a) Categorical, nominal scale.
 (b) Numerical, continuous, ratio scale.
 (c) Categorical, nominal scale.
 (d) Numerical, discrete, ratio scale.
 (e) Categorical, nominal scale.
- 1.7 (a) numerical, continuous, ratio scale *
 (b) categorical, nominal scale
 (c) categorical, nominal scale
 (d) numerical, discrete, ratio scale

*Some researchers consider money as a discrete numerical variable because it can be “counted.”

- 1.8 (a) numerical, continuous, ratio scale *
 (b) numerical, discrete, ratio scale
 (c) numerical, continuous, ratio scale *
 (d) categorical, nominal

*Some researchers consider money as a discrete numerical variable because it can be “counted.”

40 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.

- 1.19 (b) Row 29: 12 47 83 76 22 99 65 93 10 65 83 61 36 98 89 58 86
 cont. 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.

42 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 four strata will be proportionately represented.
- (b) Since the stratum may differ in the invoice amount, it may be more important to sample a larger percentage of invoices in stratum 1 and stratum 2, and smaller percentages in stratum 3 and stratum 4.
- For example, $\frac{50}{5000} = 1\%$ so 1% of 500 = 5 invoices should be selected from stratum 1; similarly 10% = 50 should be selected from stratum 2, 20% = 100 from stratum 3, and 69% = 345 from stratum 4.
- (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 smart phone and tablet users, 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 Both nominal scaled and ordinal scaled variables are categorical variables but no ranking is implied in nominal scaled variable such as male or female while ranking is implied in ordinal scaled variable such as a student's grade of A, B, C, D and F.
- 1.36 Both interval scaled and ratio scaled variables are numerical variables in which the difference between measurements is meaningful but an interval scaled variable does not involve a true zero such as standardized exam scores while a ratio scaled variable involves a true zero such as height.
- 1.37 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.
- 1.38 Microsoft Excel or Minitab could be used to perform various statistical computations that were possible only with a slide-rule or hand-held calculator in the old days.

44 Chapter 1: Defining and Collecting Data

- 1.39 (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 participated in Adobe System study.
(c) A parameter of interest is the proportion of all tablet users in the population who use their device to purchase products 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 products and services.
- 1.40 The answers to this question depend on which article and its corresponding data set is being selected.
- 1.41 (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.
(c) A parameter of interest is the proportion of supply chain executives in the population who acknowledge that the 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 the supply chain is seen as a strategic asset in their company.
- 1.42 The answers to this question depend on which data set is being selected.
- 1.43 (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?
- 1.44 (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
- 1.45 (a) (i) categorical (ii) categorical
(iii) numerical, discrete (iv) categorical
(b) The answers will vary.
(c) The answers will vary.
- 1.46 Microsoft Excel:
This product features a spreadsheet-based interface that allows users to organize, calculate, and organize data. Excel also contains many statistical functions to assist in the description of a dataset. Excel can be used to develop worksheets and workbooks to calculate a variety of statistics including introductory and advanced statistics. Excel also includes interactive tools to create graphs, charts, and pivot tables. Excel can be used to summarize data to better understand a population of interest, compare across groups, predict outcomes, and to develop forecasting models. These capabilities represent those that are generally relevant to the current course. Excel also includes many other statistical capabilities that can be further explored on the Microsoft Office Excel official website.

1.46 Minitab 18:

cont. Minitab 18 has a comprehensive set of statistical methods including introductory and advanced statistical procedures. Minitab 18 features include basic descriptive statistical procedures, graph and chart creation, diagnostic tests, analysis of variance, regression, time series and forecasting analyses, nonparametric analyses, cross-tabulation, chi-square and related tests, and other statistical procedures. Minitab 18 utilizes a user friendly interface that allows one to quickly identify the appropriate procedure. The interface also allows one to easily export results including charts and graphs to facilitate the creation of presentations and reports. These Minitab 18 features would allow one to summarize data to better understand a population of interest, compare across groups, predict outcomes, and to develop forecasting models. These capabilities represent those that are generally relevant to the current course. Minitab 18 also includes many other statistical capabilities that can be further explored on the Minitab official website.

JMP:

JMP has a comprehensive set of statistical methods including introductory and advanced statistical procedures. JMP features include basic descriptive statistical procedures, graph and chart creation, diagnostic tests, analysis of variance, regression, time series and forecasting analyses, nonparametric analyses, cross-tabulation, chi-square and related tests, and other statistical procedures. JMP utilizes a user friendly interface that allows one to quickly identify the appropriate procedure. JMP also contains predictive analytic tools such as classification trees to classify data into groups. These JMP features would allow one to summarize data to better understand a population of interest, compare across groups, predict outcomes, and to develop forecasting models. These capabilities represent those that are generally relevant to the current course. JMP also includes many other statistical capabilities that can be further explored on the JMP official website.

- 1.47 (a) The population of interest include banking executives representing institutions of various sizes and U.S. geographic locations.
 (b) The collected sample includes 163 banking executives from institutions of various sizes and U.S. geographic locations.
 (c) A parameter of interest is the percentage of the population of banking executives that identify customer experience initiatives as an area where increased spending is expected.
 (d) A statistic used to the estimate the parameter in (c) is the percentage of the 163 banking executives included in the sample who identify customer experience initiatives as an area where increased spending is expected. In this case, the statistic is 55%.

1.48 The answers are based on an article titled “U.S. Satisfaction Still Running at Improved Level” and written by Lydia Saad (August 15, 2018). The article is located on the following site:
https://news.gallup.com/poll/240911/satisfaction-running-improved-level.aspx?g_source=link_NEWSV9&g_medium=NEWSFEED&g_campaign=item_&g_content=U.S.%2520Satisfaction%2520Still%2520Running%2520at%2520Improved%2520Level

The population of interest includes all individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.

The collected sample includes a random sample of 1,024 individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.

A parameter of interest is the percentage of the population of individuals aged 18 and older and live within the 50 U.S. states and the District of Columbia who are satisfied with the direction of the U.S.

A statistic used to the estimate the parameter in (c) is the percentage of the 1,024 individuals included in the sample. In this case, the statistic is 36%.

46 Chapter 1: Defining and Collecting Data

- 1.49 The answers were based on information obtained from the following site:
<https://www.pwc.com/gx/en/ceo-survey/2017/pwc-ceo-20th-survey-report-2017.pdf>
- (a) The population of interest included CEOs representing a mix of industries from 79 countries.
 - (b) The sample included 1,379 CEOs. The percentage of CEOs by continent were as follows: North America (11%), Western Europe (21%), Central and Eastern Europe (11%), Latin America (12%), Middle East and Africa (9%), and Asia Pacific (36%).
 - (c) A parameter of interest would be the percentage of CEOs among the population of interest that believe social media could have a negative impact on the level of stakeholder trust in their industry over the next few years.
 - (d) The statistic used to estimate the parameter in (c) is the percentage of CEOs among the 1,379 CEOs included in the sample who believe social media could have a negative impact on the level of stakeholder trust in their industry over the next few years. In this case, the statistic is 87%.
- 1.50
- (a) One variable collected with the American Community Survey is marital status with the following possible responses: now married, widowed, divorced, separated, and never married.
 - (b) The variable in (a) represents a categorical variable.
 - (c) Because the variable in (a) is a categorical, this question is not applicable. If one had chosen age in years from the American Community Survey as the variable, the answer to (c) would be discrete.
- 1.51 Answers will vary depending on the specific sample survey used. The below answers were based on the sample survey located at: <http://www.zarca.com/Online-Surveys-Non-Profit/association-salary-survey.html>
- (a) An example of a categorical variable included in the survey would be whether one had obtained an undergraduate degree with yes or no as possible answers.
 - (b) An example of a numerical variable included in the survey would be the respondent's annual base salary for the past year.
- 1.52
- (a) The population of interest consisted of 10,000 benefited employees of the University of Utah.
 - (b) The sample consisted of 3,095 employees of the University of Utah.
 - (c) Gender, marital status, and employment category represent categorical variables. Age in years, education level in years completed, and household income represent numerical variables.
- 1.53
- (a) Key social media platforms used represents a categorical variable. The frequency of social media usage represents a discrete numerical variable. Demographics of key social media platform users represent categorical variables.
 - (b)
 1. Which of the following is your preferred social media platform: YouTube, Facebook, or Twitter?
 2. What time of the day do you spend the most amount of time using social media: morning, afternoon, or evening?
 3. Please indicate your ethnicity?
 4. Which of the following do you most often use to access social media: mobile device, laptop computer, desktop computer, other device?
 5. Please indicate whether you are a home owner: Yes or No?
 - (c)
 1. For the past week, how many hours did you spend using social media?
 2. Please indicate your current age in years.
 3. What was your annual income this past year?
 4. Currently, how many friends have you accepted on Facebook?
 5. Currently, how many twitter followers do you have?

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 in 2011 Android, iOS, and OtherOS dominated the market in 2011. In 2012, 2013, 2014, and 2015 Android and iOS dominated the market. Android has increased its market share from 49.2% in 2011 to 80.7% in 2015. iOS has seen a slight decrease in market share from 18.8% in 2011 to 17.7% in 2015. OtherOS market share has declined from 19.8% in 2011 to 0.2% in 2015. Blackberry has also seen a significant decrease from 10.3% in 2011 to 0.3% in 2015. Microsoft reached its highest market share in 2013 with 3.3% and its lowest in 2015 with 1.1%.

(b) iOS increased its market share from 14.8% in 2014 to 17.7% in 2015. Android’s market share has remained steady from 2014 to 2015 while Microsoft, Blackberry, and OtherOS have all lost market share.

48 Chapter 2: Organizing and Visualizing Variables

2.4 (a)

Category	Total	Percentages
Bank Account or Service	202	9.330%
Consumer Loan	132	6.097%
Credit Card	175	8.083%
Credit Reporting	581	26.836%
Debt Collection	486	22.448%
Mortgage	442	20.416%
Student Loan	75	3.464%
Other	72	3.326%
Grand Total	2165	

(b) There are more complaints for credit reporting, debt collection, and mortgage than the other categories. These categories account for about 70% of all the complaints.

(c)

Company	Total	Percentage
Bank of America	42	3.64%
Capital One	93	8.07%
Citibank	59	5.12%
Ditech Financial	31	2.69%
Equifax	217	18.82%
Experian	177	15.35%
JPMorgan	128	11.10%
Nationstar Mortgage	39	3.38%
Navient	38	3.30%
Ocwen	41	3.56%
Synchrony	43	3.73%
Trans-Union	168	14.57%
Wells Fargo	77	6.68%
Grand Total	1153	

(d) Equifax, Trans-Union, and Experian, all of which are credit score companies, have the most complaints.

2.5 Executives anticipate Artificial Intelligence/Machine Learning technology to have the greatest disruptive impact on their firm in the next decade followed by Digital Technologies such as mobile, social media and IoT. They anticipate Financial Tech Solutions and Cloud computing will have some disruptive impact while Blockchain and other technologies to have little impact.

2.6 The largest sources of summer power-generating capacity in the United States are natural gas followed by coal. Nuclear, hydro, wind and other generate about the same, and solar generates very little.

2.7 (a)

Technologies	Frequency	Percentage
Wearable technology	9	10.00%
Blockchain technology	9	10.00%
Artificial Intelligence	17	18.89%
lot: retail insurance	23	25.56%
lot: commerical insurance	5	5.56%
Social media	27	30.00%
Grand Total	90	

(b) Professionals expect to be using Social media and Iot: retail insurance technologies the most over the next year followed by Artificial Intelligence. Professionals do not expect to be using Wearable, Blockchain, and Iot: commercial insurance technologies much over the next year.

2.8 (a) Table of row percentages:

OVERLOADED	GENDER		Total
	Male	Female	
Yes	44.08%	55.92%	100.00%
No	53.54%	46.46%	100.00%
Total	51.64%	48.36%	100.00%

Table of column percentages:

OVERLOADED	GENDER		Total
	Male	Female	
Yes	17.07%	23.13%	20.00%
No	82.93%	76.87%	80.00%
Total	100.00%	100.00%	100.00%

Table of total percentages:

OVERLOADED	GENDER		Total
	Male	Female	
Yes	8.82%	11.18%	20.00%
No	42.83%	37.17%	80.00%
Total	51.64%	48.36%	100.00%

(b) Approximately the same percentages of males and females as a percentage of the total number of people surveyed feel overloaded with too much information. As percentages of those who do and do not feel overloaded, the genders differ mildly. However, four times as many people do not feel overloaded at work than those that do.

50 Chapter 2: Organizing and Visualizing Variables

2.9 (a)

Column Percentage

CATEGORY	OUTCOME		Total
	Successful	Not Successful	
Film & Video	36.02%	36.81%	36.51%
Games	15.44%	18.24%	17.19%
Music	40.20%	24.38%	30.34%
Technology	8.34%	20.56%	15.96%
Total	100.00%	100.00%	100.00%

Row Percentages

CATEGORY	OUTCOME		Total
	Successful	Not Successful	
Film & Video	37.15%	62.85%	100.00%
Games	33.84%	66.16%	100.00%
Music	49.91%	50.09%	100.00%
Technology	19.69%	80.31%	100.00%
Total	37.67%	62.33%	100.00%

Total Percentages

CATEGORY	OUTCOME		Total
	Successful	Not Successful	
Film & Video	13.57%	22.95%	36.51%
Games	5.82%	11.37%	17.19%
Music	15.14%	15.20%	30.34%
Technology	3.14%	12.82%	15.96%
Total	37.67%	62.33%	100.00%

- (b) The row percentages are most informative because they provide a percentage of successful projects within each category which allows one to compare across categories.
- (c) Music kick starter projects were the most successful with approximately 50% of the projects succeeding compared to less than 20% of the Technology projects. The Film & Video and Games categories had success rates in between the Music and Technology categories, with success rates of 37% and 34% respectively.

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) $(166 + 100)/591 * 100 = 45.01\%$
 (b) $(124 + 77)/591 * 100 = 34.01\%$
 (c) $(59 + 65)/591 * 100 = 20.98\%$
 (d) 45% of the incidents took fewer than 2 days and 66% of the incidents were detected in less than 8 days. 79% of the incidents were detected in less than 31 days.

2.14 $\frac{216,000 - 61,000}{6} = 33,333.33$ so choose 40,000 as interval width

- (a) \$60,000 – under \$100,000; \$100,000 – under \$140,000; \$140,000 – under \$180,000; \$180,000 – under \$220,000; \$220,000 – under \$260,000; \$260,000 – under \$300,000
 (b) \$40,000
 (c) $\frac{60,000 + 100,000}{2} = 80,000$ similarly, the remaining class midpoints are \$120,000; \$160,000; \$200,000; \$240,000; \$280,000

2.15 (a)

222.67	262.50	262.67	276.40	278.00	290.83	292.87
298.00	318.67	324.33	332.93	345.09	346.70	380.67
398.55	418.14	422.45	423.50	429.00	441.00	492.71
505.77	539.68	571.50	585.20	696.33	718.50	726.40
789.20	878.20					

(b)

NBA Cost		
Cost \$	Frequency	Percentage
200 but less than 300	8	27%
300 but less than 400	7	23%
400 but less than 500	6	20%
500 but less than 600	4	13%
600 but less than 700	1	3%
700 but less than 800	3	10%
800 but less than 900	1	3%
Total	30	100%

- (c) 70% of the costs to attend a NBA basketball game are between \$200 and \$500 with 27% of the costs between \$200 and \$300. Three teams or 10% of the NBA teams have costs between \$700 and \$800.

52 Chapter 2: Organizing and Visualizing Variables

2.16 (a)

Electricity Costs		
Electricity Costs	Frequency	Percentage
\$80 but less than \$100	4	8%
\$100 but less than \$120	7	14%
\$120 but less than \$140	9	18%
\$140 but less than \$160	13	26%
\$160 but less than \$180	9	18%
\$180 but less than \$200	5	10%
\$200 but less than \$220	3	6%

(b)

Electricity Costs	Frequency	Percentage	Cumulative %
\$ 99	4	8.00%	8.00%
\$119	7	14.00%	22.00%
\$139	9	18.00%	40.00%
\$159	13	26.00%	66.00%
\$179	9	18.00%	84.00%
\$199	5	10.00%	94.00%
\$219	3	6.00%	100.00%

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

2.17 (a)

Commuting Time (minutes)	Frequency	Percentage
200 but less than 230	12	40%
230 but less than 260	9	30%
260 but less than 290	4	13%
290 but less than 320	3	10%
320 but less than 350	1	3%
350 but less than 380	1	3%
	30	100%

(b)

Commuting Time (minutes)	Frequency	Percentage	Cumulative %
200 but less than 230	12	40%	40%
230 but less than 260	9	30%	70%
260 but less than 290	4	13%	83%
290 but less than 320	3	10%	93%
320 but less than 350	1	3%	97%
350 but less than 380	1	3%	100%
	30	100%	

2.17 (c) cont. The majority of commuters living in or near cities spend from 200 up to 230 minutes commuting each week. 70% of commuters spend from 200 up to 260 minutes commuting each week.

2.18 (a), (b)

Credit Score	Frequency	Percent (%)	Cumulative Percent (%)
560 – under 580	4	0.16	0.16
580 – under 600	24	0.93	1.09
600 – under 620	68	2.65	3.74
620 – under 640	290	11.28	15.02
640 – under 660	548	21.32	36.34
660 – under 680	560	21.79	58.13
680 – under 700	507	19.73	77.86
700 – under 720	378	14.71	92.57
720 – under 740	168	6.54	99.11
740 – under 760	22	0.86	99.96
760 – under 780	1	0.04	100.00

(c) The average credit scores are concentrated between 620 and 720.

2.19 (a), (b)

Bin	Frequency	Percentage	Cumulative %
-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)

Time in Seconds	Frequency	Percent (%)
5 – under 10	8	16%
10 – under 15	8	30%
15 – under 20	8	36%
20 – under 25	8	12%
25 – under 30	8	6%

54 Chapter 2: Organizing and Visualizing Variables

2.20 (b)
cont.

Time in Seconds	Percentage Less Than
5	0
10	16
15	46
20	82
25	94
30	100

(c) The target is being met since 82% of the calls are being answered in less than 20 seconds

2.21 (a)

Call Duration (seconds)	Frequency	Percentage
60 up to 119	7	14%
120 up to 179	12	24%
180 up to 239	11	22%
240 up to 299	11	22%
300 up to 359	4	8%
360 up to 419	3	6%
420 and longer	2	4%
	50	100%

(b)

Call Duration (seconds)	Frequency	Percentage	Cumulative %
60 up to 119	7	14%	14%
120 up to 179	12	24%	38%
180 up to 239	11	22%	60%
240 up to 299	11	22%	82%
300 up to 359	4	8%	90%
360 up to 419	3	6%	96%
420 and longer	2	4%	100%
	50	100%	

(c) The call center's target of call duration less than 240 seconds is only met for 60% of the calls in this data set.

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%

2.22 (a) Manufacturer B:
cont.

Bin Cell	Frequency	Percentage	Cumulative Pctage.
7,500 but less than 8,500	2	5.00%	5.00%
9,500 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

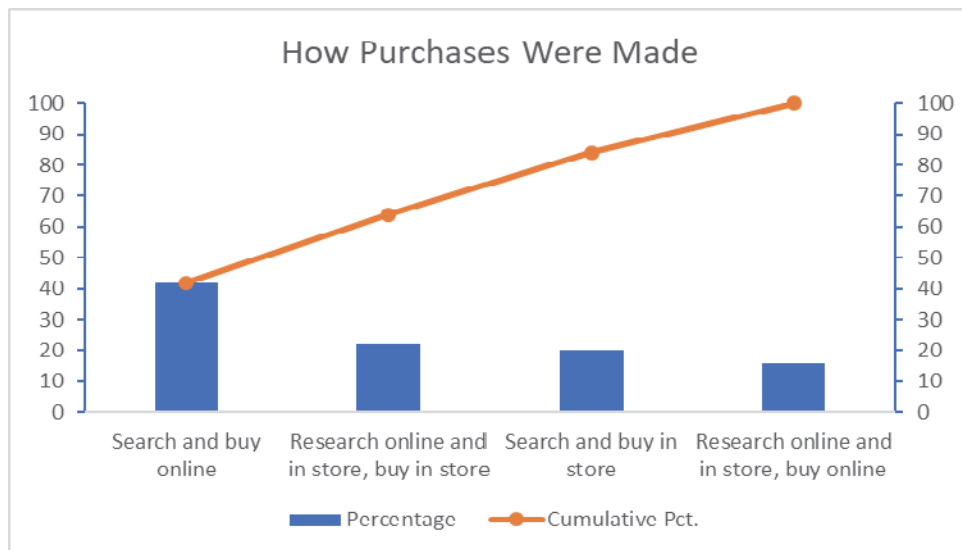
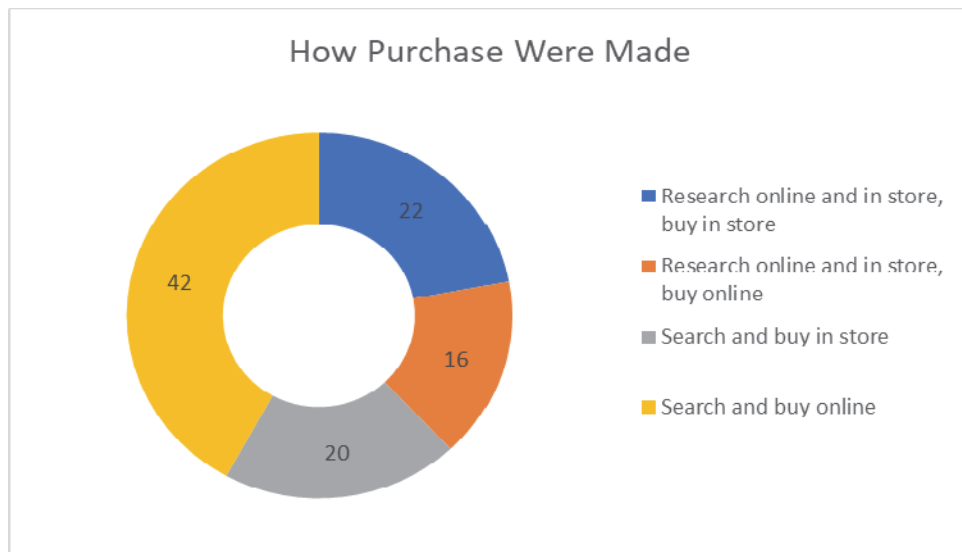
2.23 (a)

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%

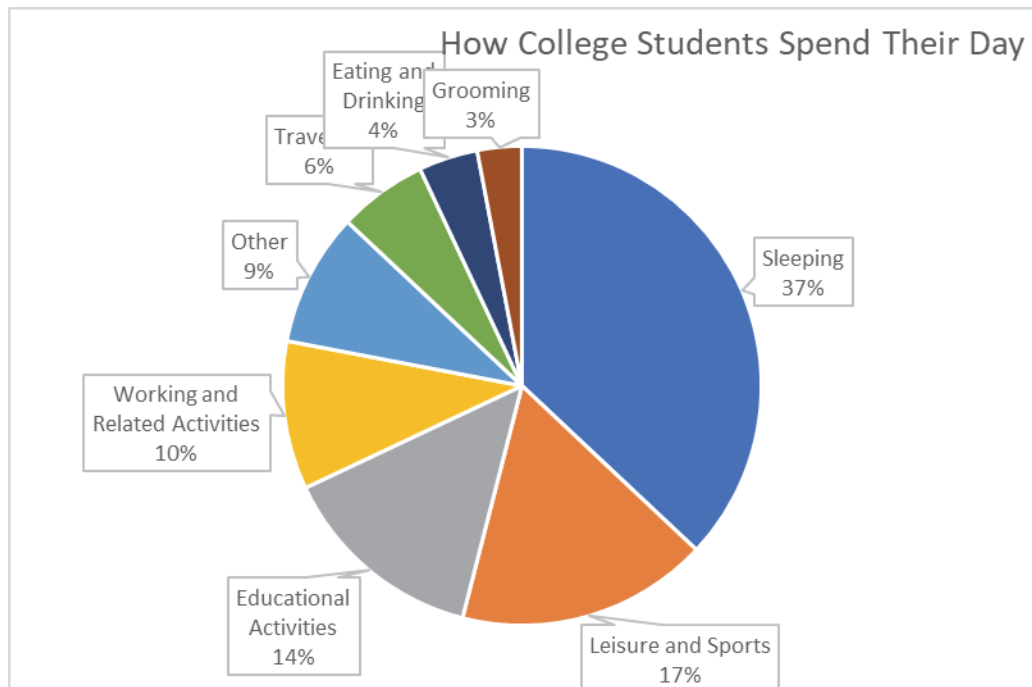
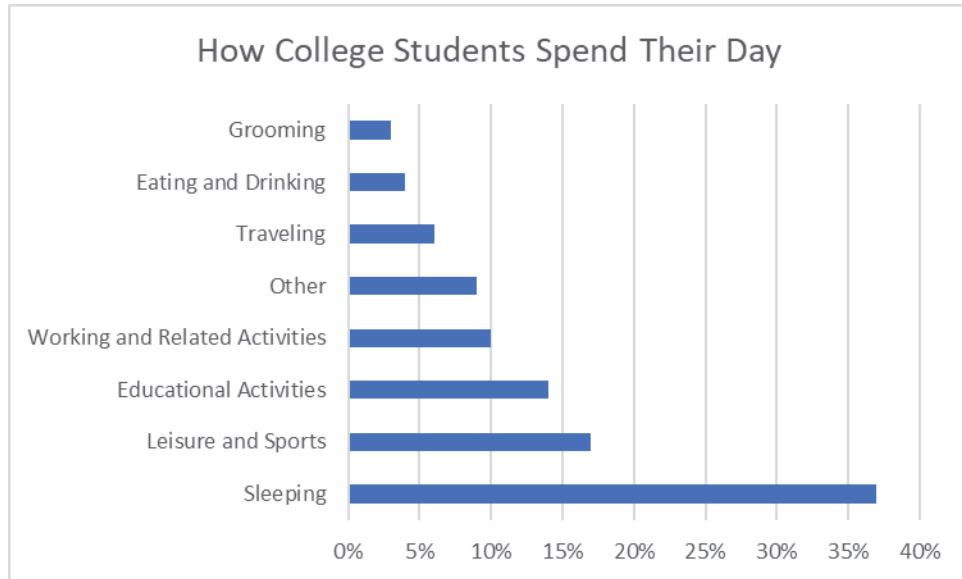
(b) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.24 (a)



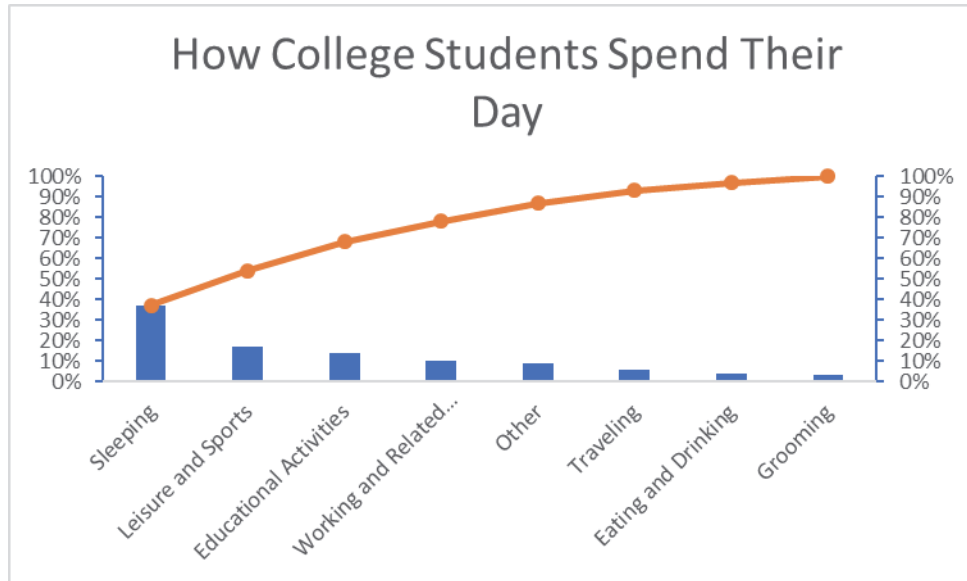
- 2.24 (b) The Pareto chart is best for portraying these data because it not only sorts the frequencies in descending order but also provides the cumulative line on the same chart.
- (c) You can conclude that searching and buying online was the highest category and the other three were equally likely.

2.25 (a)



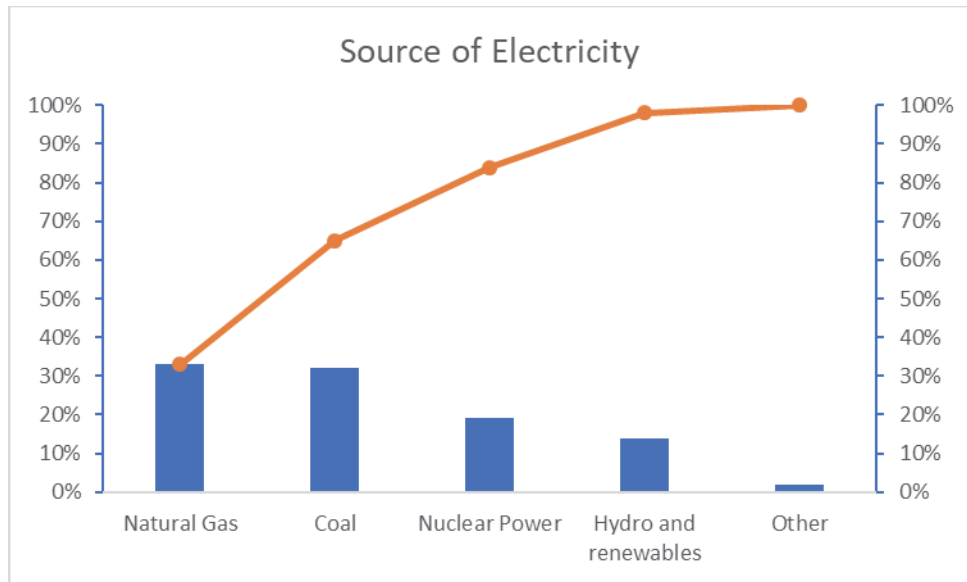
58 Chapter 2: Organizing and Visualizing Variables

2.25 (a)
cont.



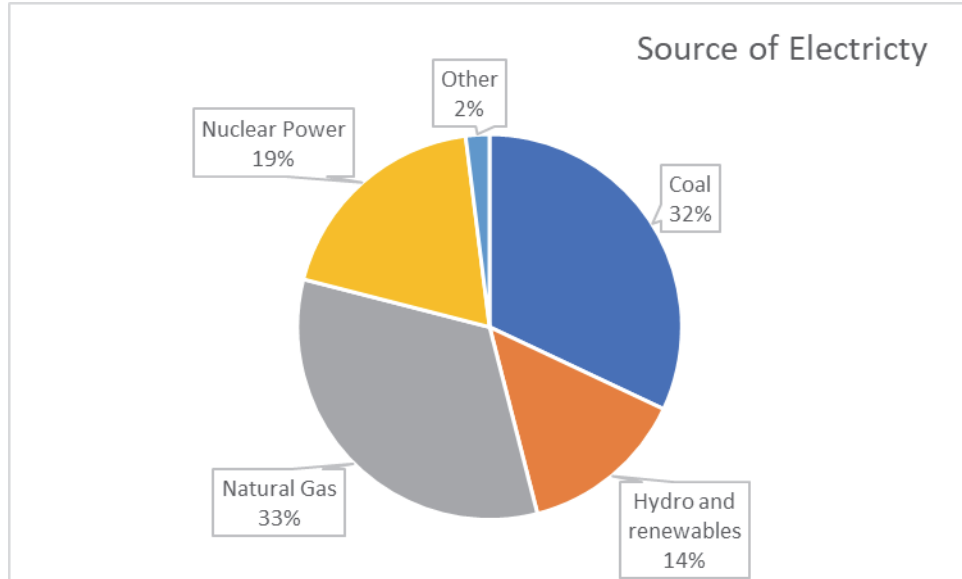
- (b) The Pareto diagram is better than the pie chart or the bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.
- (c) From the Pareto diagram it is obvious that more than 50% of their day is spent sleeping and taking part in leisure and sports.

2.26 (a)



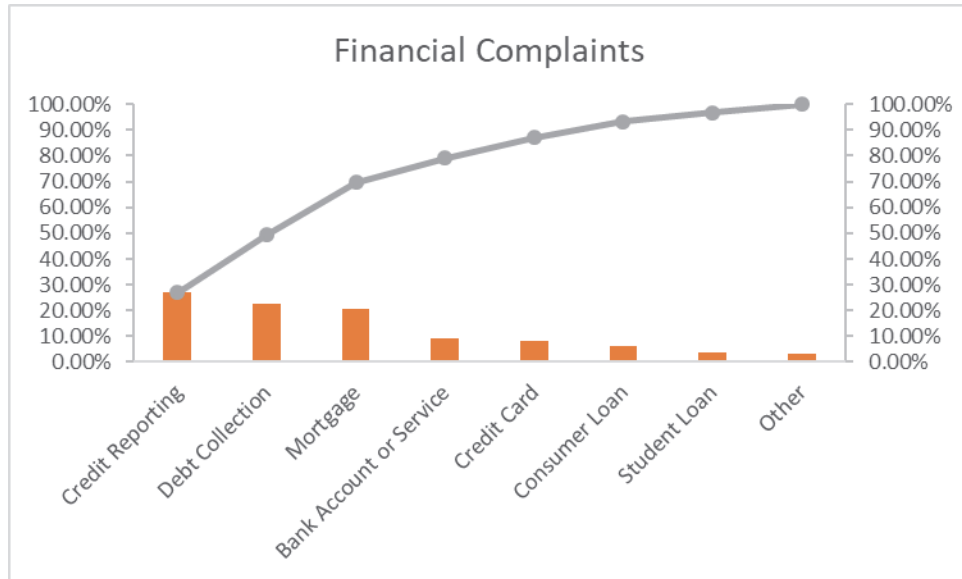
- (b) $32\% + 19\% + 33\% = 84\%$

2.26 (c)
cont.



(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.

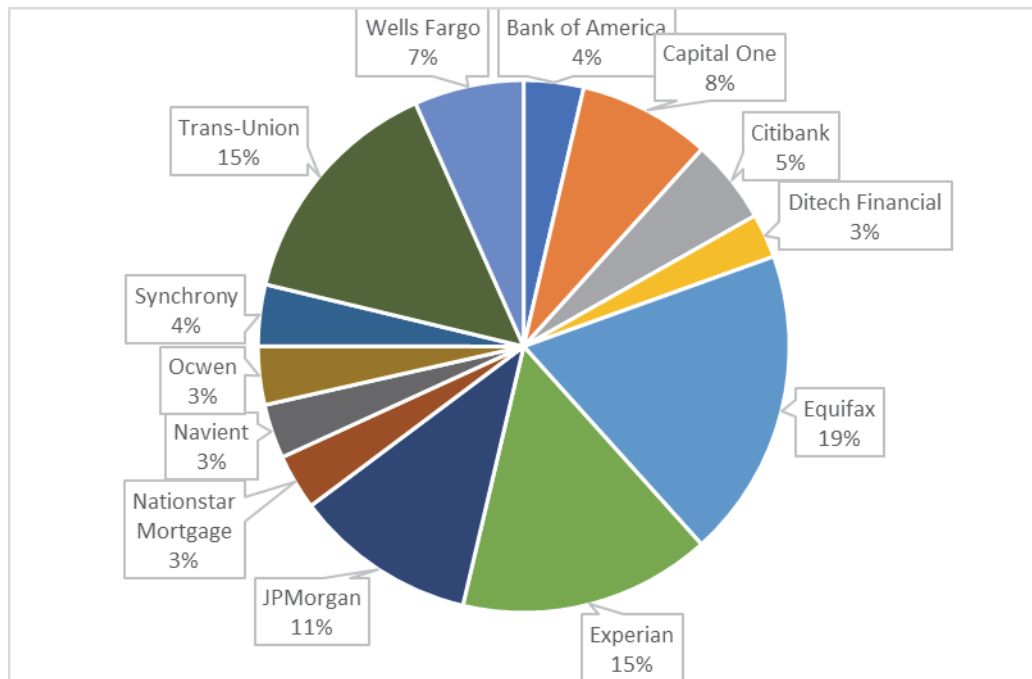
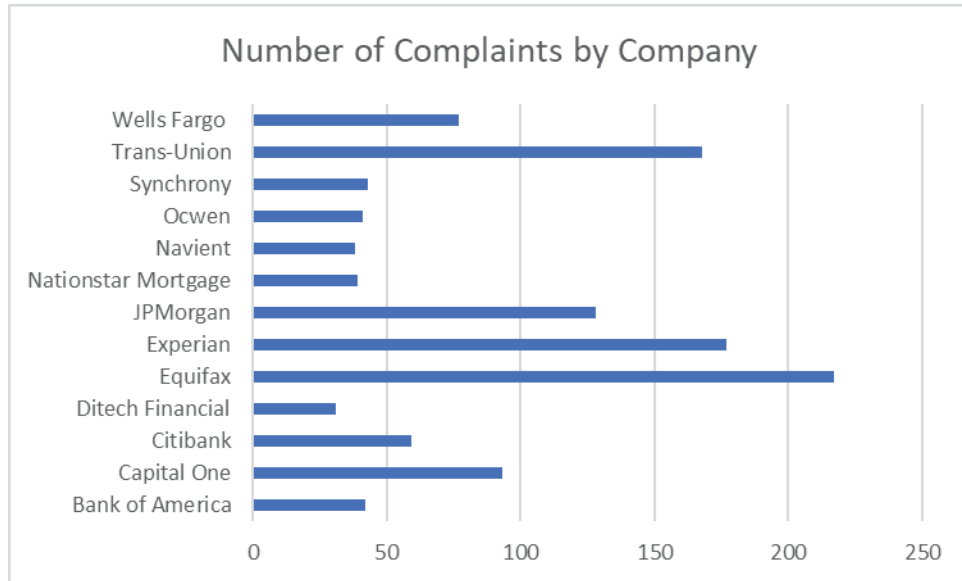
2.27 (a)



(b) The “vital few” reasons for the categories of complaints are “Credit Reporting”, “Debt Collection”, and “Mortgage” which account for 70% of the complaints. The remaining are the “trivial many” which make up 30% of the complaints.

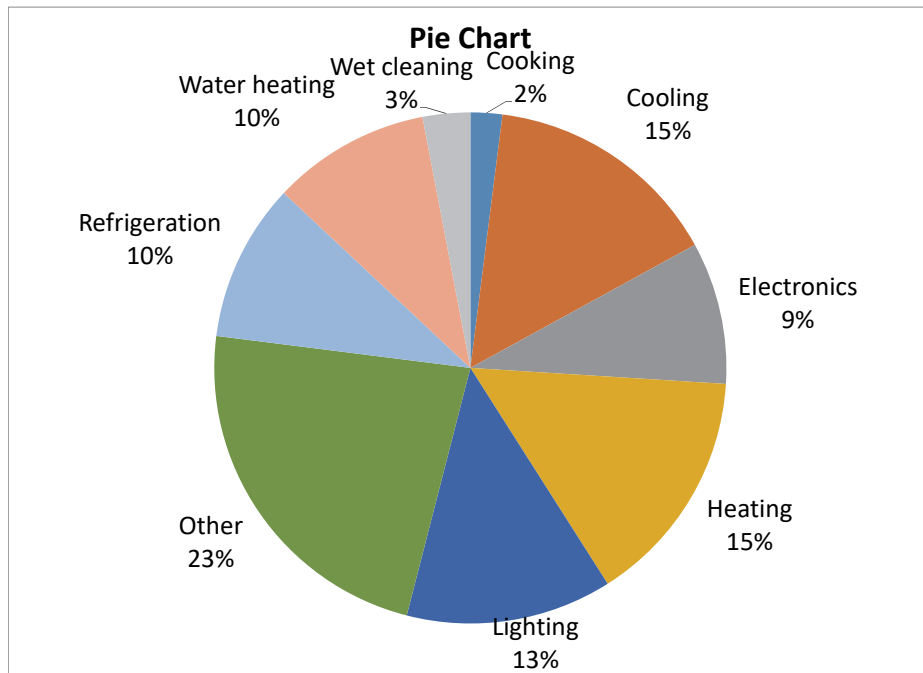
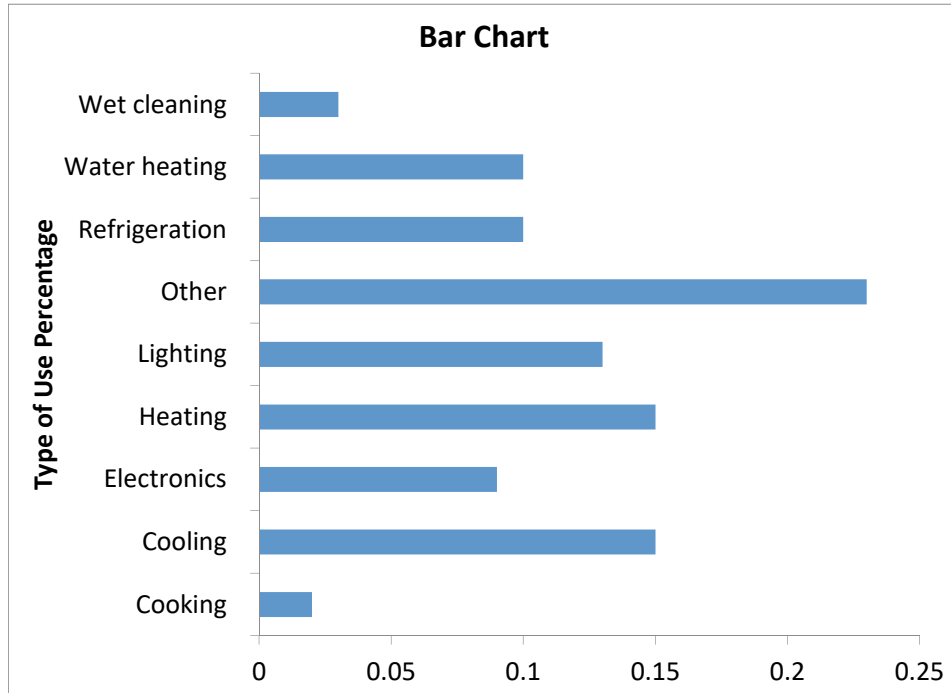
60 Chapter 2: Organizing and Visualizing Variables

2.27 (c)
cont.



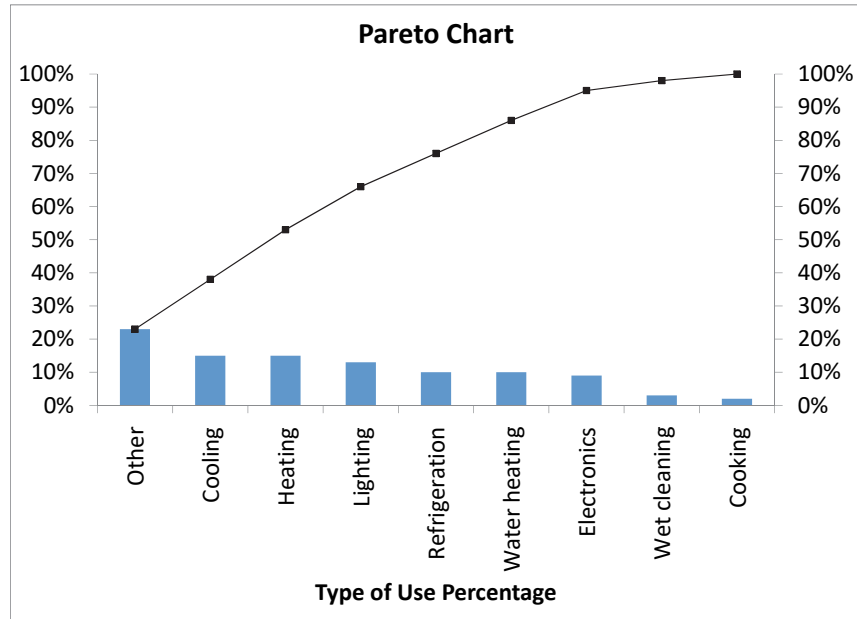
(d) The Pareto diagram is better than the pie chart and bar chart because it allows you to see which companies account for most of the complaints.

2.28 (a)



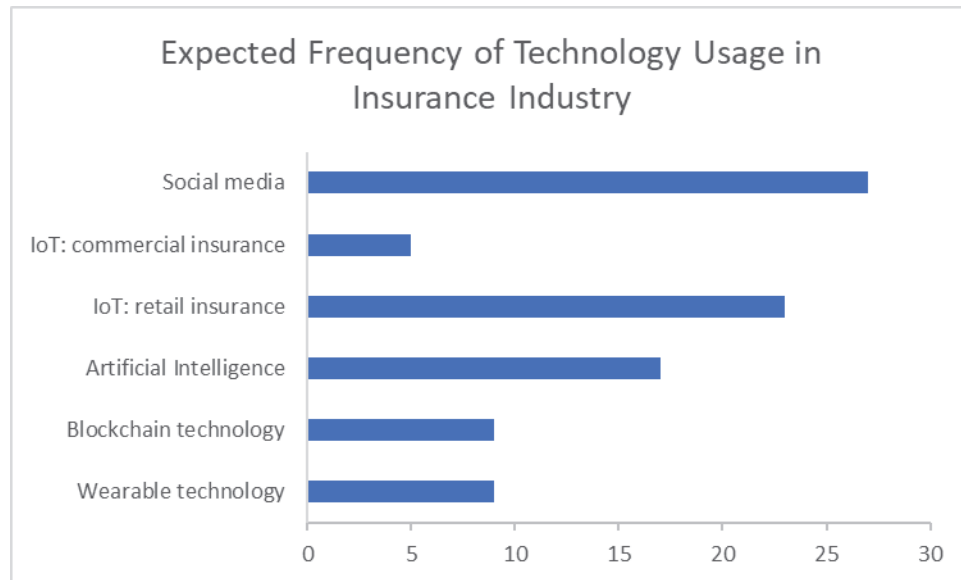
62 Chapter 2: Organizing and Visualizing Variables

2.28 (a)
cont.

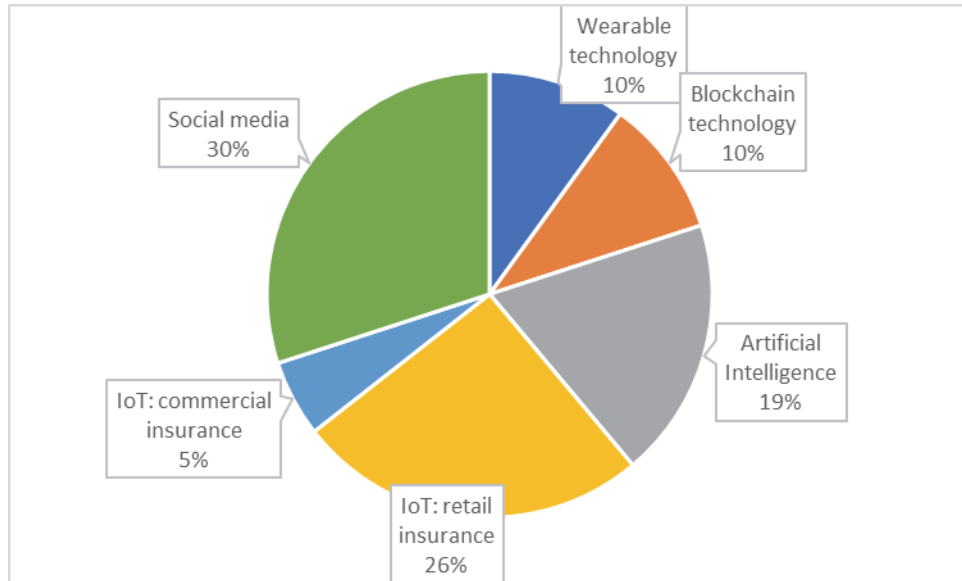


- (b) The Pareto diagram is better than the pie chart and bar chart because it not only sorts the frequencies in descending order; it also provides the cumulative polygon on the same scale.
- (c) Other, cooling, heating and lighting accounted for 66% of the residential electricity consumption in the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

2.29 (a)

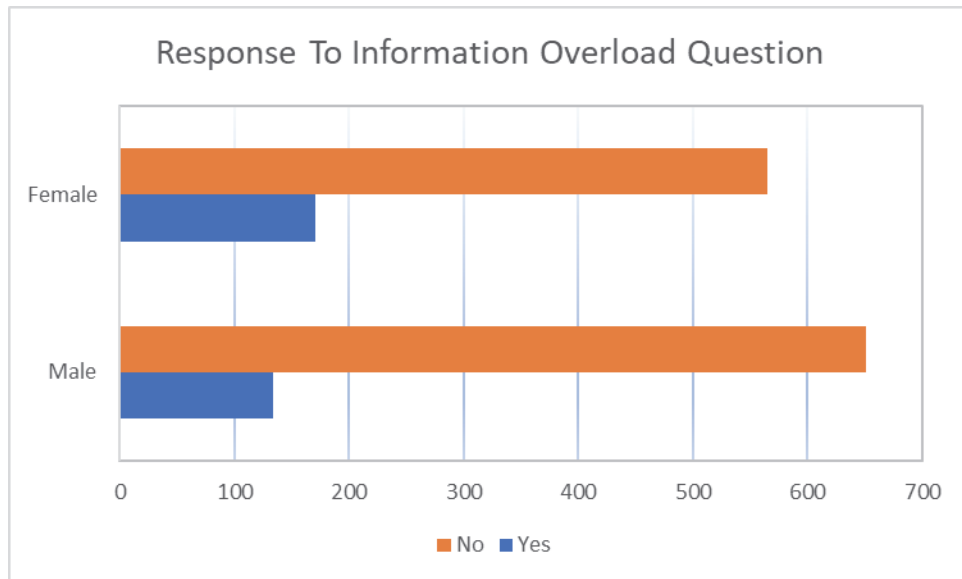


2.29 (a)
cont.



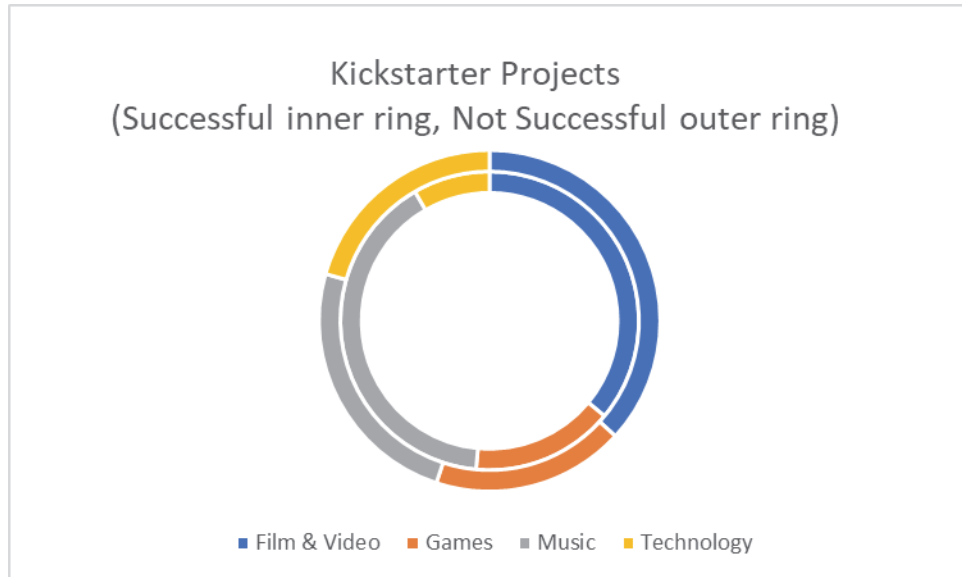
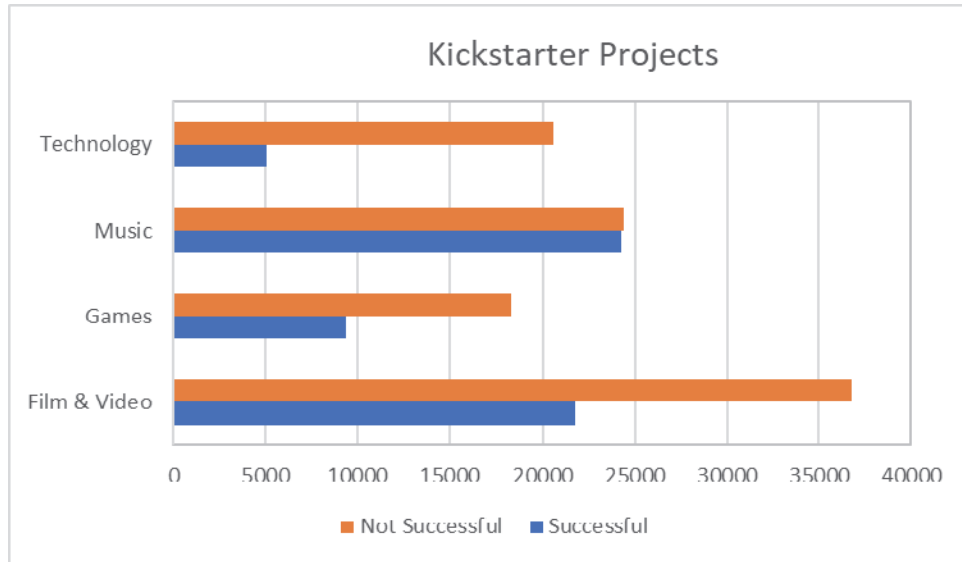
(b) Insurance professionals expect Social Media, AI, and IOT retail insurance to be most used in the insurance industry in the coming year.

2.30 (a)



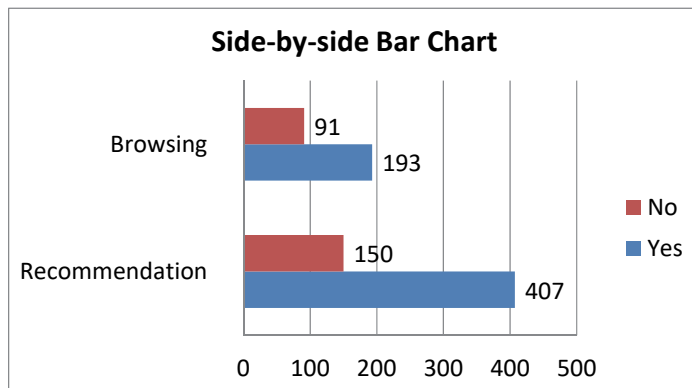
(b) Females are more likely to be overloaded with information.

2.31 (a)



(b) Of the successful kickstarter projects, music projects make up the largest part.

2.32 (a)



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

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 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)

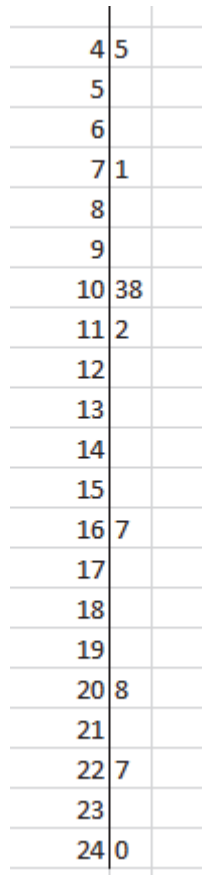
Stem Unit	100
2	2 6 6 8 8 9
3	0 2 2 3 5 5 8
4	0 2 2 2 3 4 9
5	1 4 7 9
6	
7	0 2 3 9
8	8

(b) The results are concentrated between \$220 and \$490.

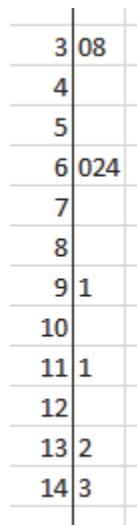
2.37 (a) Download Speed 4.5 7.1 10.3 10.8 11.2 16.7 20.8 22.7 24.0
Upload Speed 3.0 3.8 6.0 6.2 6.4 9.1 11.1 13.2 14.3

66 Chapter 2: Organizing and Visualizing Variables

2.37 (b) Download Speeds: Stem unit :1
cont.

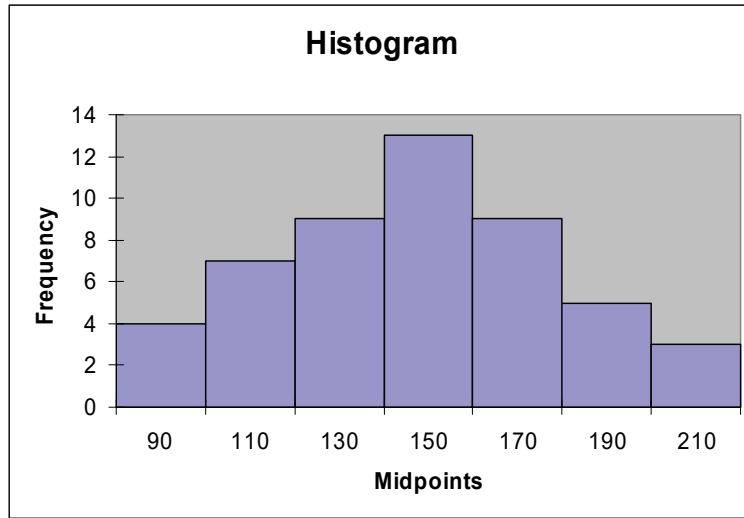


Upload Speeds: Stem unit 1

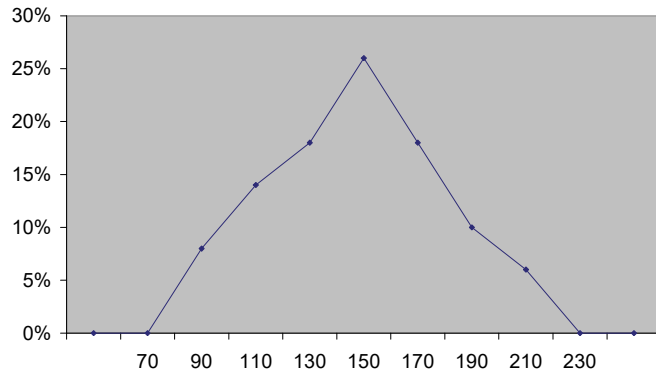


- (b) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.
- (c) Download speeds are concentrated around 10 mbs and Upload speeds are concentrated around 6 mbs.

2.38 (a)

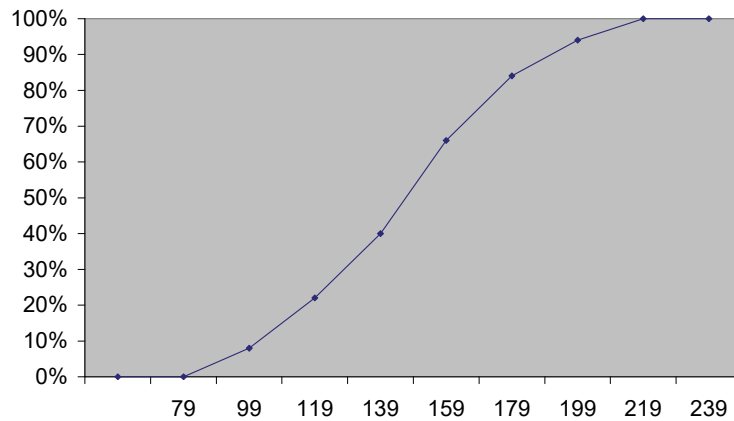


Percentage Polygon



(b)

Cumulative Percentage Polygon



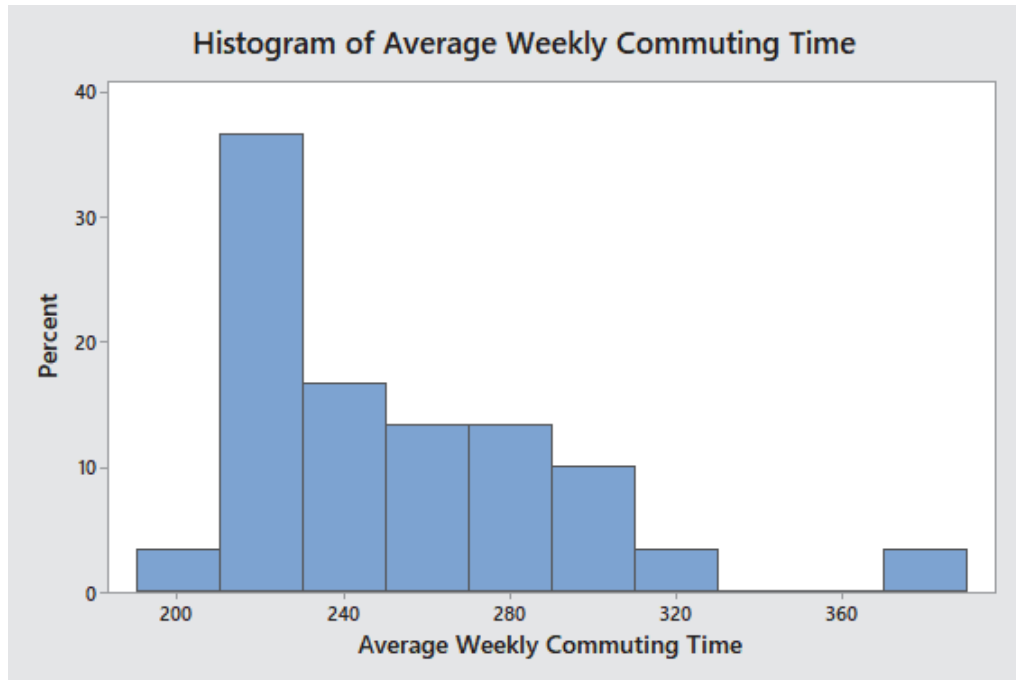
(c) The majority of utility charges are clustered between \$120 and \$180.

2.39 The cost of attending a baseball game is concentrated around \$65 with twelve teams at that cost. Four teams have costs of \$85 and one team is has the highest cost of \$115.

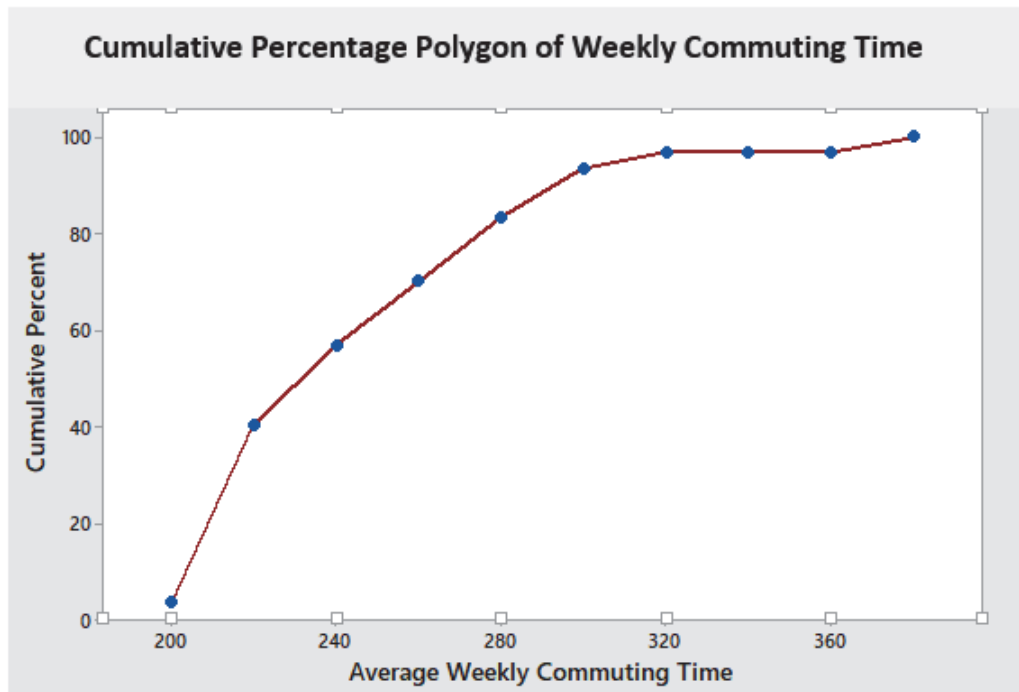
68 Chapter 2: Organizing and Visualizing Variables

2.40 Property taxes on a \$176K home seem concentrated between \$700 and \$2,200 and also between \$3,200 and \$3,700.

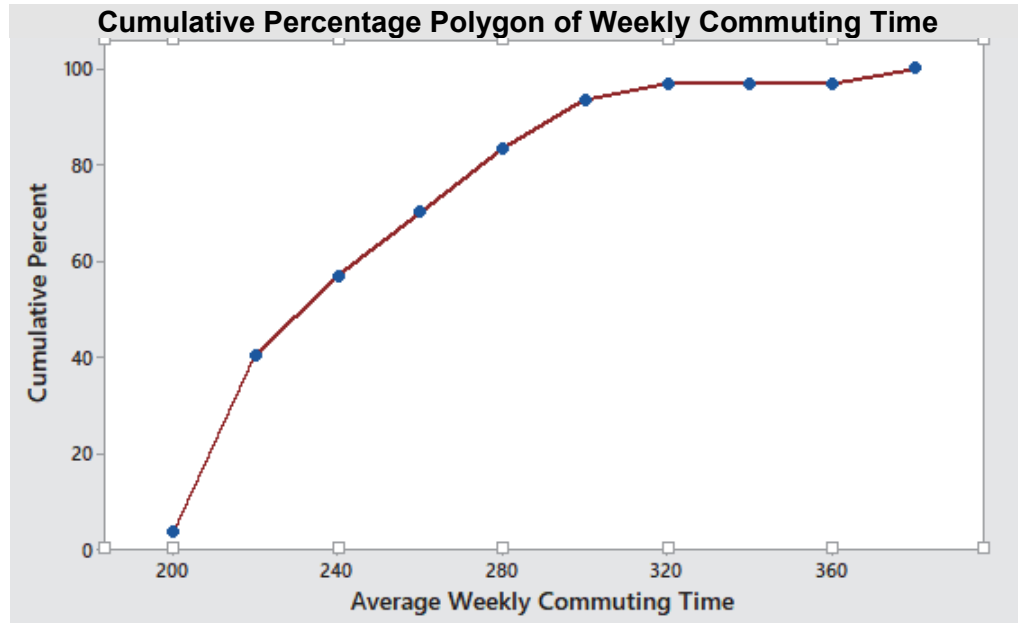
2.41 (a)



(b)

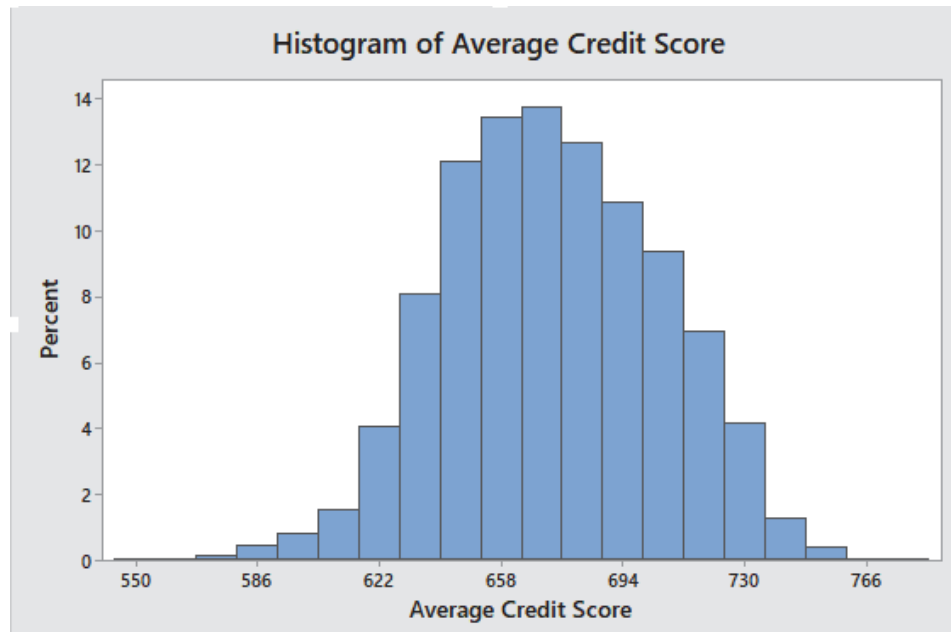


2.41 (b)
cont.



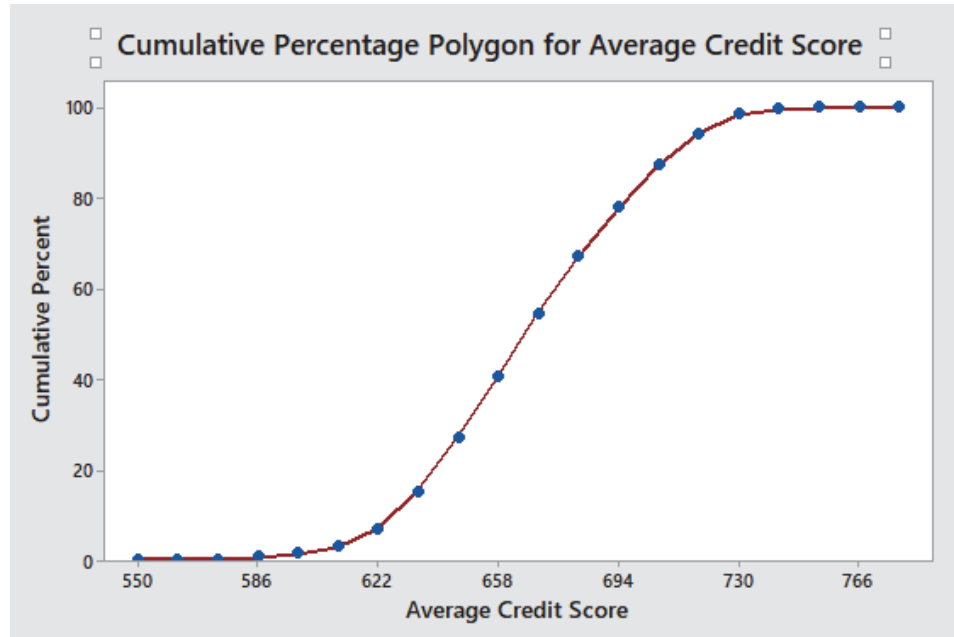
(c) The majority of Americans living in cities spend an average of 280 minutes or less commuting each week. Approximately 38% spend between 210 and 230 minutes commuting each week with a small percentage commuting spending between 370 to 380 minutes commuting each week.

2.42 (a)



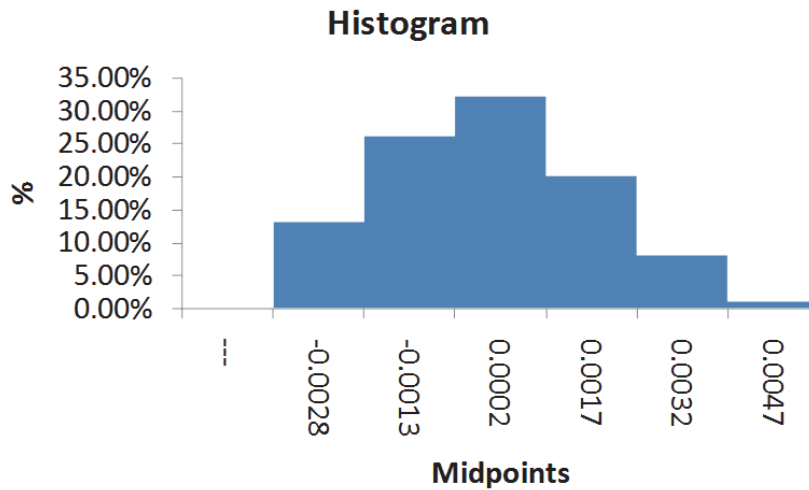
70 Chapter 2: Organizing and Visualizing Variables

2.42 (b)
cont.



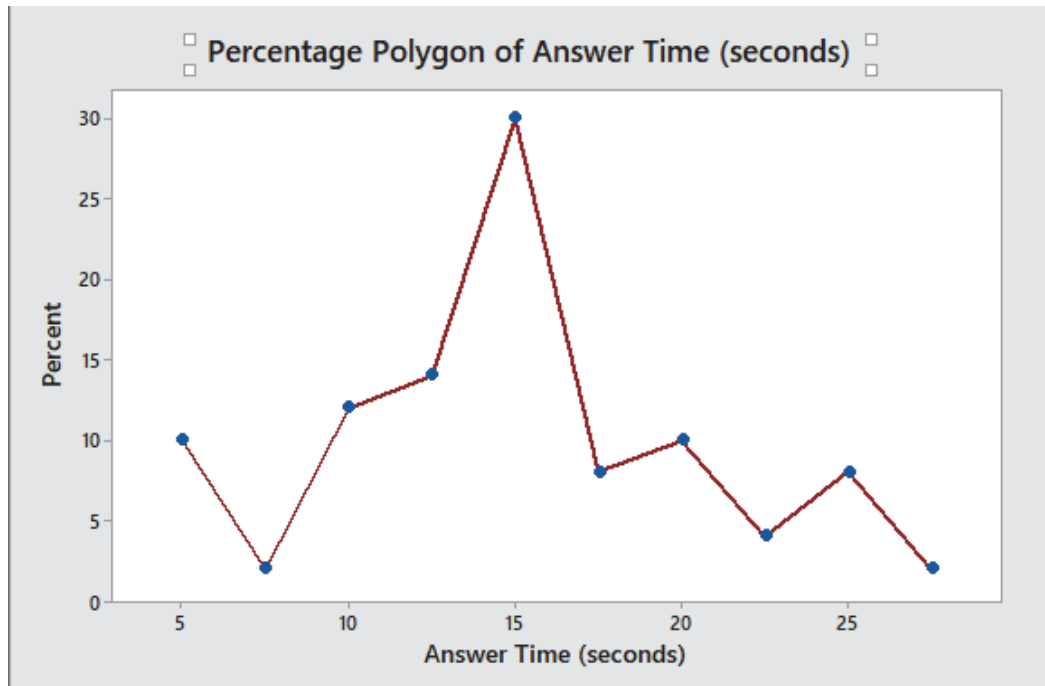
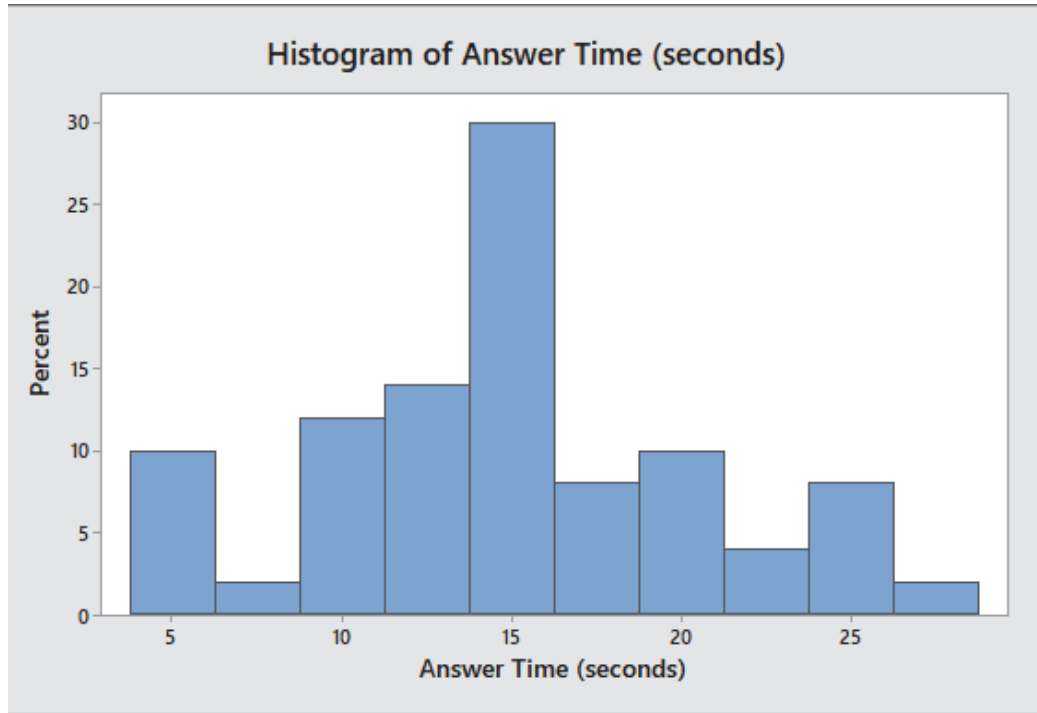
(c) The average credit scores are concentrated between 622 and 730.

2.43 (a)



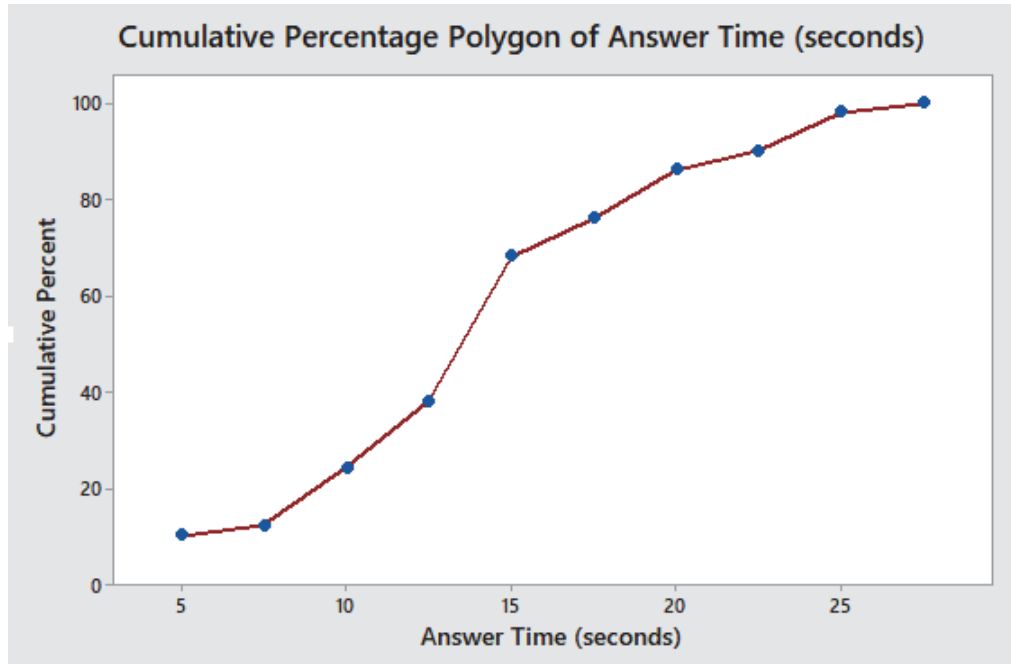
(b) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.44 (a)



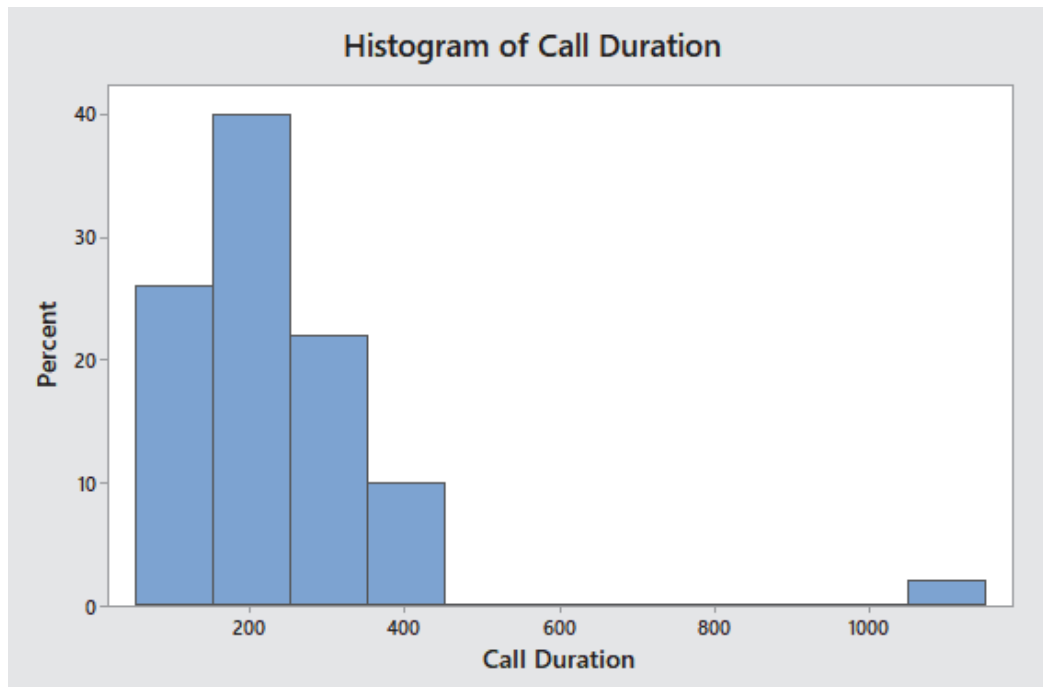
72 Chapter 2: Organizing and Visualizing Variables

2.44 (b)
cont.

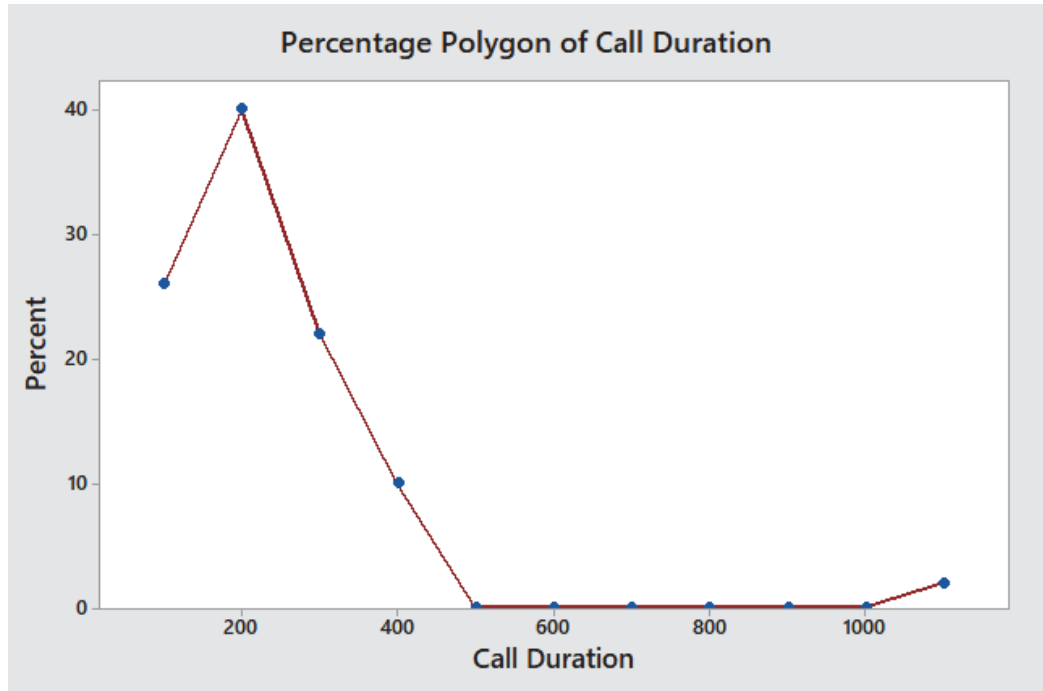


(c) The target is being met since 82% of the calls are being answered in less than 20 seconds.

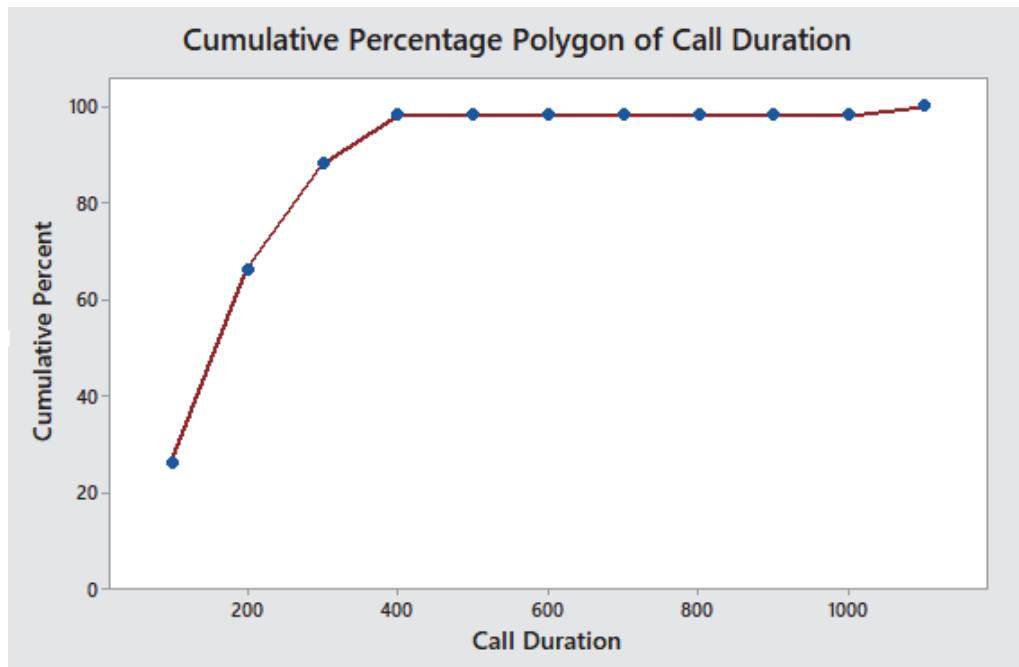
2.45 (a)



2.45 (a)
cont.

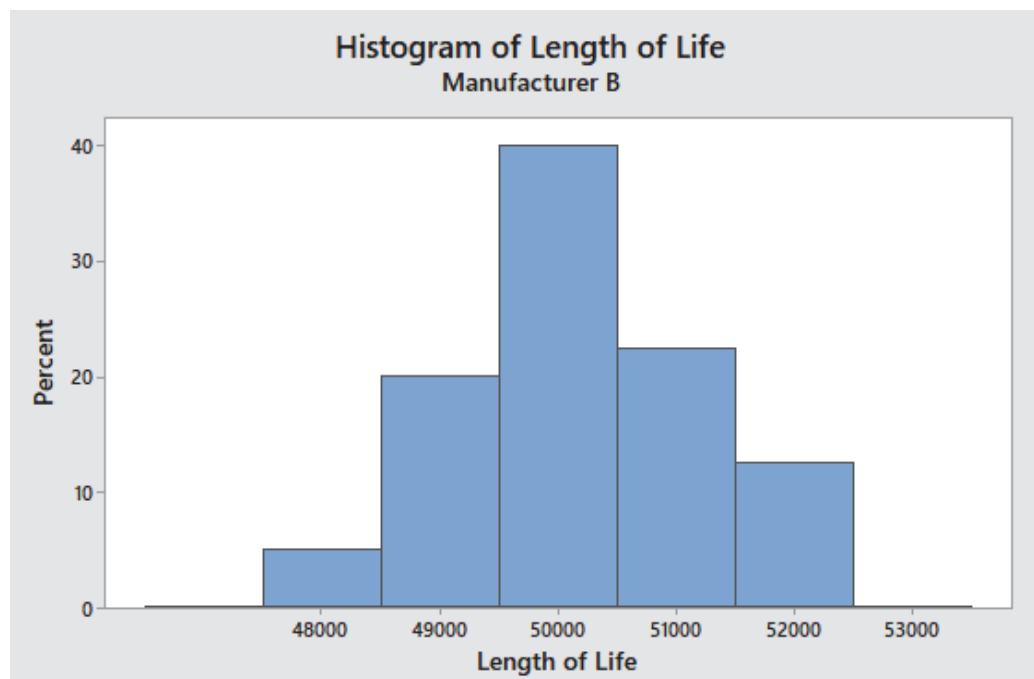
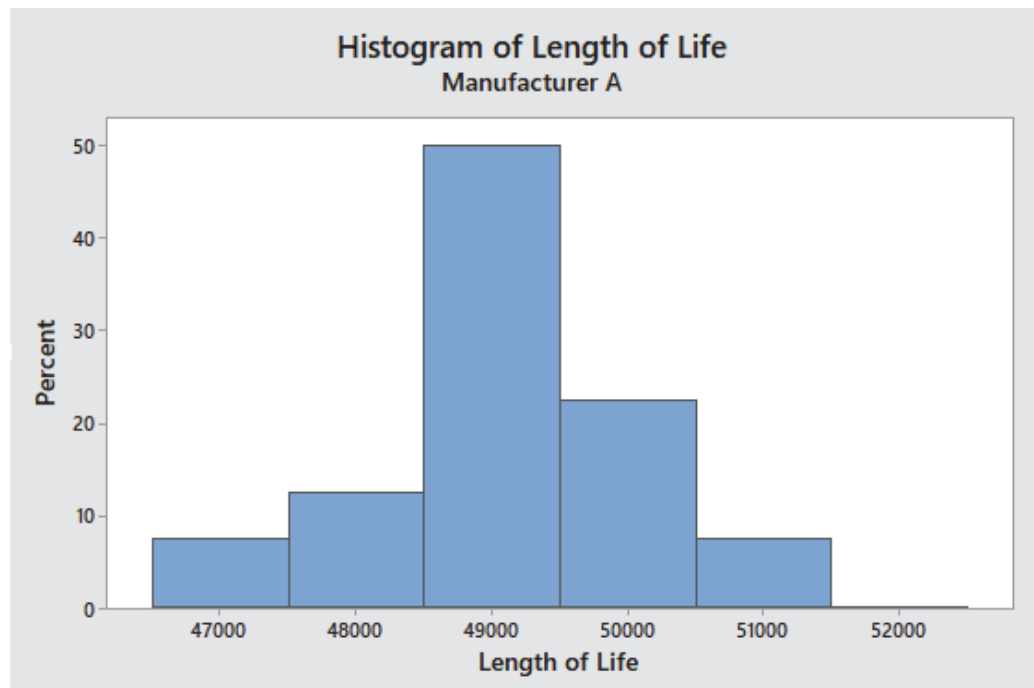


(b)

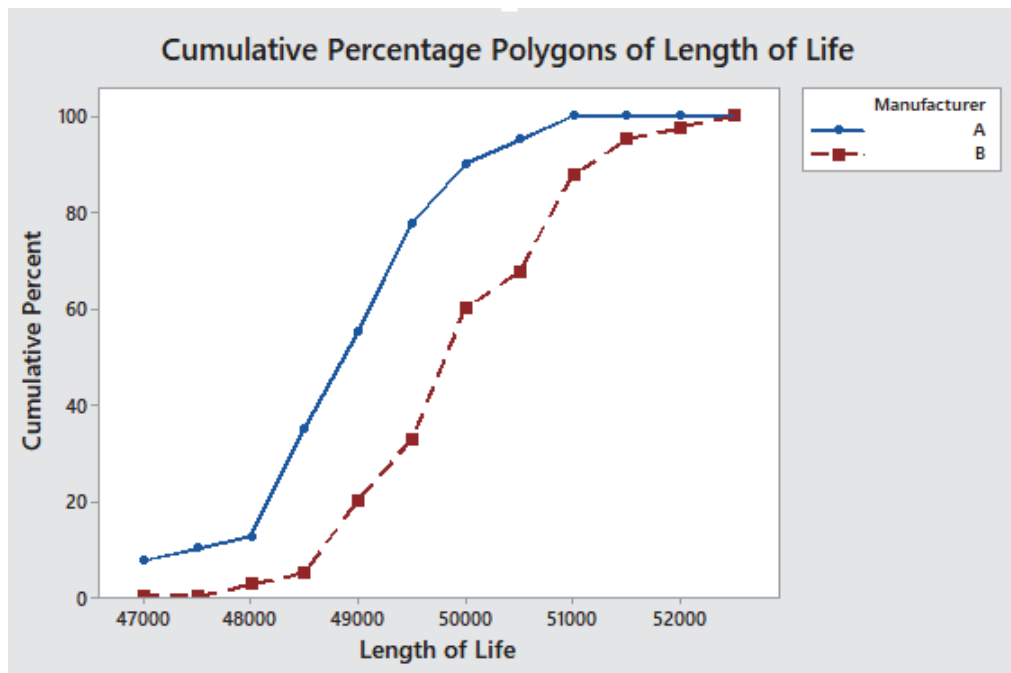
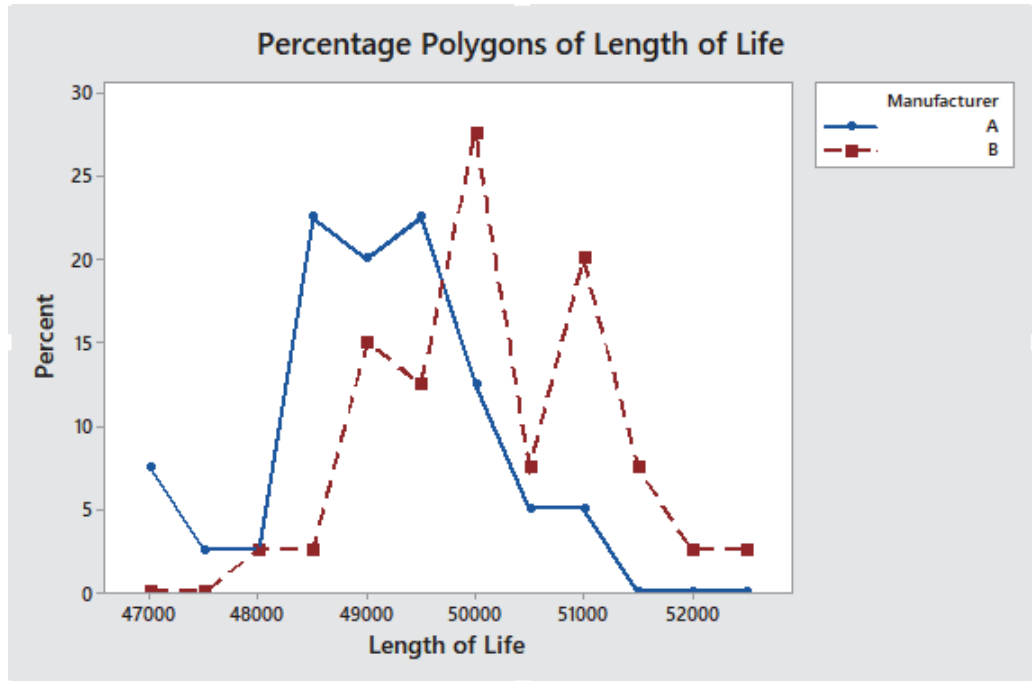


(c) The call center's target of call duration less than 240 seconds is only met for 60% of the calls in this data set.

2.46 (a)

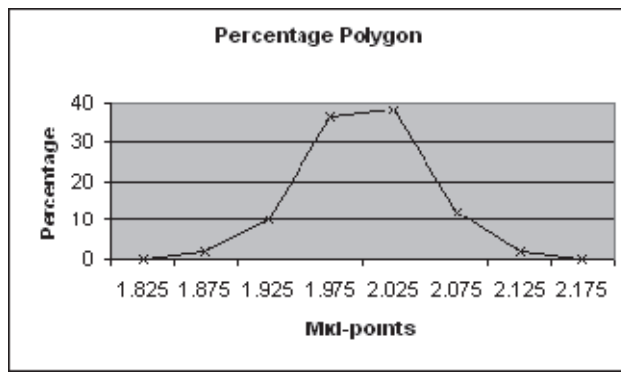
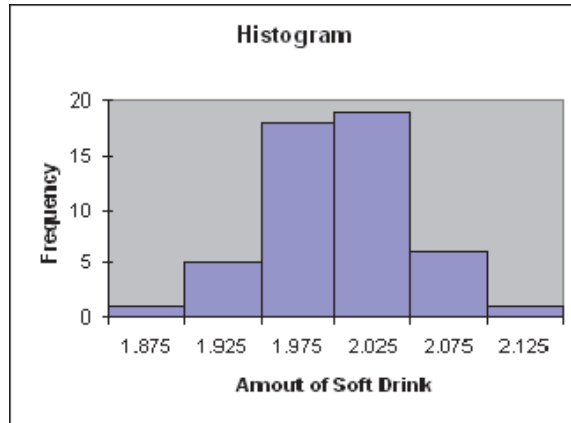


2.46 (b)
cont.



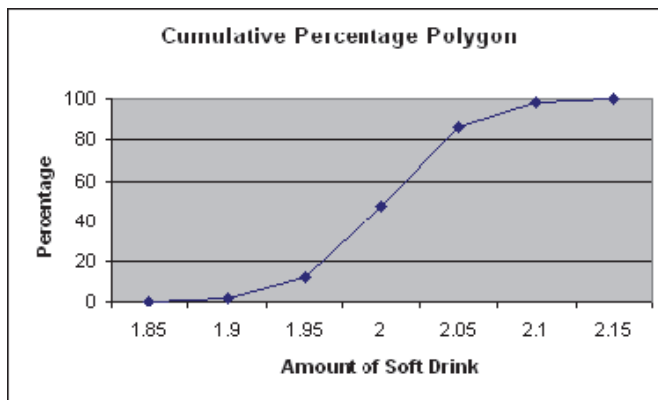
(c) Manufacturer B produces bulbs with longer lives than Manufacturer A

2.47 (a)



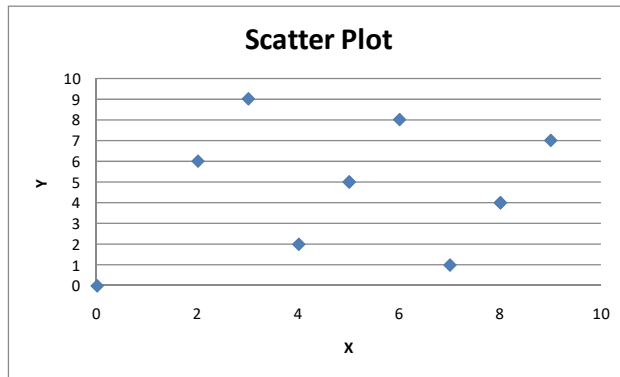
(b)

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



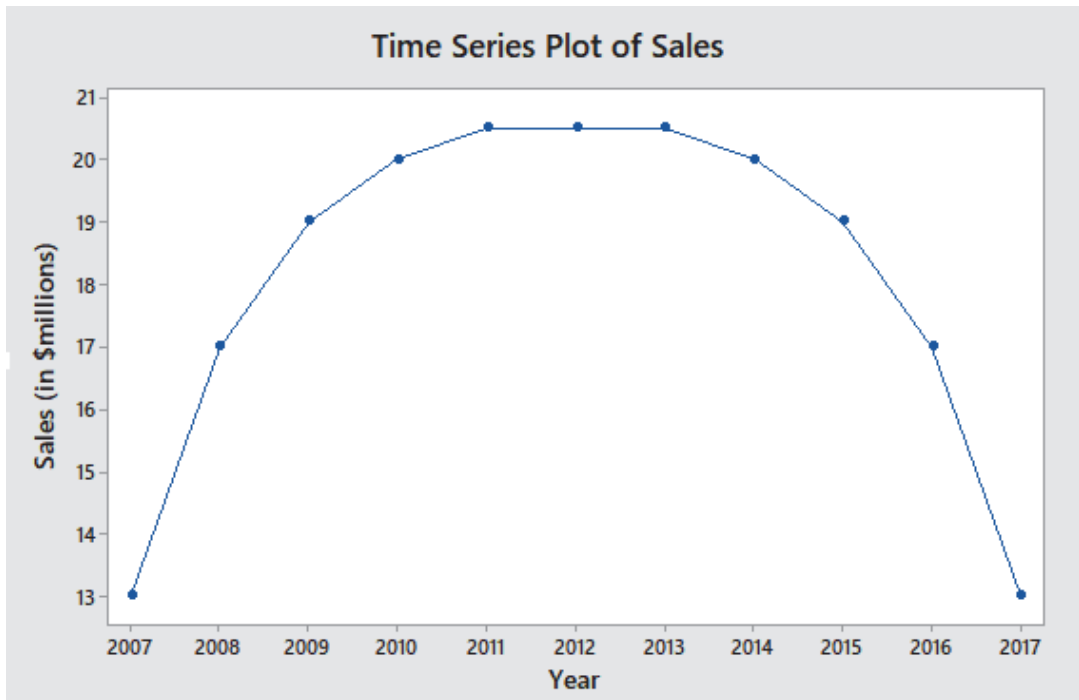
(c) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.48 (a)



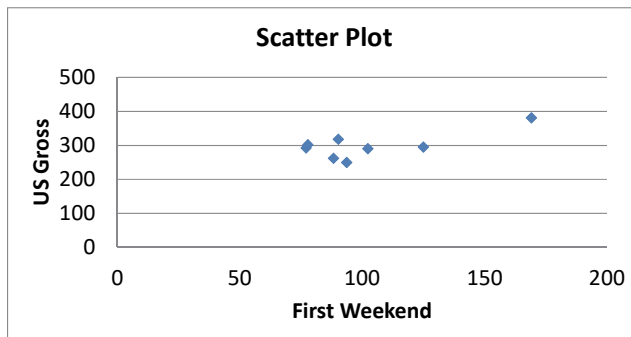
(b) There is no relationship between X and Y .

2.49 (a)



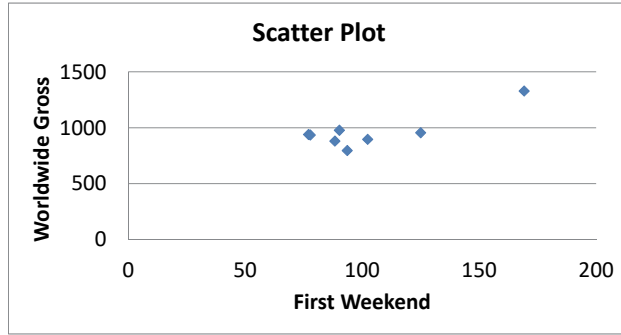
(b) Annual sales appear to be increasing until 2011 then remain flat from 2011 to 2013 followed by a decline after 2013.

2.50 (a)



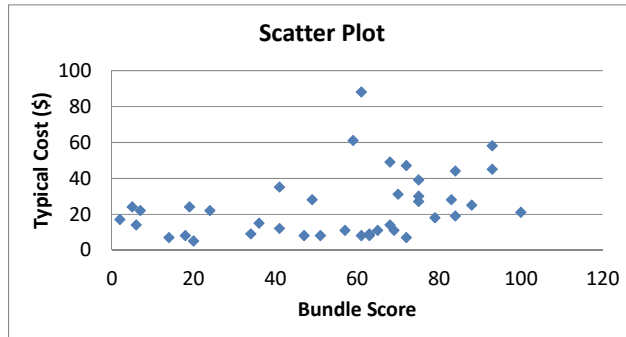
78 Chapter 2: Organizing and Visualizing Variables

2.50 (b)
cont.



(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*.

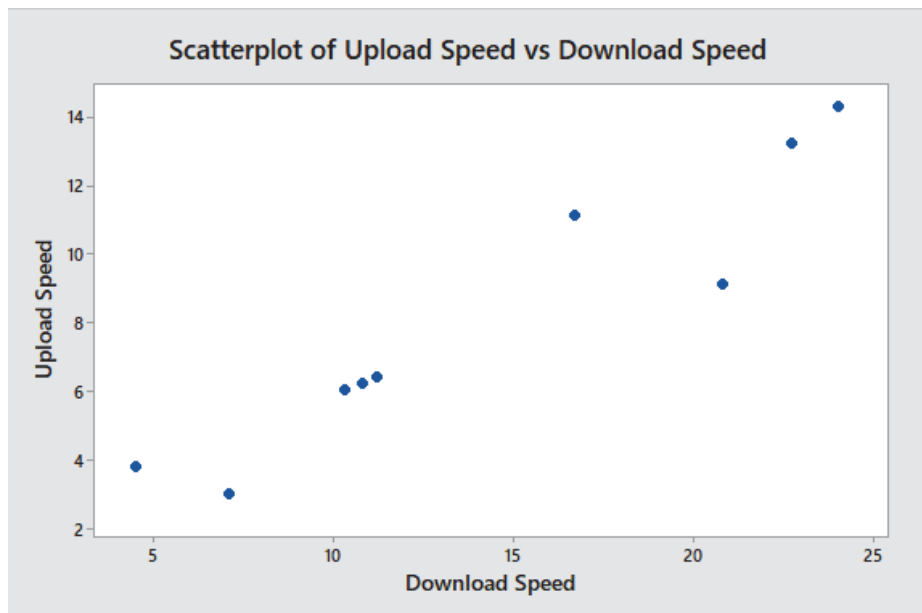
2.51 (a)



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

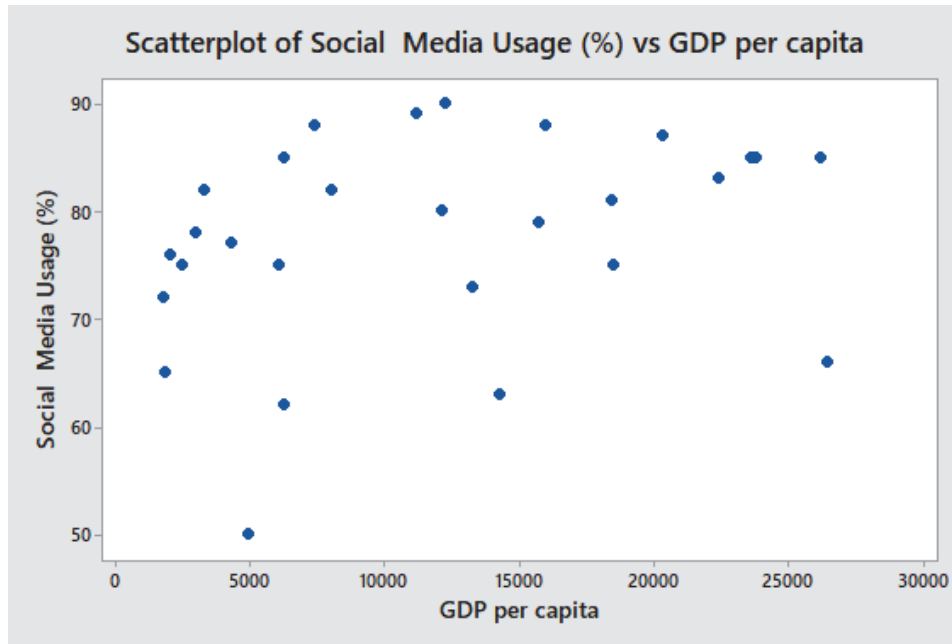
2.52 (a) There appears to be a positive relationship between the download speed and the upload speed.

(b)



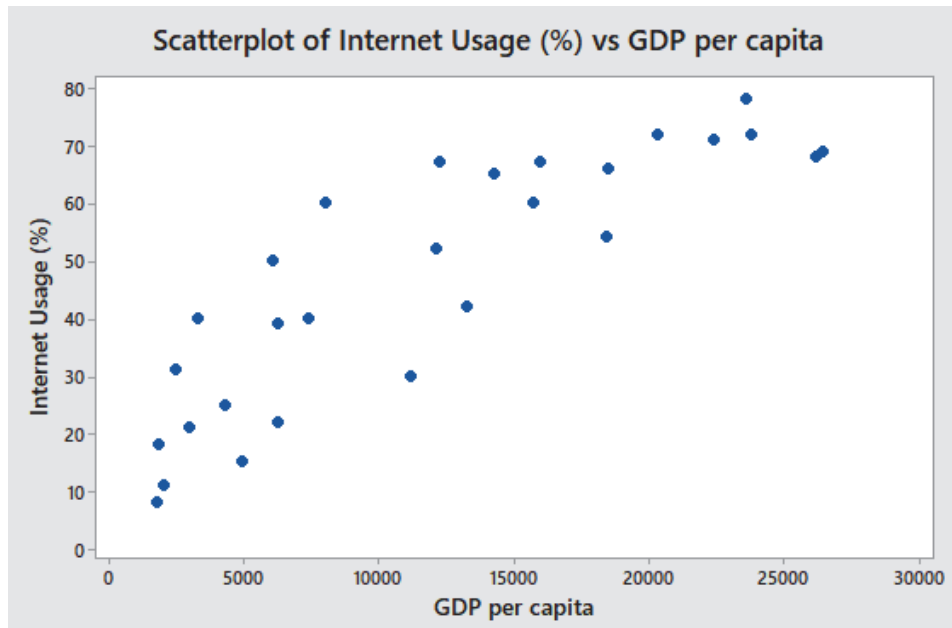
2.52 (c) Yes, this is borne out by the data
cont.

2.53 (a)



(b) There does not appear to be a relationship between GDP and social media usage.

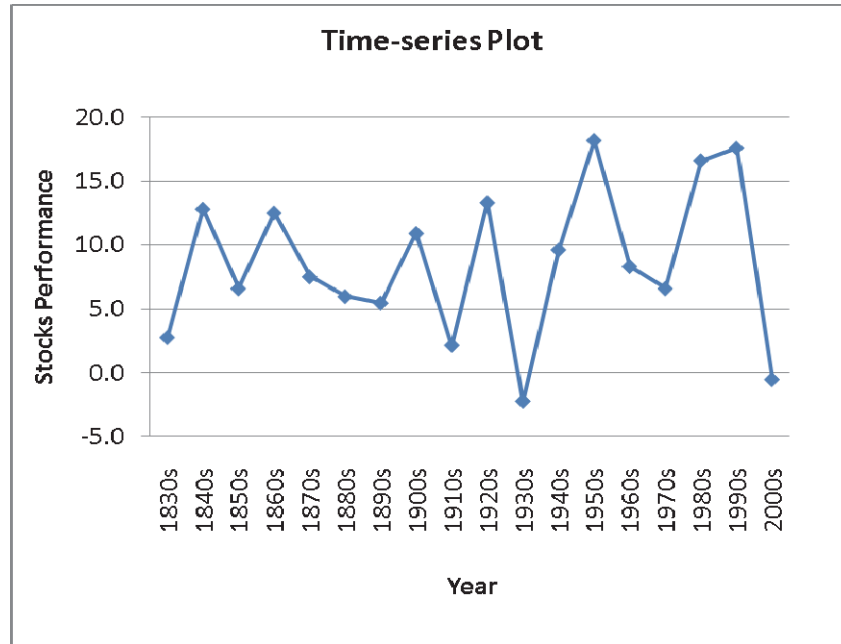
(c)



(d) There is a positive relationship between GDP and internet usage.

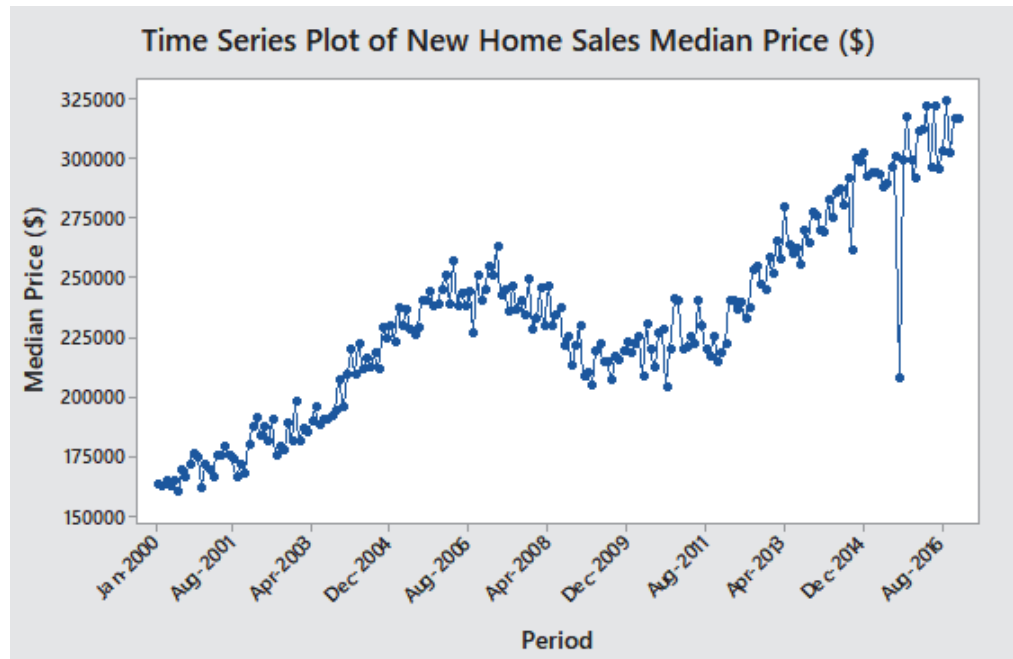
80 Chapter 2: Organizing and Visualizing Variables

2.54 (a) Excel output:



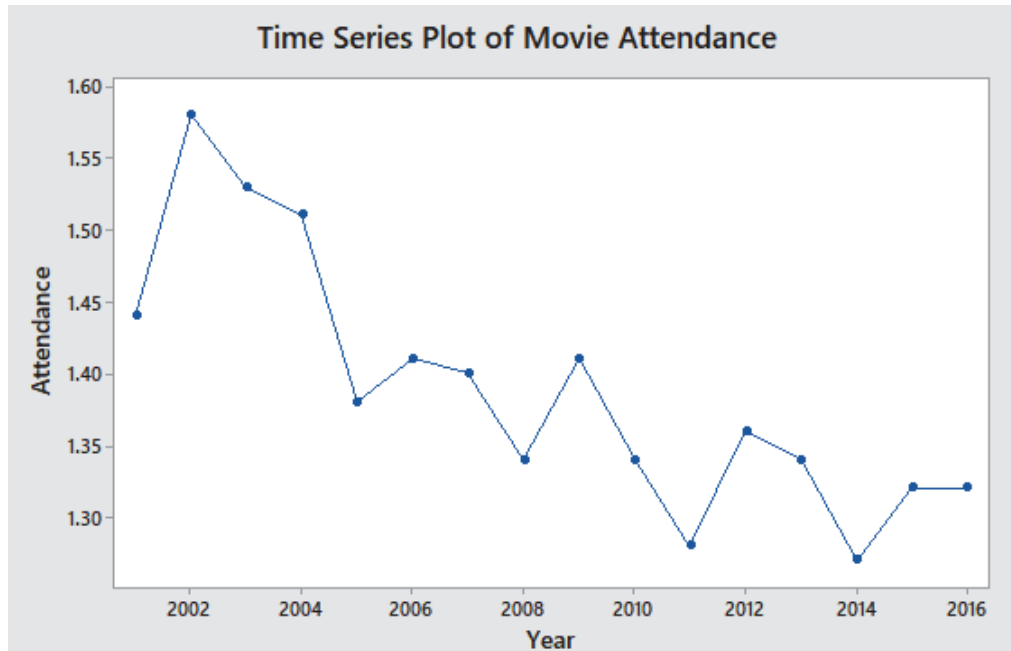
(b) There is a great deal of variation in the returns from decade to decade. Most of the returns are between 5% and 15%. The 1950s, 1980s, and 1990s had exceptionally high returns, and only the 1930s and 2000s had negative returns.

2.55 (a)



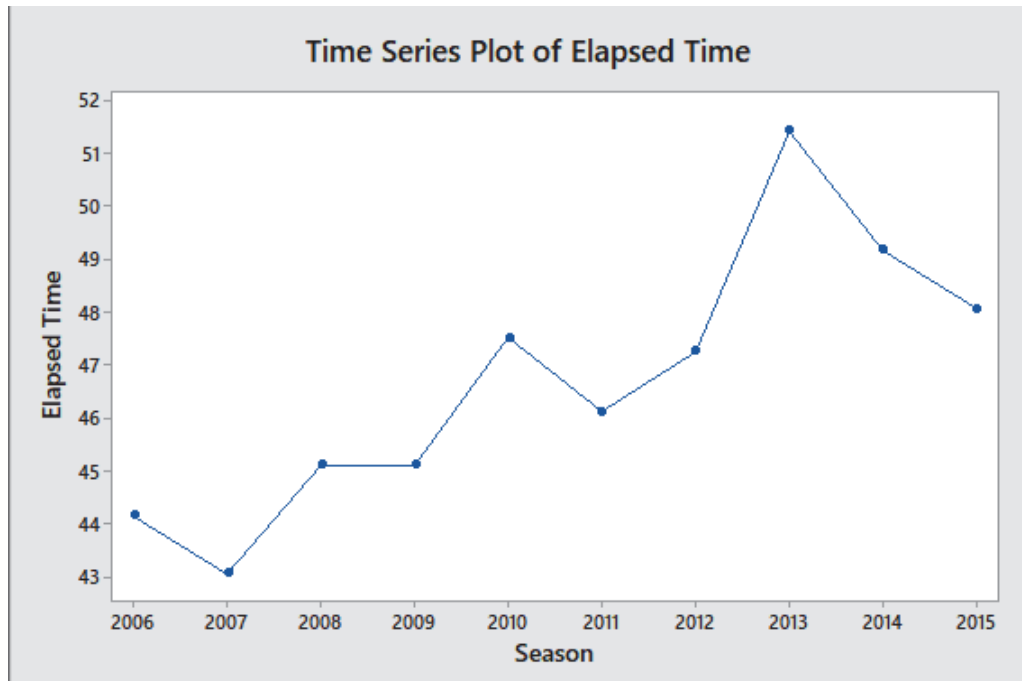
(b) There is an upward trend in home sales price until 2006. Prices decline or remain flat from 2006 – 2011. From 2011 – 2016 there is an upward trend in median price of new home sales. There is a huge drop in median prices in September 2015. This should be investigated further and may be just an error in the data file.

2.56 (a)



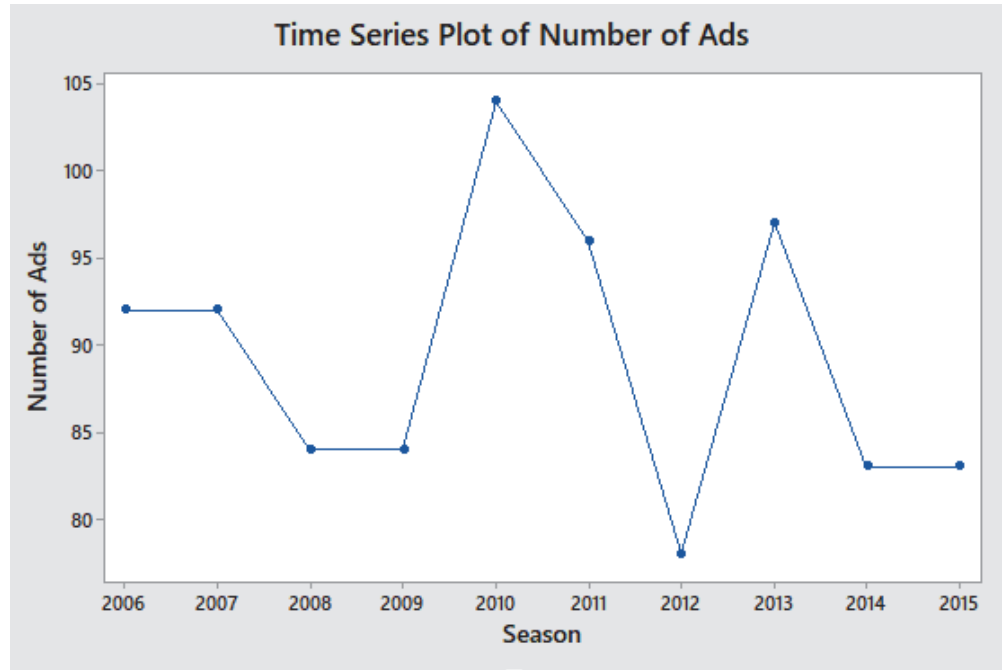
(b) There was a decline in movie attendance from 2001 to 2016. During that time, movie attendance increased from 2001 to 2002 but then by 2016 decreased to a level below 2001.

2.57 (a)



82 Chapter 2: Organizing and Visualizing Variables

2.57 (a)
cont.



- (b) There does not appear to be a pattern present in the number of Ads ran between the opening kickoff and the final whistle over the years from 2006 to 2015.
- (c) The total elapse run time (in minutes) of commercials increased from 2007 to 2013 followed by a declined in 2014 and 2015.

2.58 (a) Pivot Table in terms of %

Count of Type Type	Star Rating					Grand Total
	One	Two	Three	Four	Five	
Growth	5.43%	17.12%	27.35%	11.27%	2.71%	63.88%
Large	3.76%	7.72%	13.57%	5.43%	1.67%	32.15%
Mid-Cap	1.25%	5.43%	7.52%	3.13%	0.63%	17.96%
Small	0.42%	3.97%	6.26%	2.71%	0.42%	13.78%
Value	2.92%	10.65%	13.99%	7.31%	1.25%	36.12%
Large	2.09%	6.68%	9.19%	3.97%	1.25%	23.18%
Mid-Cap	0.63%	2.09%	2.71%	1.04%	0.00%	6.47%
Small	0.21%	1.88%	2.09%	2.30%	0.00%	6.48%
Grand Total	8.35%	27.77%	41.34%	18.58%	3.97%	100.00%

- (b) The growth and value funds have similar patterns in terms of star rating and type. Both growth and value funds have more funds with a rating of three. Very few funds have ratings of five.

2.58 (c) Pivot Table in terms of Average Three-Year Return
cont.

Count of Type Type	Star Rating					Grand Total
	One	Two	Three	Four	Five	
Growth	5.41	7.04	8.94	10.14	12.83	8.51
Large	6.97	9.43	10.62	11.83	14.25	10.30
Mid-Cap	2.27	5.07	7.93	8.77	11.22	6.93
Small	0.78	5.09	6.52	8.35	9.53	6.39
Value	4.43	5.49	7.29	8.34	10.23	6.84
Large	5.23	6.05	7.58	8.85	10.23	7.29
Mid-Cap	2.79	5.77	7.32	9.26	–	6.69
Small	1.33	3.20	5.93	7.04	–	5.39
Grand Total	5.07	6.45	8.38	9.43	12.01	7.91

(d) There are 65 large cap growth funds with a rating of three. Their average three year return is 10.62.

2.59 (a) Pivot table of tallies in terms of counts:

Count of Star Rating	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Large	14	45	109	69	28	265
Low	8	24	50	25	9	116
Average	3	20	55	40	11	129
High	3	1	4	4	8	20
MidCap	3	20	49	36	9	117
Low	2	11	7	5	1	26
Average	1	9	34	15	3	62
High			8	16	5	29
Small	2	24	40	28	3	97
Low	1	3		1		5
Average			13	15	5	33
High	1	8	25	22	3	59
Grand Total	19	89	198	133	40	479

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

Count of Star Rating	Column Labels					
Row Labels	Five	Four	Three	Two	One	Grand Total
Large	2.92%	9.39%	22.76%	14.41%	5.85%	55.32%
Low	1.67%	5.01%	10.44%	5.22%	1.88%	24.22%
Average	0.63%	4.18%	11.48%	8.35%	2.30%	26.93%
High	0.63%	0.21%	0.84%	0.84%	1.67%	4.18%
MidCap	0.63%	4.18%	10.23%	7.52%	1.88%	24.43%
Low	0.42%	2.30%	1.46%	1.04%	0.21%	5.43%
Average	0.21%	1.88%	7.10%	3.13%	0.63%	12.94%
High	0.00%	0.00%	1.67%	3.34%	1.04%	6.05%
Small	0.42%	5.01%	8.35%	5.85%	0.63%	20.25%
Low	0.21%	0.63%	0.00%	0.21%	0.00%	1.04%
Average	0.00%	2.71%	3.13%	1.04%	0.00%	6.89%
High	0.21%	1.67%	5.22%	4.59%	0.63%	12.32%
Grand Total	3.97%	18.58%	41.34%	27.77%	8.35%	100.00%

84 Chapter 2: Organizing and Visualizing Variables

2.59 (b) For the large-cap funds, the three-star rating category had the highest percentage of cont. funds, followed by two-star, four-star, one-star, and five-star. Very few large-cap funds had ratings of five. This pattern was also seen with the mid-cap funds as a group. The same pattern was observed with the small-cap funds. However, the pattern was more subtle in that the differences in percentage were less in many cases.

Within the large-cap fund category, the highest percentage of funds were in the average-risk category followed by the low-risk and high-risk categories. Within the mid-cap category, the highest percentage of funds were in the average-risk category followed by the high and low risk categories. Within the small-cap category, the highest percentage of funds were in the high-risk category followed by the average and low risk categories.

(c)

Average of 3YrReturn	Column Labels						
Row Labels	Five	Four	Three	Two	One	Grand Total	
Large	12.53	10.57	9.39	7.86	6.35	9.04	
Low	12.36	9.91	8.57	7.59	6.94	8.77	
Average	10.73	11.35	10.00	7.75	7.03	9.27	
High	14.80	10.92	11.36	10.68	4.76	9.07	
MidCap	11.22	8.89	7.77	5.27	2.44	6.87	
Low	11.74	9.02	8.48	7.13	3.90	8.52	
Average	10.18	8.73	7.40	5.74	0.72	6.91	
High			8.73	4.24	3.18	5.29	
Small	9.53	7.75	6.38	4.48	0.96	6.07	
Low	9.09	5.98		7.60		6.92	
Average		7.73	6.61	2.85		6.48	
High	9.96	8.44	6.24	4.71	0.96	5.76	
Grand Total	12.01	9.43	8.38	6.45	5.07	7.91	

(d) There are four high-risk large-cap funds with a three-star rating. Their average three-year return is 11.36.

2.60

Count of Type Type	Star Rating					Grand Total
	One	Two	Three	Four	Five	
Growth	5.43%	17.12%	27.35%	11.27%	2.71%	63.88%
Large	1.25%	2.09%	4.80%	3.55%	1.46%	13.15%
Mid-Cap	1.67%	7.72%	15.87%	6.05%	0.42%	31.73%
Small	2.51%	7.31%	6.68%	1.67%	0.84%	19.00%
Value	2.92%	10.65%	13.99%	7.31%	1.25%	36.12%
Large	0.84%	4.38%	7.10%	4.38%	0.84%	17.54%
Mid-Cap	1.25%	4.80%	5.85%	2.71%	0.42%	15.03%
Small	0.84%	1.46%	1.04%	0.21%	0.00%	3.55%
Grand Total	8.35%	27.77%	41.34%	18.58%	3.96%	100.00%

(b) Patterns of star rating conditioned on risk:
For the growth funds as a group, most are rated as three-star, followed by two-star, four-star, one-star, and five-star. The pattern of star rating is different among the various risk growth funds.

For the value funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. Among the high-risk value funds, more are two-star than three-star.

Most of the growth funds are rated as average-risk, followed by high-risk and then low-risk. The pattern is not the same among all the rating categories.

2.60 (b) cont. 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 three-star, four-star, and five-star value funds. Among the one-star and two-star funds, there are more average risk funds than low risk funds.

(c)

Count of Type Type	Star Rating					Grand Total
	One	Two	Three	Four	Five	
Growth	5.41	7.04	8.94	10.14	12.83	8.51
Large	7.53	8.60	9.89	10.29	12.64	9.87
Mid-Cap	6.17	7.99	9.28	10.43	11.96	9.06
Small	3.83	5.59	7.45	8.76	13.59	6.64
Value	4.43	5.49	7.29	8.34	10.23	6.84
Large	5.29	7.00	7.66	8.57	10.74	7.76
Mid-Cap	5.01	4.98	6.97	7.96	9.23	6.41
Small	2.71	2.63	6.53	8.39	–	4.13
Grand Total	5.07	6.45	8.38	9.43	12.01	7.91

The three-year returns for growth funds is higher than for value funds. The return is higher for funds with higher ratings than lower ratings. This pattern holds for the growth funds for each risk level. For the low risk and average risk value funds, the return is lowest for the funds with a two-star rating.

(d) There are 32 growth funds with high risk with a rating of three. These funds have an average three-year return of 7.45.

2.61 (a) Pivot table of tallies in terms of counts:

Row Labels	Five	Four	Three	Two	One	Grand Total
Growth	13	54	131	82	26	306
Large	8	26	65	37	18	154
High	3	1	3	4	6	17
Average	1	16	43	25	6	91
Low	4	9	19	8	6	46
MidCap	3	15	36	26	6	86
High			8	13	4	25
Average	1	7	24	11	2	45
Low	2	8	4	2		16
Small	2	13	30	19	2	66
High	1	7	21	18	2	49
Average		6	9	1		16
Low	1					1
Value	6	35	67	51	14	173
Large	6	19	44	32	10	111
High			1		2	3
Average	2	4	12	15	5	38
Low	4	15	31	17	3	70
MidCap		5	13	10	3	31
High				3	1	4
Average		2	10	4	1	17
Low		3	3	3	1	10
Small		11	10	9	1	31
High		1	4	4	1	10
Average		7	6	4		17
Low		3		1		4
Grand Total	19	89	198	133	40	479

86 Chapter 2: Organizing and Visualizing Variables

2.61 (a) Pivot table of tallies in terms of % of grand total:
cont.

Count of Star Rating	Column Labels						
Row Labels	Five	Four	Three	Two	One	Grand Total	
Growth	2.71%	11.27%	27.35%	17.12%	5.43%	63.88%	
Large	1.67%	5.43%	13.57%	7.72%	3.76%	32.15%	
High	0.63%	0.21%	0.63%	0.84%	1.25%	3.55%	
Average	0.21%	3.34%	8.98%	5.22%	1.25%	19.00%	
Low	0.84%	1.88%	3.97%	1.67%	1.25%	9.60%	
MidCap	0.63%	3.13%	7.52%	5.43%	1.25%	17.95%	
High	0.00%	0.00%	1.67%	2.71%	0.84%	5.22%	
Average	0.21%	1.46%	5.01%	2.30%	0.42%	9.39%	
Low	0.42%	1.67%	0.84%	0.42%	0.00%	3.34%	
Small	0.42%	2.71%	6.26%	3.97%	0.42%	13.78%	
High	0.21%	1.46%	4.38%	3.76%	0.42%	10.23%	
Average	0.00%	1.25%	1.88%	0.21%	0.00%	3.34%	
Low	0.21%	0.00%	0.00%	0.00%	0.00%	0.21%	
Value	1.25%	7.31%	13.99%	10.65%	2.92%	36.12%	
Large	1.25%	3.97%	9.19%	6.68%	2.09%	23.17%	
High	0.00%	0.00%	0.21%	0.00%	0.42%	0.63%	
Average	0.42%	0.84%	2.51%	3.13%	1.04%	7.93%	
Low	0.84%	3.13%	6.47%	3.55%	0.63%	14.61%	
MidCap	0.00%	1.04%	2.71%	2.09%	0.63%	6.47%	
High	0.00%	0.00%	0.00%	0.63%	0.21%	0.84%	
Average	0.00%	0.42%	2.09%	0.84%	0.21%	3.55%	
Low	0.00%	0.63%	0.63%	0.63%	0.21%	2.09%	
Small	0.00%	2.30%	2.09%	1.88%	0.21%	6.47%	
High	0.00%	0.21%	0.84%	0.84%	0.21%	2.09%	
Average	0.00%	1.46%	1.25%	0.84%	0.00%	3.55%	
Low	0.00%	0.63%	0.00%	0.21%	0.00%	0.84%	
Grand Total	3.97%	18.58%	41.34%	27.77%	8.35%	100.00%	

(b) For growth funds, most are rated as three-star followed by two-star, four-star, one-star and five-star. Among the growth funds, large-cap and mid-cap had the same pattern of star-rating as observed for growth funds in general. Small-cap growth funds had the same pattern with the exception of having the same the number of funds rated as one-star and five-star. The pattern of star-rating is different among the various risk levels within the large-cap, mid-cap and small-cap growth funds.

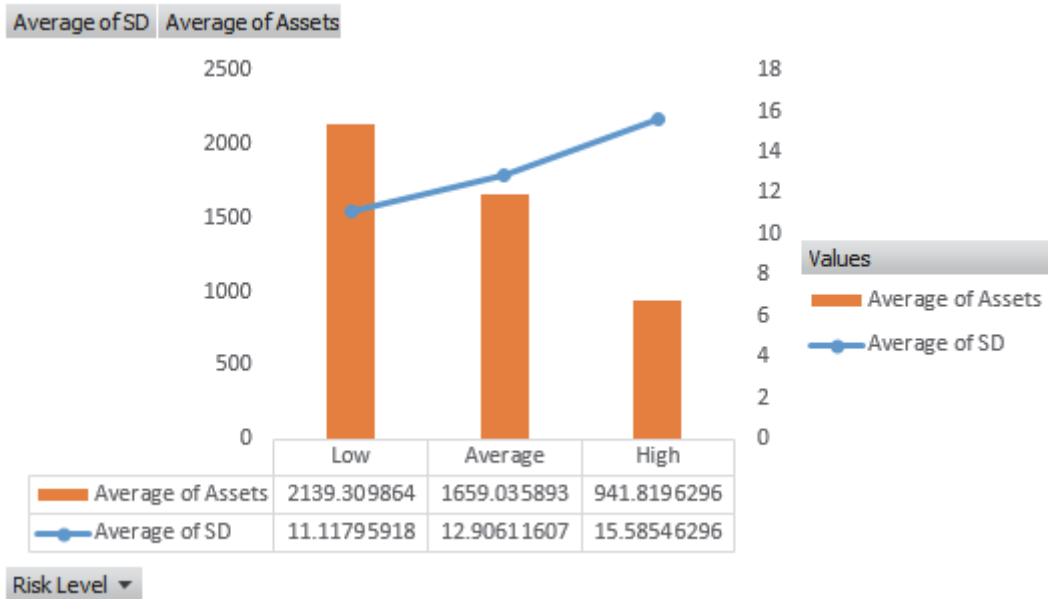
For value funds, most are rated as three-star followed by two-star, four-star, one-star, and five-star. Among the value funds, the pattern is the same for large-cap and mid-cap funds. Small-cap value funds have a different pattern. The pattern of star-rating is different among the various risk levels within the large-cap, mid-cap and small-cap funds.

(c) The tables in 2.58 through 2.60 are easier to interpret because they contain fewer fields. The table in 2.61 tallies star rating across three fields: market type, market cap, and risk level. Problems 2.58 through 2.60 tally star rating across two fields.

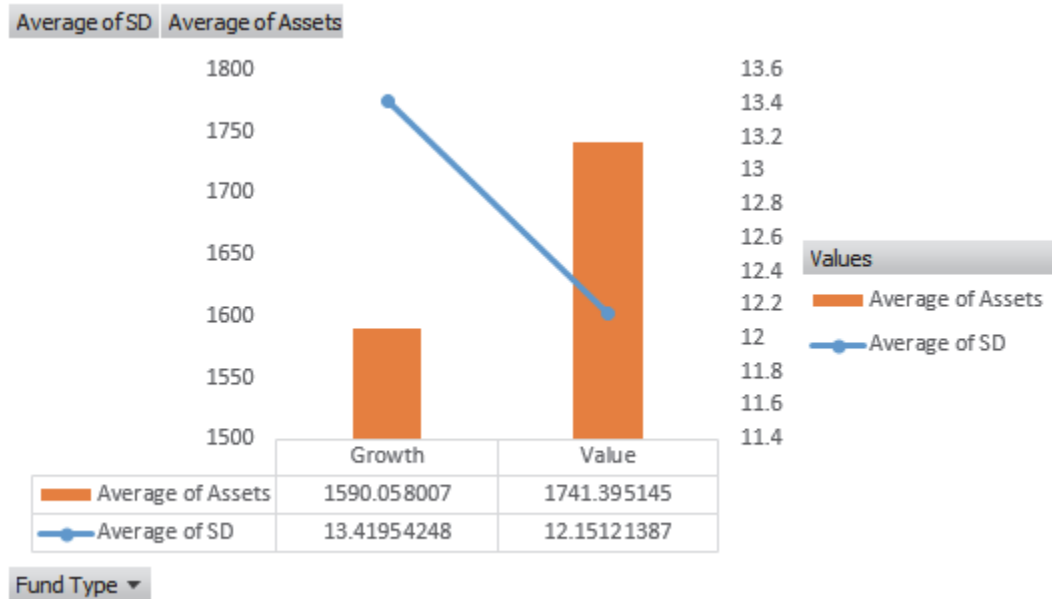
(d) Problem 2.60 reveals that most value funds are rated as low-risk followed by average-risk and high-risk. Problem 2.61 reveals that this is only the case among large-cap value funds. Most mid-cap value funds are rated as average-risk followed by low-risk and high-risk. Most small-cap value funds are rated as average-risk followed by high-risk and low-risk. Problem 2.61 also reveals that among small-cap funds rated as average-risk, most are rated as four-star, followed by three-star and two-star. Because Problem 2.61 includes four fields compared to three fields included in problems 2.58 through 2.60, additional patterns can be observed.

2.62 The fund with the highest five-year return of 15.72 is a large cap growth fund that has a four-star rating and low risk.

2.63 (a)



(b)

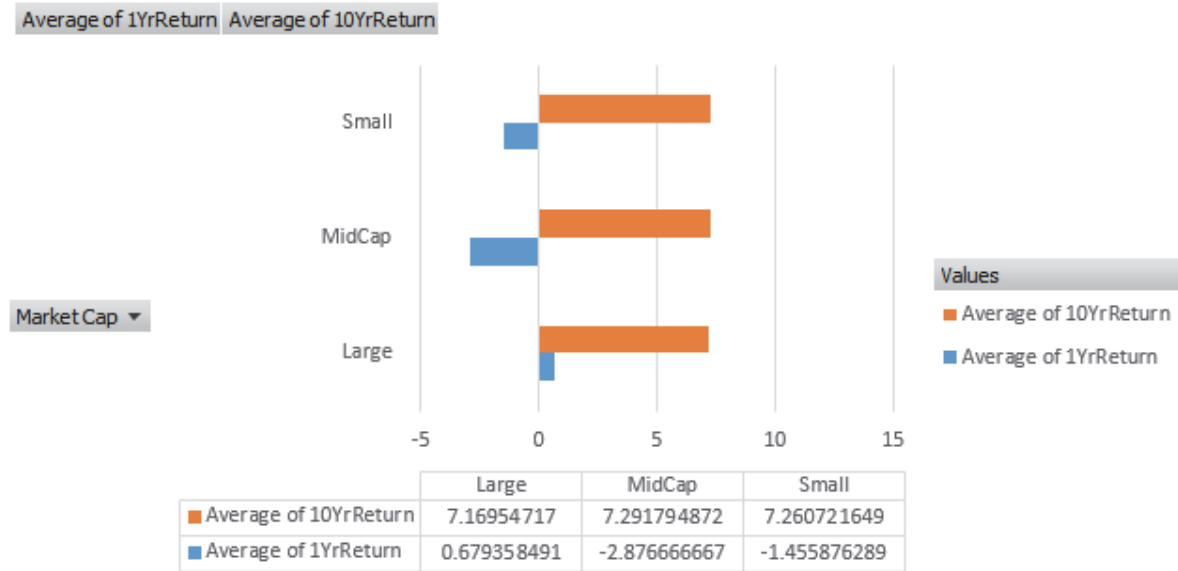


(c) The results from (a) reveal that the average of SD increases as the risk level increases while average of assets decreases as risk level increases. The results from (b) reveal that the average of SD is higher for growth funds compared to value funds. The patterns suggest that value funds are likely to be associated with less risk because the average of SD was lower among value funds and low risk funds.

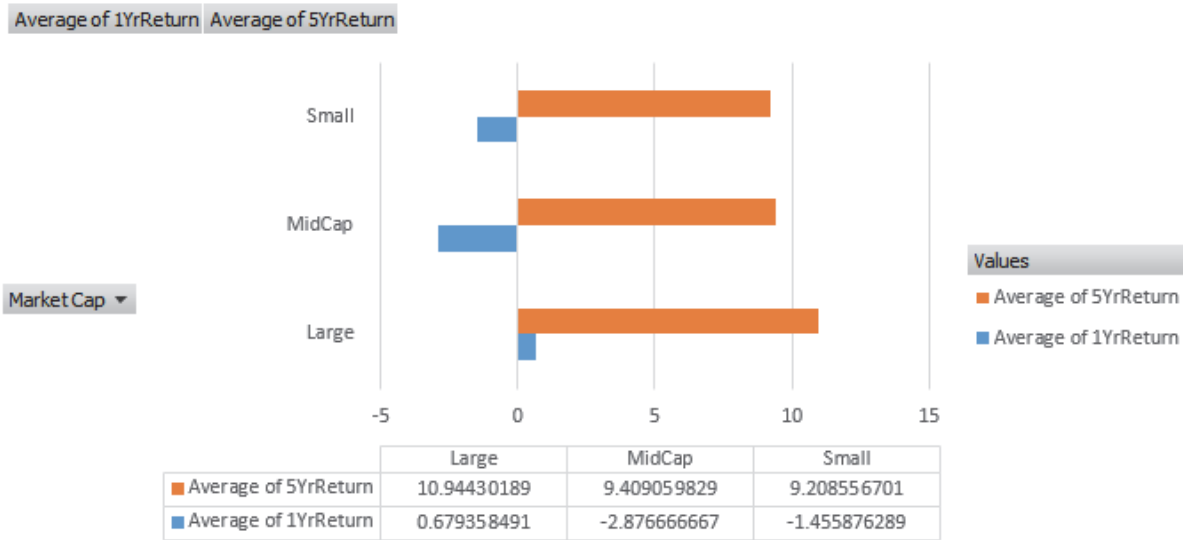
2.64 Funds 479, 471, 347, 443, and 477 have the lowest five-year return.

88 Chapter 2: Organizing and Visualizing Variables

2.65 (a)



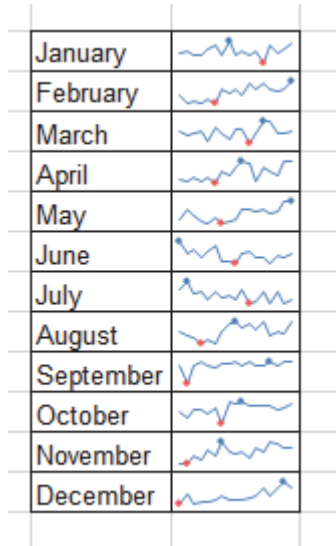
(b)



- (c) For the 1-year versus 10-year return chart, the 10-year returns are much higher than the 1-year returns with similar 5-year returns near 7 percent for all three market cap categories. For the 1-year versus 5-year chart, the returns are all higher for the 5-year returns compared to the 1-year returns. The 5-year returns are higher than the 10-year returns. The large-cap 5-year return is higher than the mid-cap and small 5-year returns.
- (d) Because the average 5-year returns were all higher than the 10-year returns for all market cap categories, one can conclude that the returns were lower in years 6 through 10. Without annual data, one cannot conclude that this was due to consistent lower returns across the years or the result of one or two years with lower returns.

2.66 The five funds with the lowest five-year return have (1) midcap growth, average risk, one-star rating, (2) midcap growth, high risk, two-star rating, (3) large value, average risk, two-star rating, (4) midcap growth, high risk, one-star rating, and (5) small value, average risk, two-star rating.

2.67 (a)



(b) The sparklines reveal that a general trend upward in home prices during the months of February, May, November, and December and they have remained steady in September after a jump from a low in 2001.

(c) In the Time-series plot one can see an upward trend in home sales price until 2006. Prices decline or remain flat from 2006 – 2011. From 2011 – 2016 there is an upward trend in median price of new home sales. With the exception of one year, the September home prices are fairly stable. This could be an error in the data.

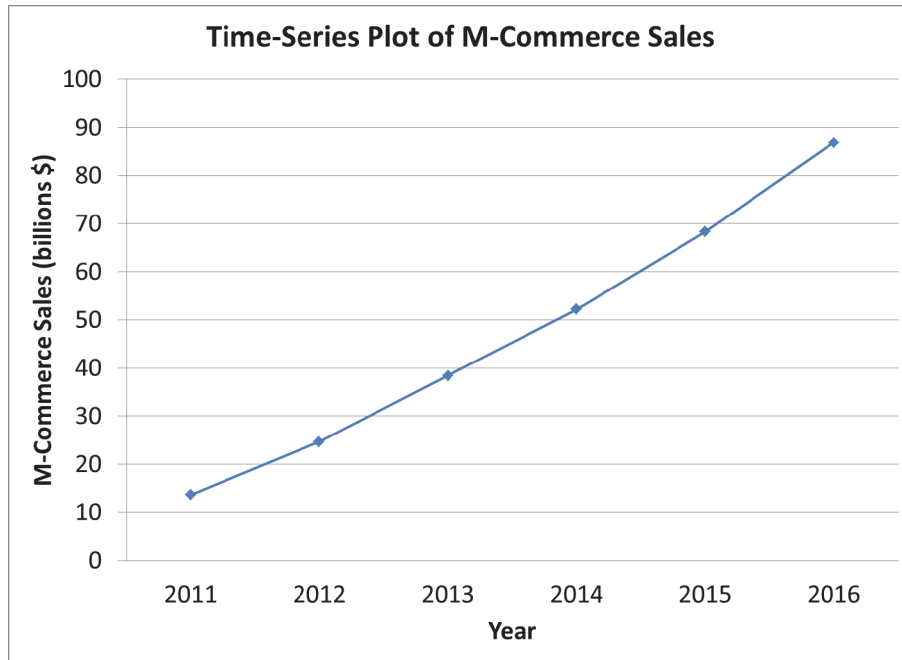
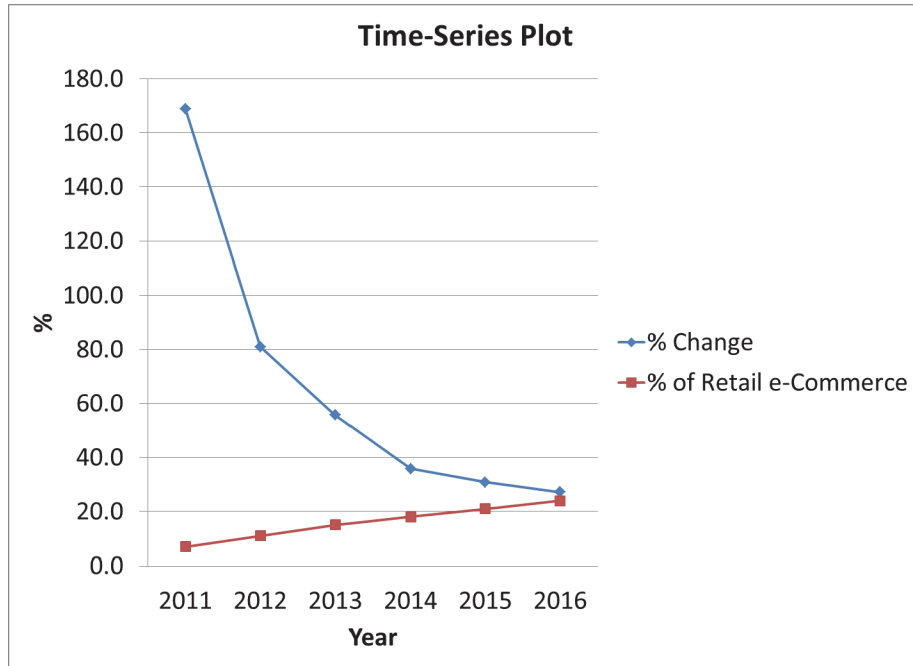
2.68 There has been a decline in the price of natural gas over time. However, there is no pattern within the years. For some years, the price is higher in the beginning of the year. For other years, the price is higher in the latter part of the year. Sometimes, there is little variation within the year.

2.69 Student project answers will vary

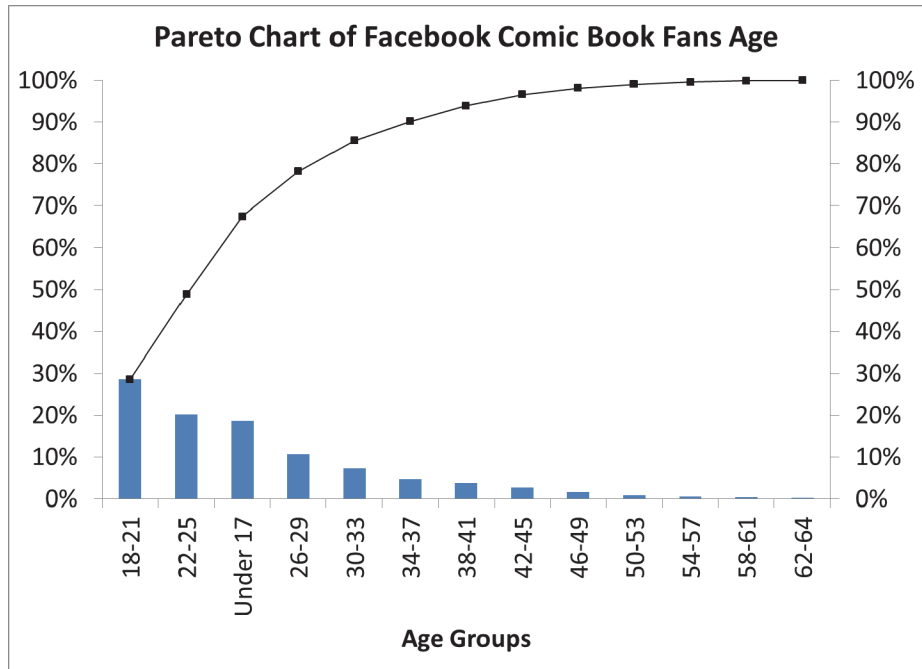
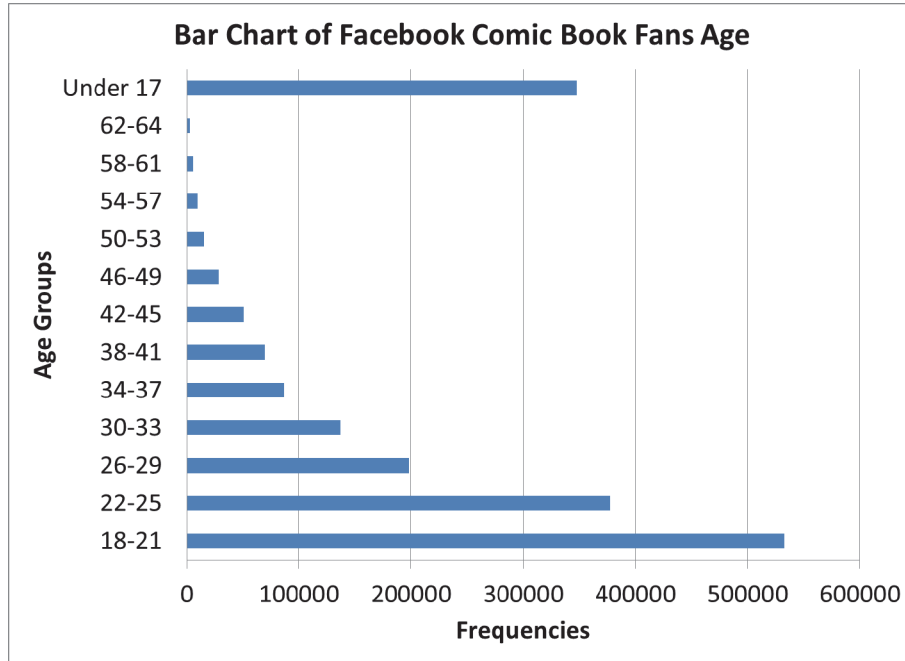
2.70 Student project answers will vary

90 Chapter 2: Organizing and Visualizing Variables

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

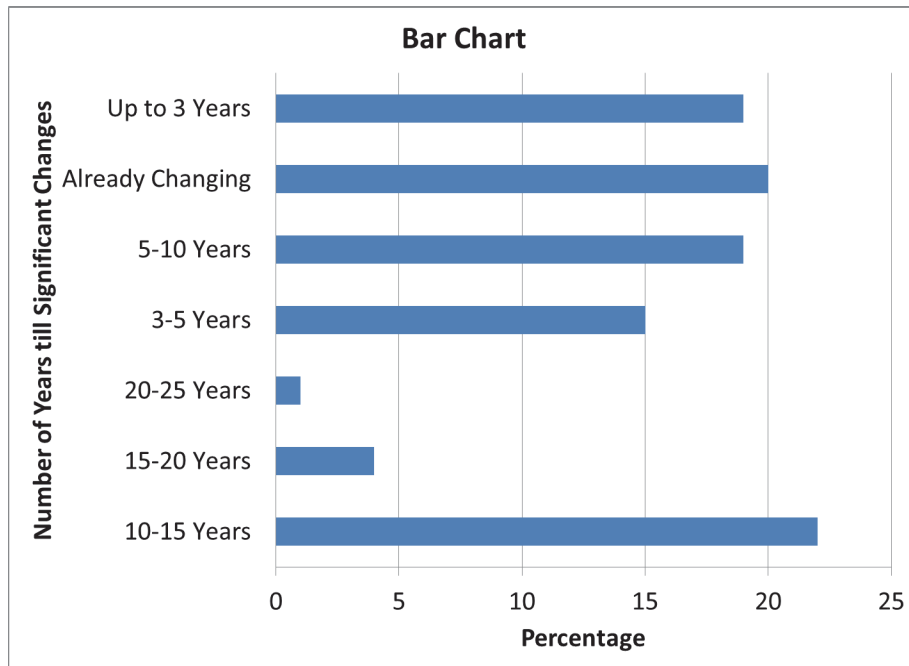
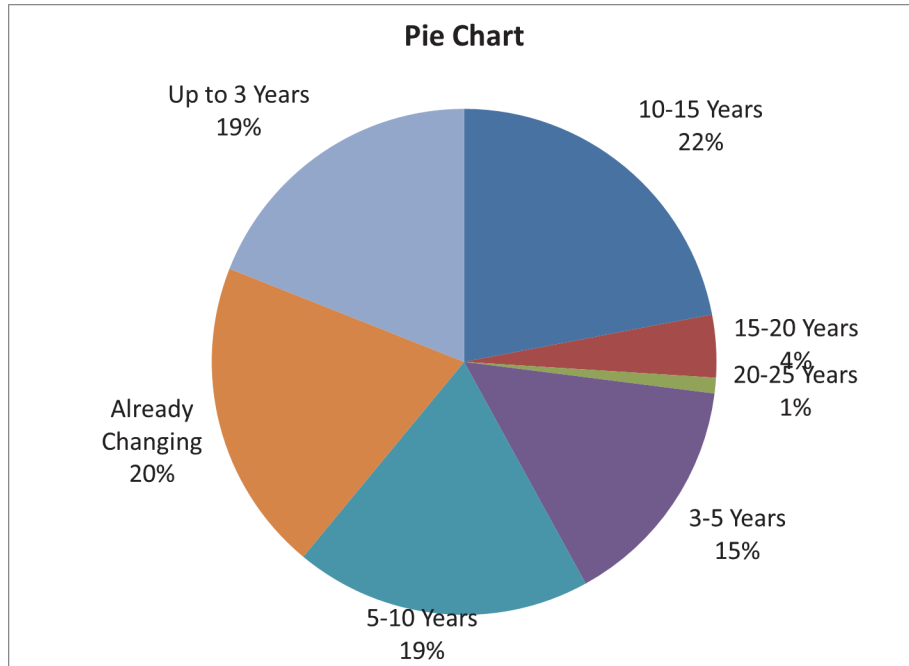


- 2.72 (a) There is a title.
 (b) The simplest possible visualization is not used.
 (c)



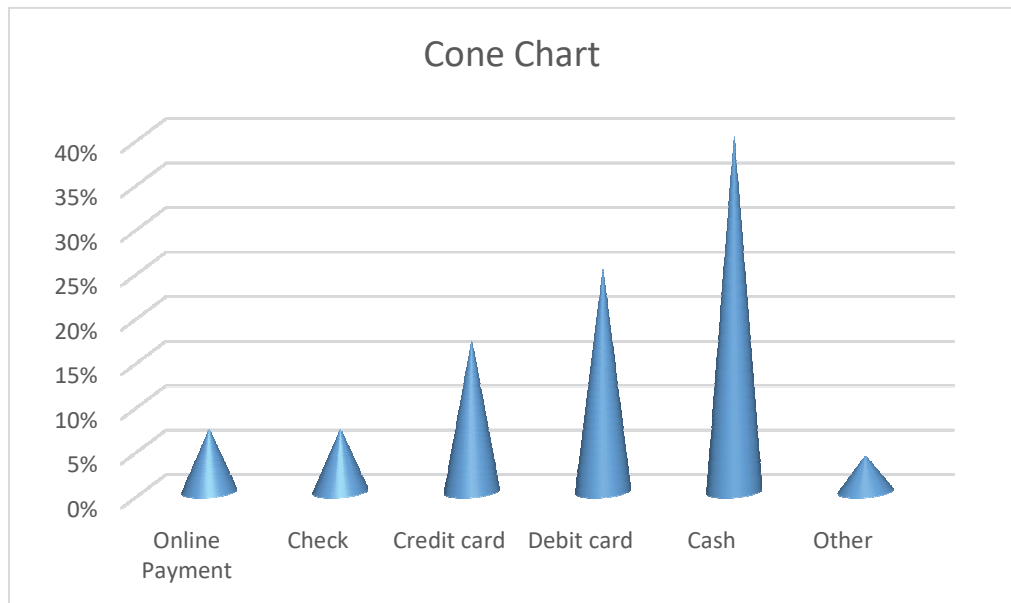
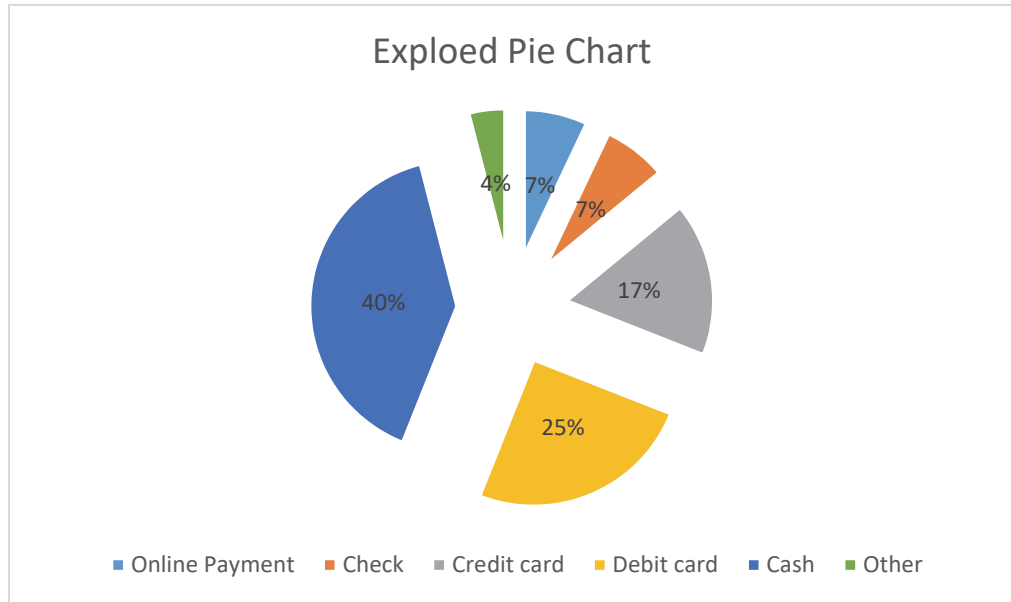
92 Chapter 2: Organizing and Visualizing Variables

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

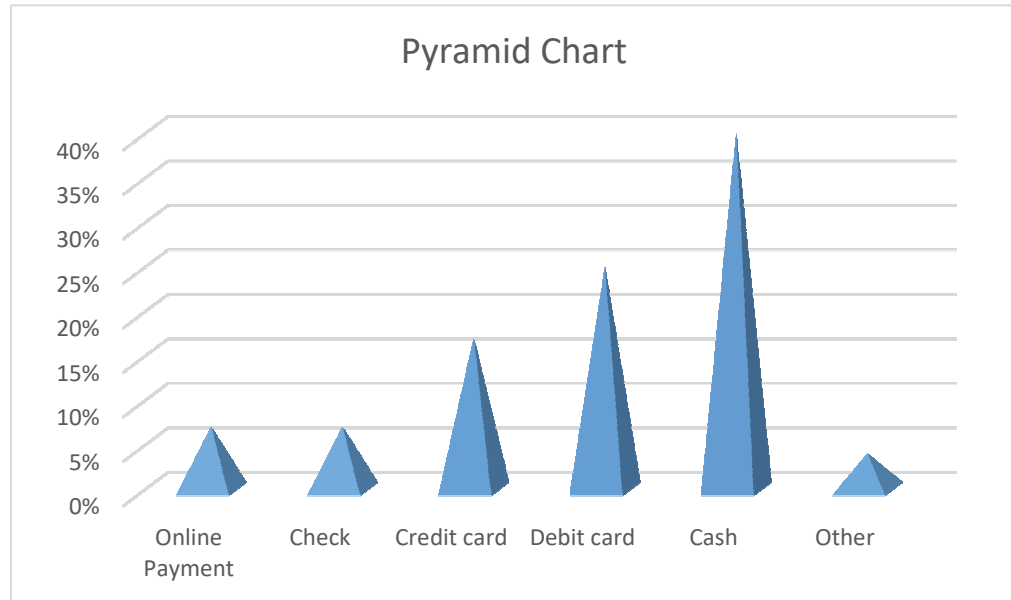


- 2.74 Answers will vary depending on selection of source.

2.75 (a)

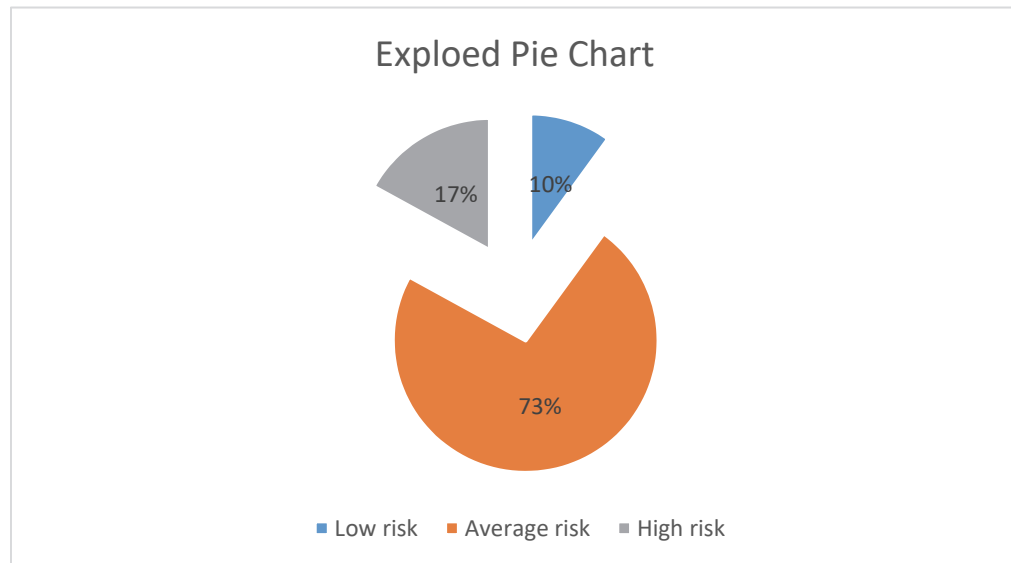


2.75 (a)
cont.

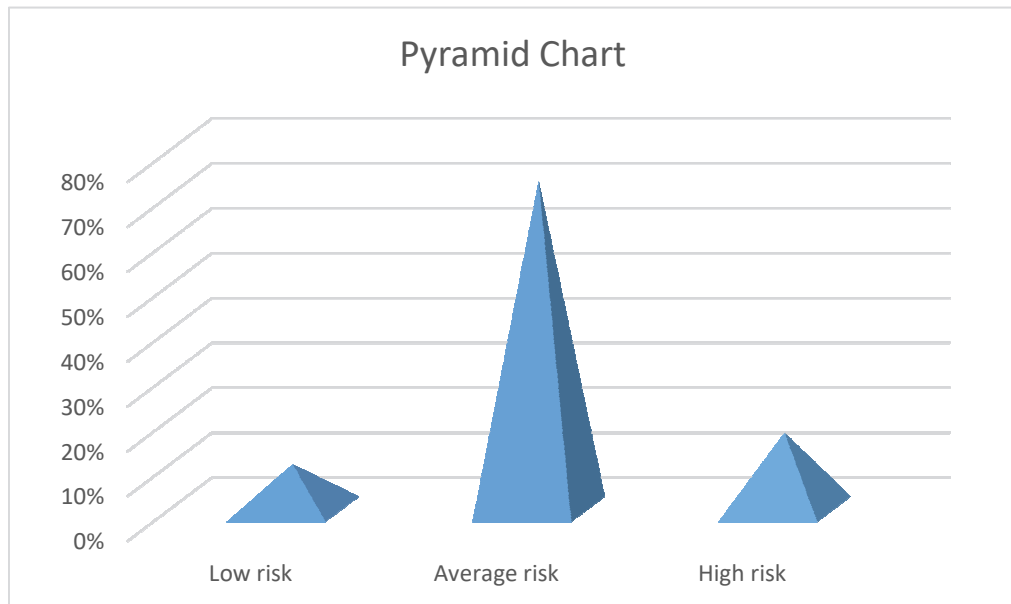
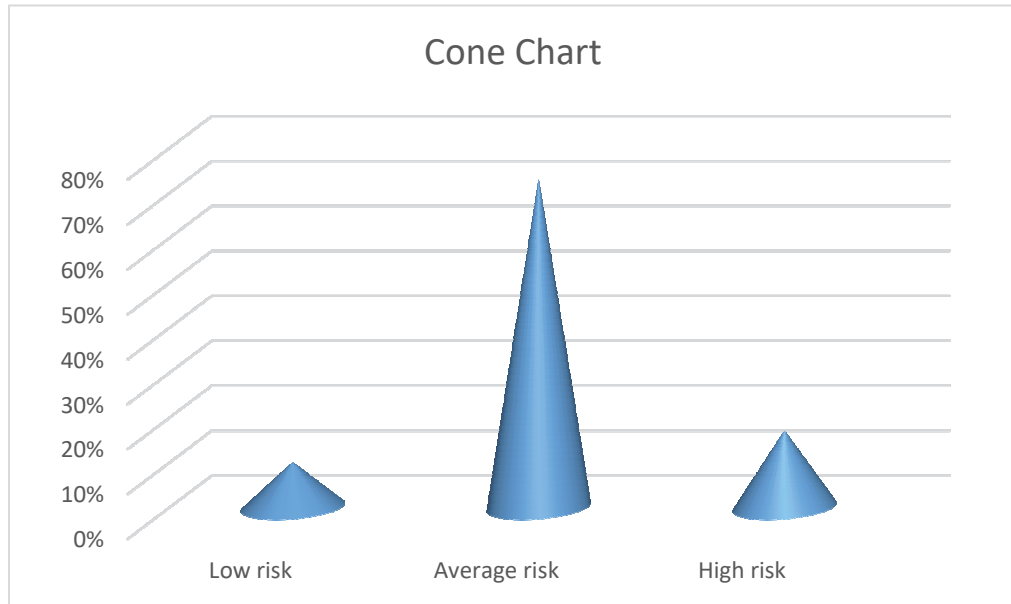


(b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.76 (a)



2.76 (a)
cont.



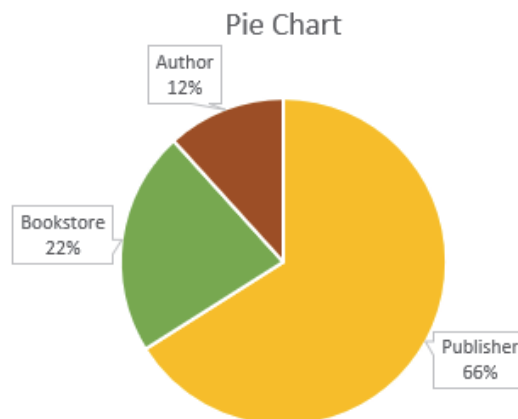
(b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.77 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.

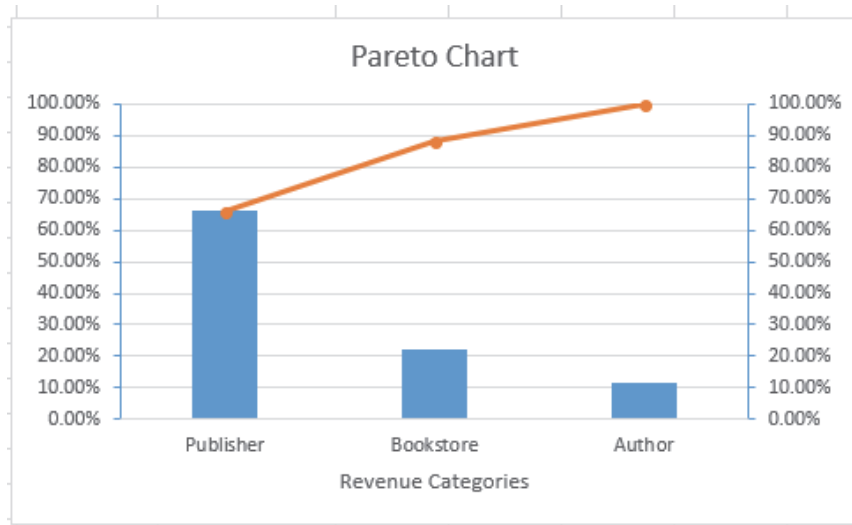
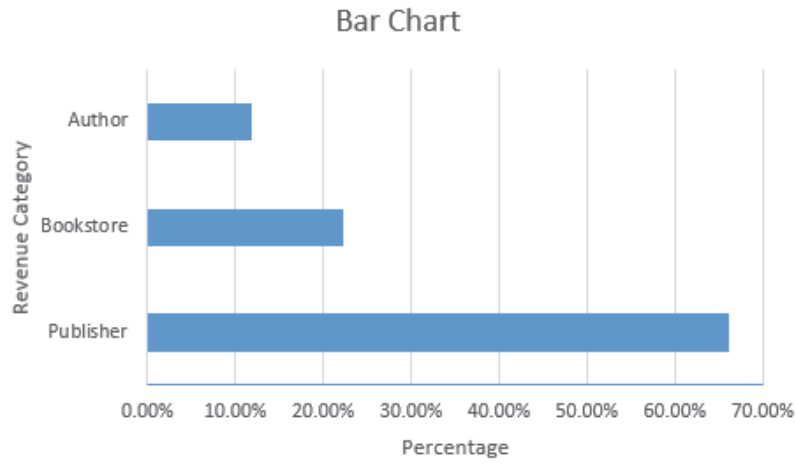
2.78 A summary table allows one to determine the frequency or percentage of occurrences in each category.

96 Chapter 2: Organizing and Visualizing Variables

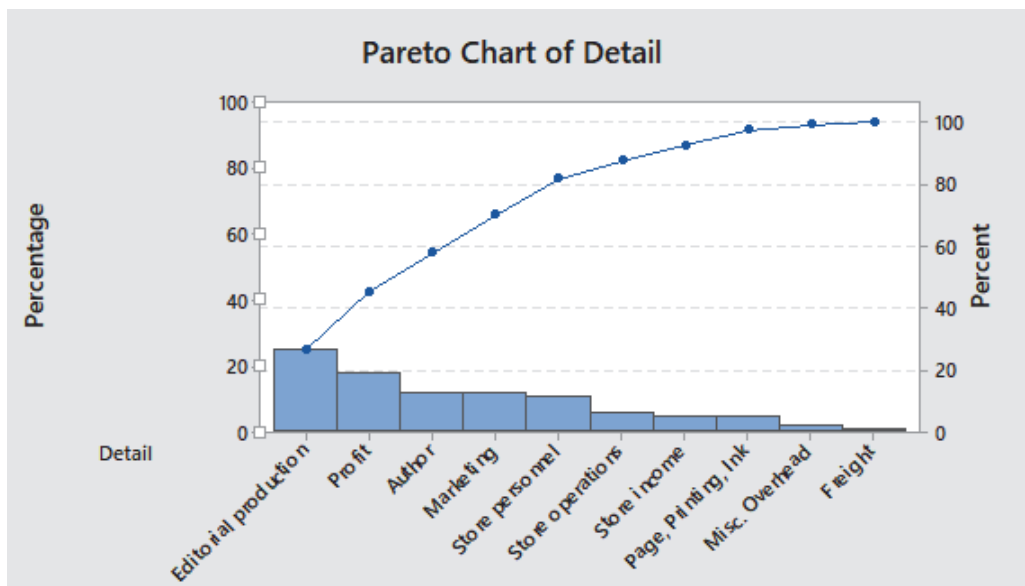
- 2.79 A bar chart is useful for comparing categories. A pie chart is useful when examining the portion of the whole that is in each category. A Pareto diagram is useful in focusing on the categories that make up most of the frequencies or percentages.
- 2.80 The bar chart for categorical data is plotted 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.
- 2.81 A time-series plot is a type of scatter diagram with time on the x-axis.
- 2.82 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.
- 2.83 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.
- 2.84 A contingency table contains information on two categorical variables whereas a multidimensional table can display information on more than two categorical variables.
- 2.85 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.
- 2.86 In a PivotTable in Excel, double-clicking a cell drills down and causes Excel to display the underlying data in a new worksheet to enable you to then observe the data for patterns. In Excel, a slicer is a panel of clickable buttons that appears superimposed over a worksheet to enable you to work with many variables at once in a way that avoids creating an overly complex multidimensional contingency table that would be hard to comprehend and interpret.
- 2.87 Sparklines are compact time-series visualizations of numerical variables. Sparklines can also be used to plot time-series data using smaller time units than a time-series plot to reveal patterns that the time-series plot may not.
- 2.88 (a)



2.88 (a)
cont.



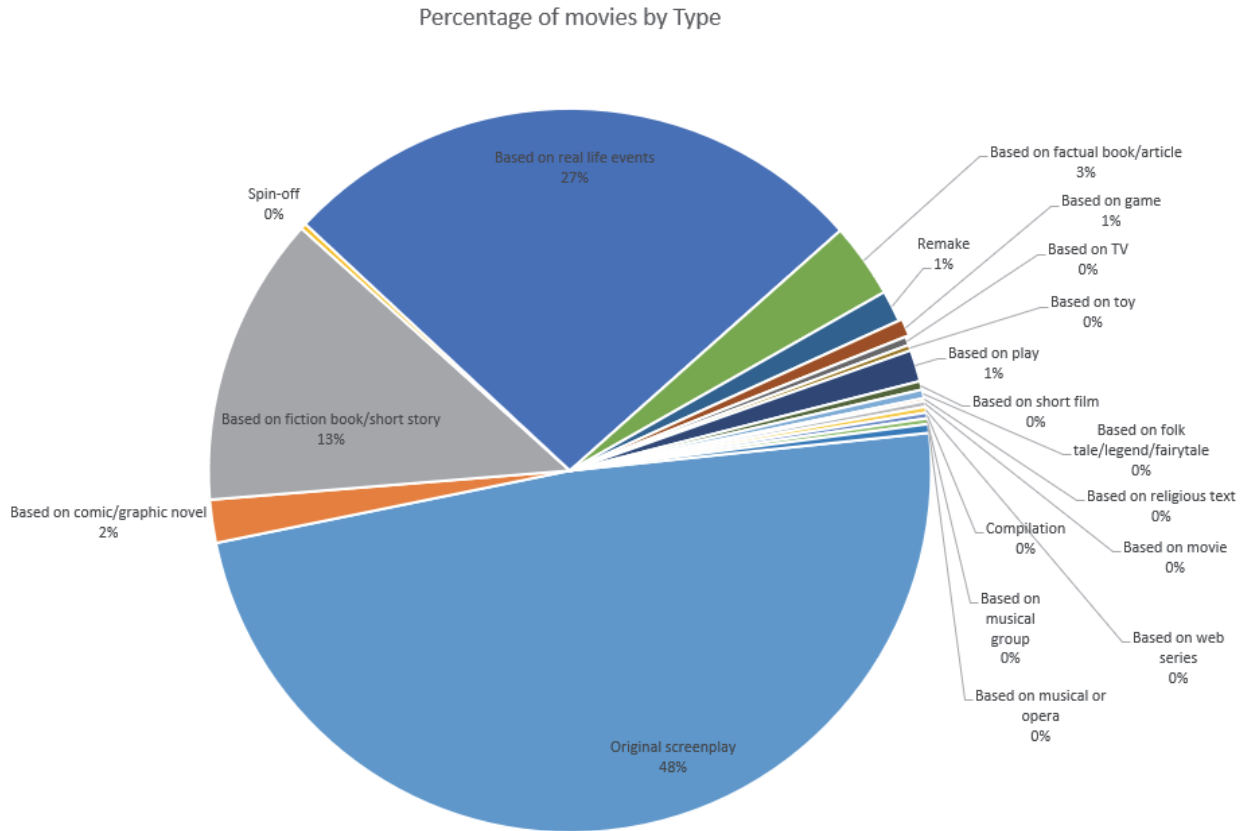
(b)



98 Chapter 2: Organizing and Visualizing Variables

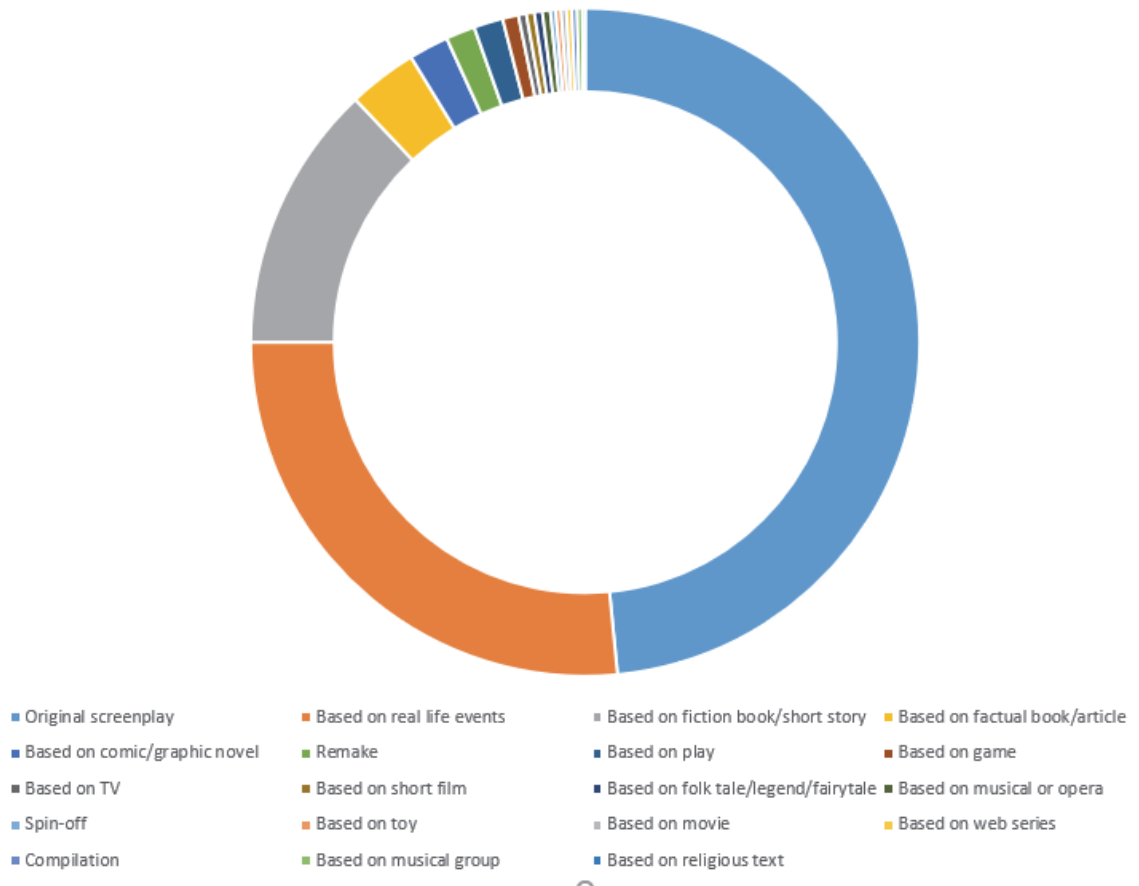
2.88 (c) cont. The publisher gets the largest portion (66.06%) of the revenue. 24.93% is editorial production manufacturing costs. The publisher’s marketing accounts for the next largest share of the revenue, at 11.6%. Author and bookstore personnel each account for around 11 to 12% of the revenue, whereas the publisher and bookstore profit and income account for more than 26% of the revenue. Yes, the bookstore gets almost twice the revenue of the authors.

2.89 (a) Number of Movies

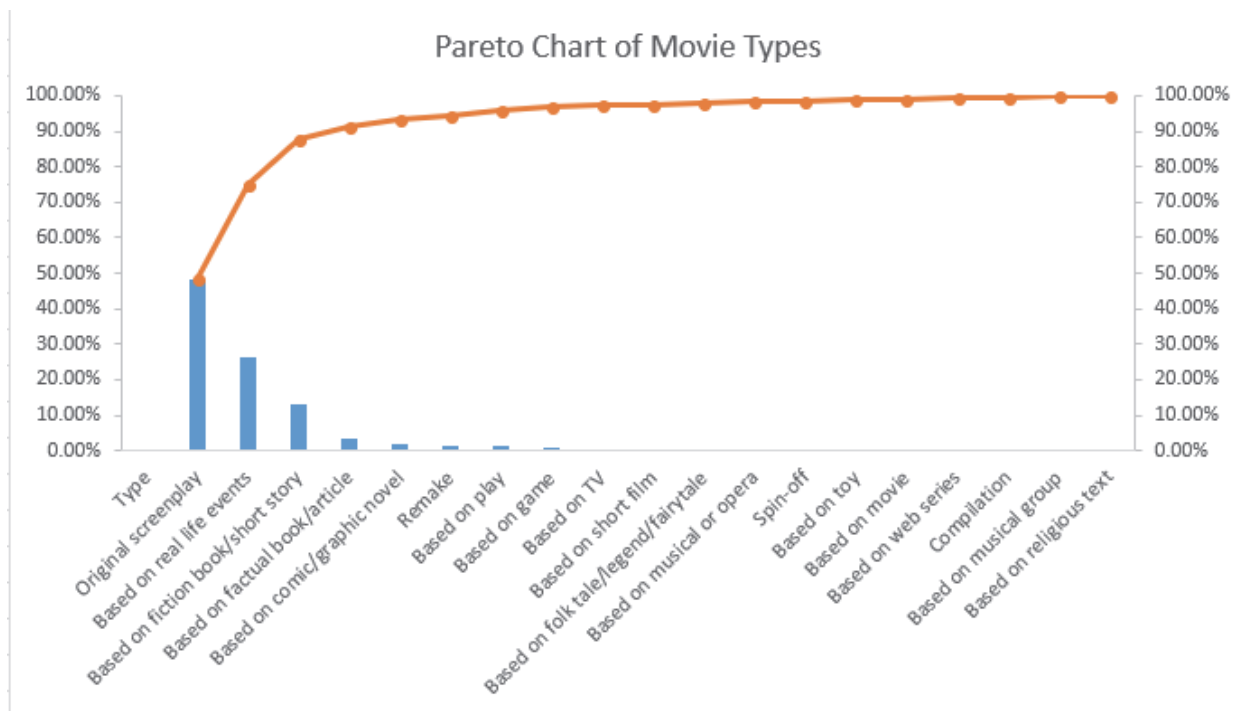
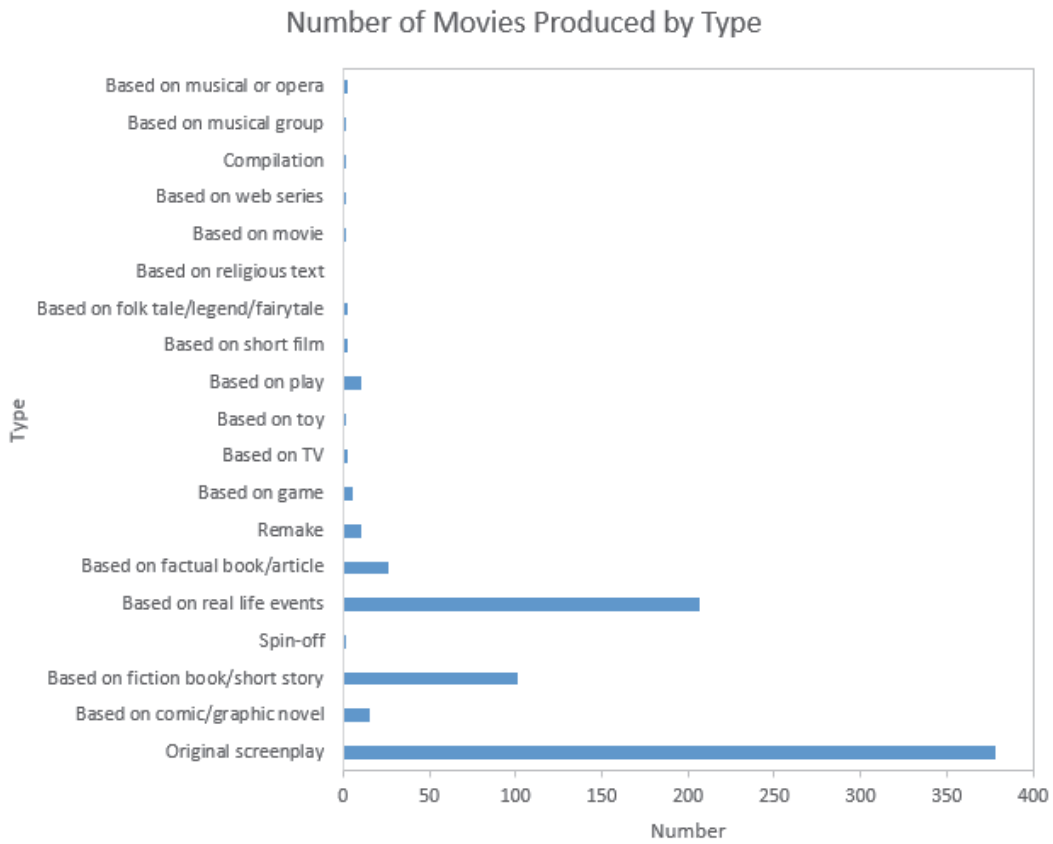


2.89 (a)
cont.

Doughnut Chart of Movies by Type

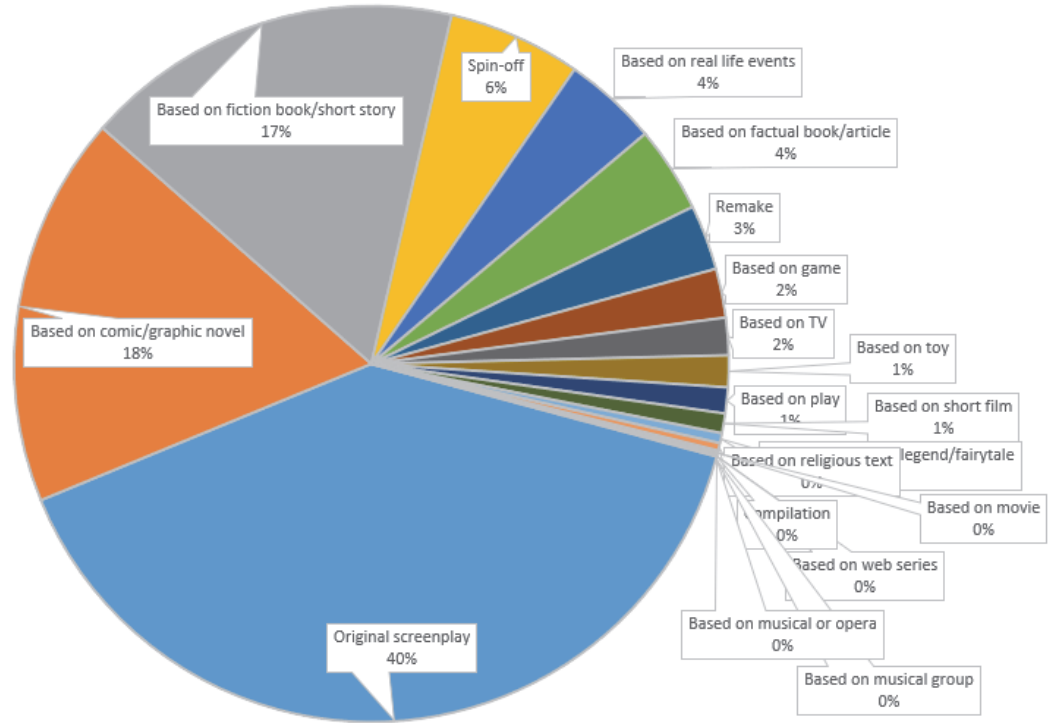


2.89 (a)
cont.

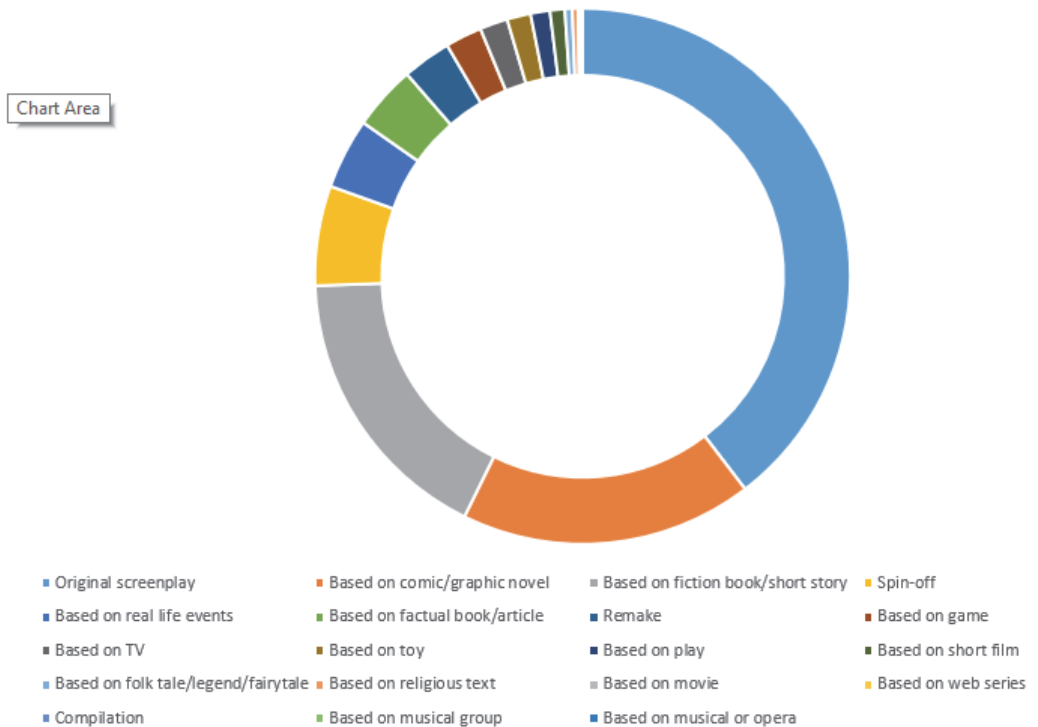


2.89 (a) Gross
cont.

Percentage of Movie Gross

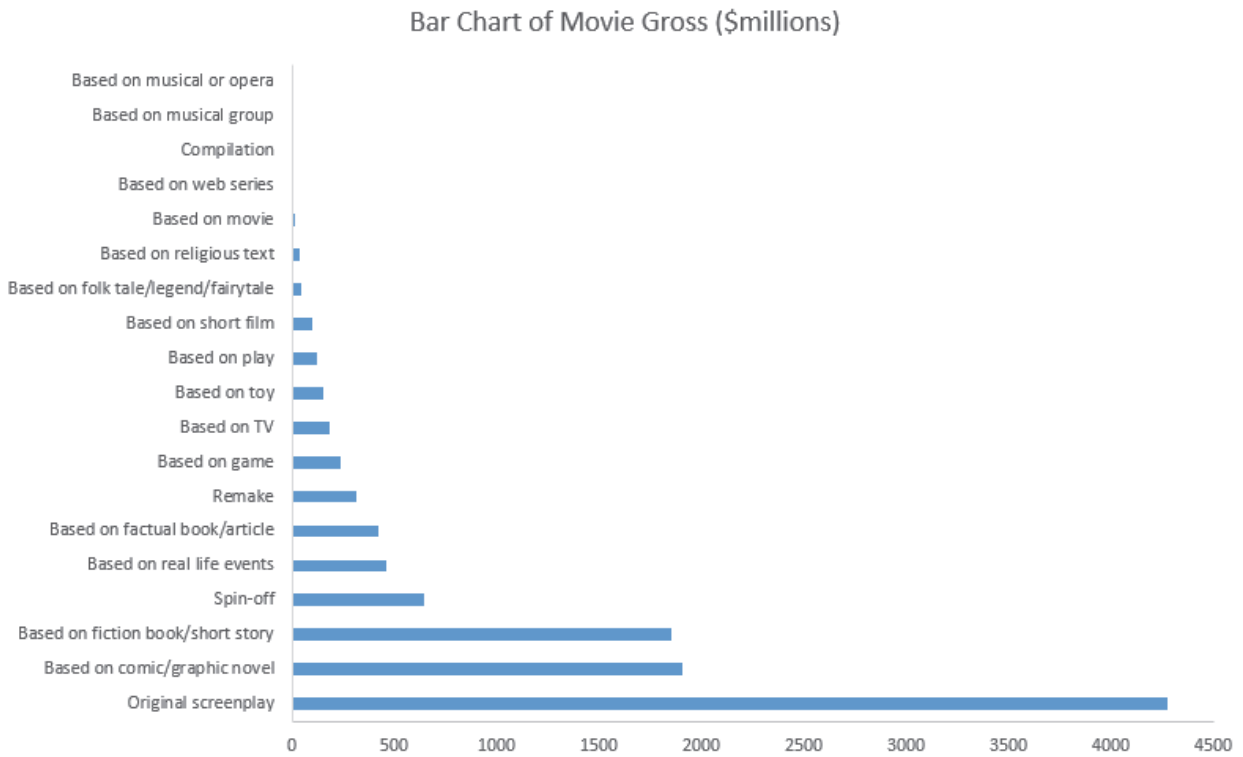


Doughnut Chart of Gross (\$millions)

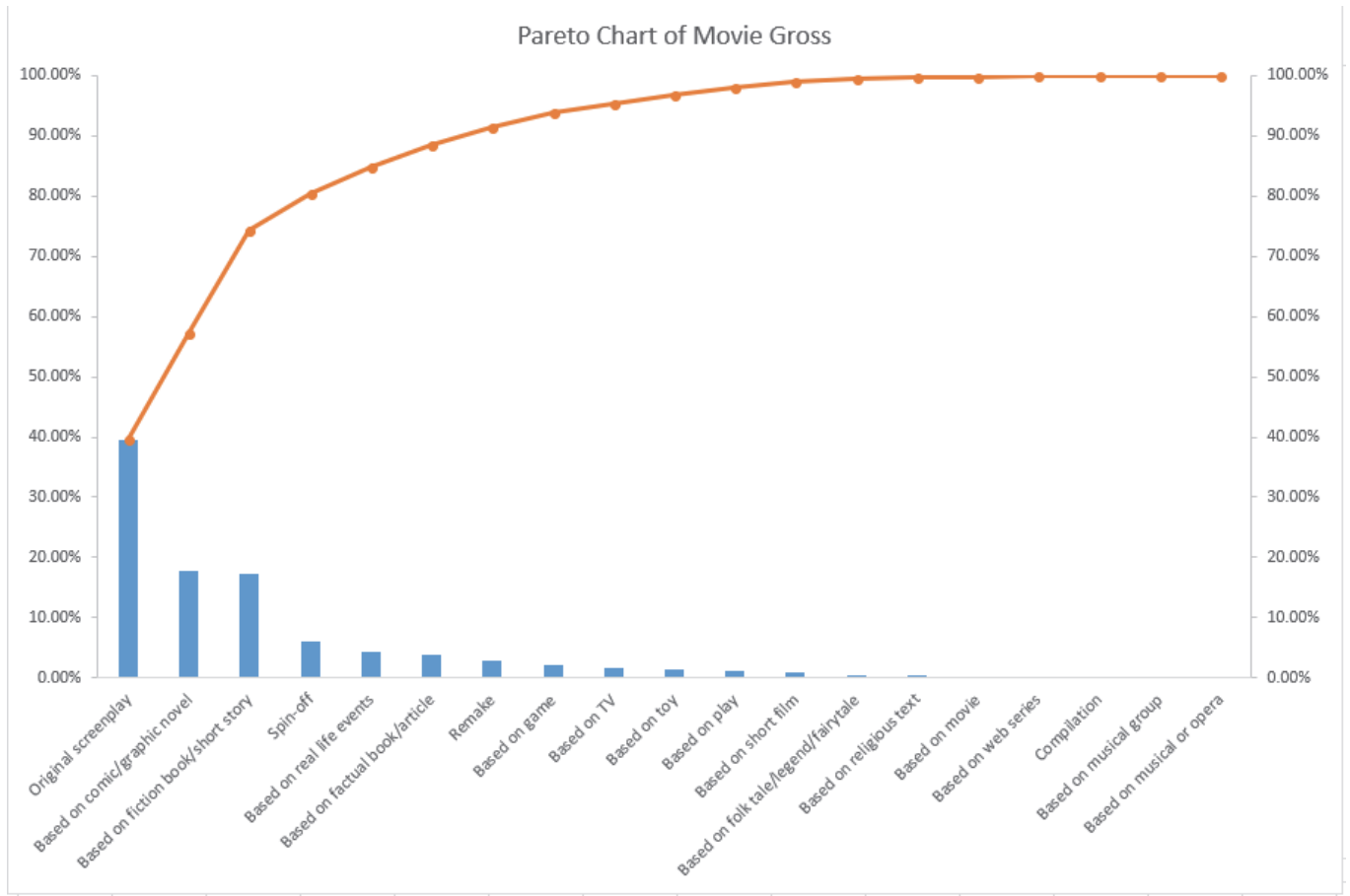


102 Chapter 2: Organizing and Visualizing Variables

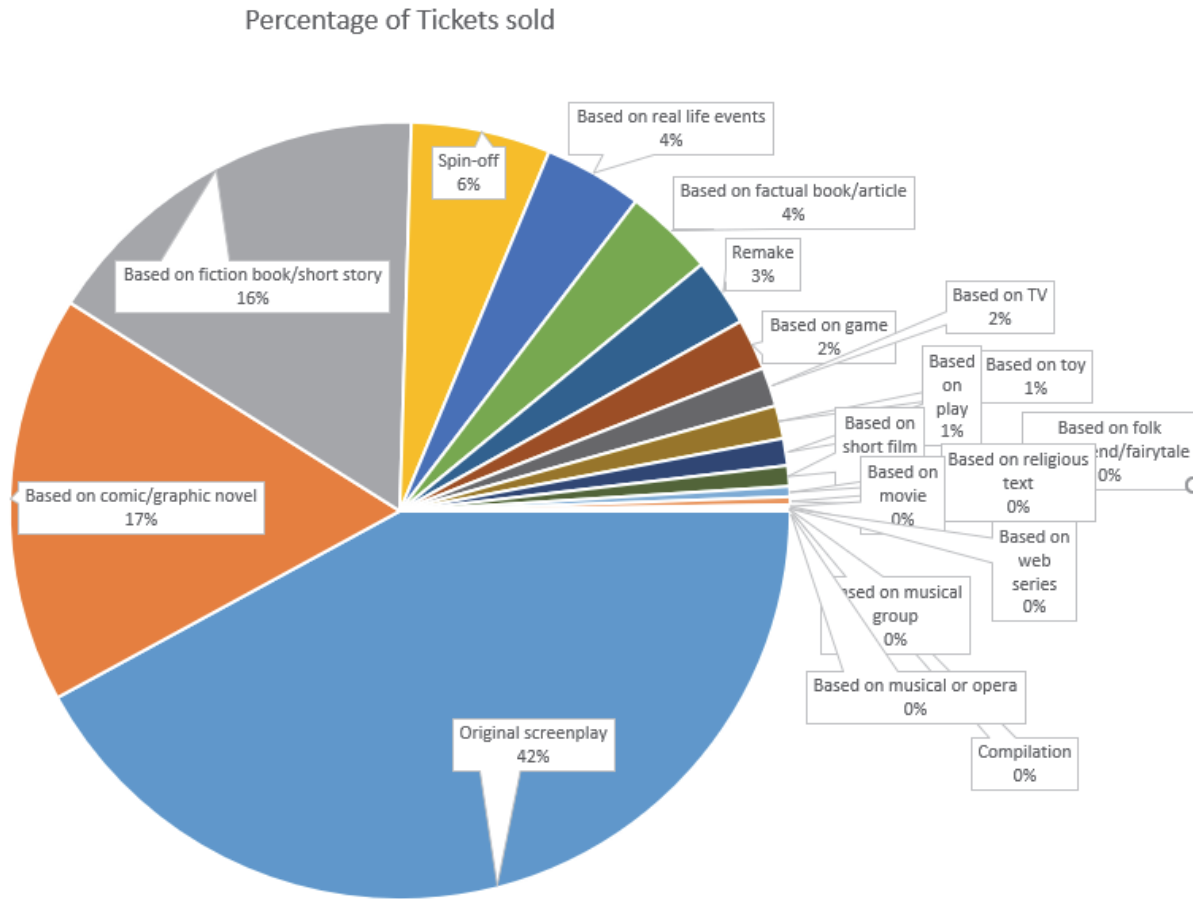
2.89 (a) Gross
cont.



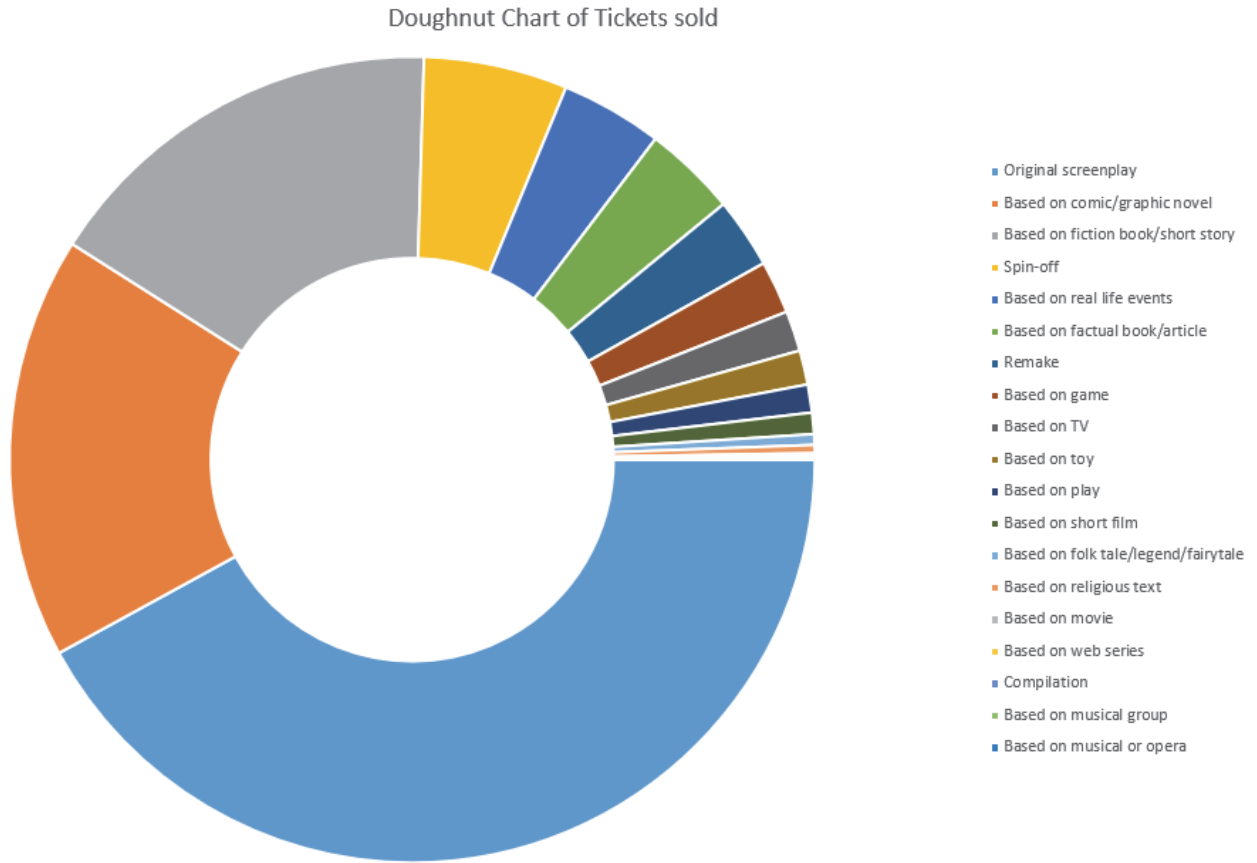
2.89 (a) Tickets Sold
cont.



2.89 (a) Tickets Sold
cont.

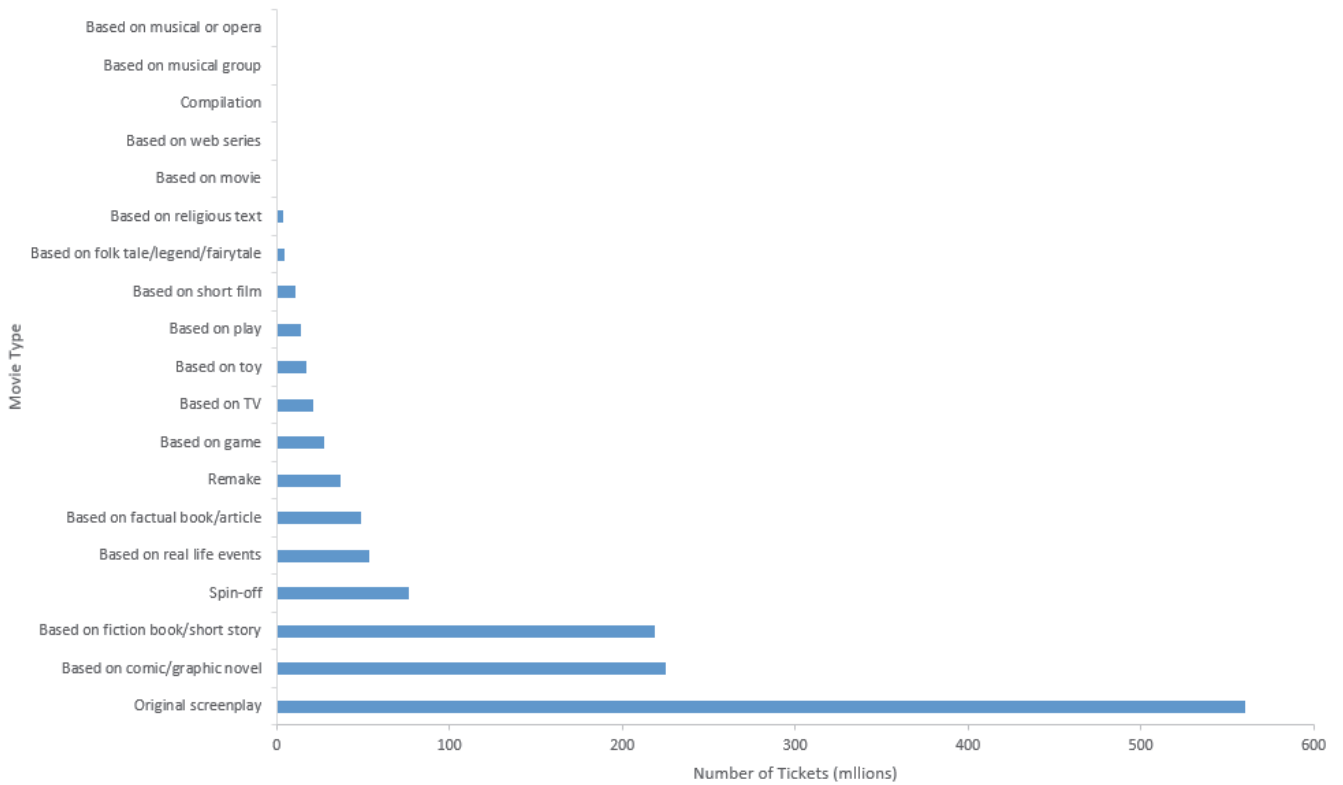


2.89 (a) Tickets Sold
cont.

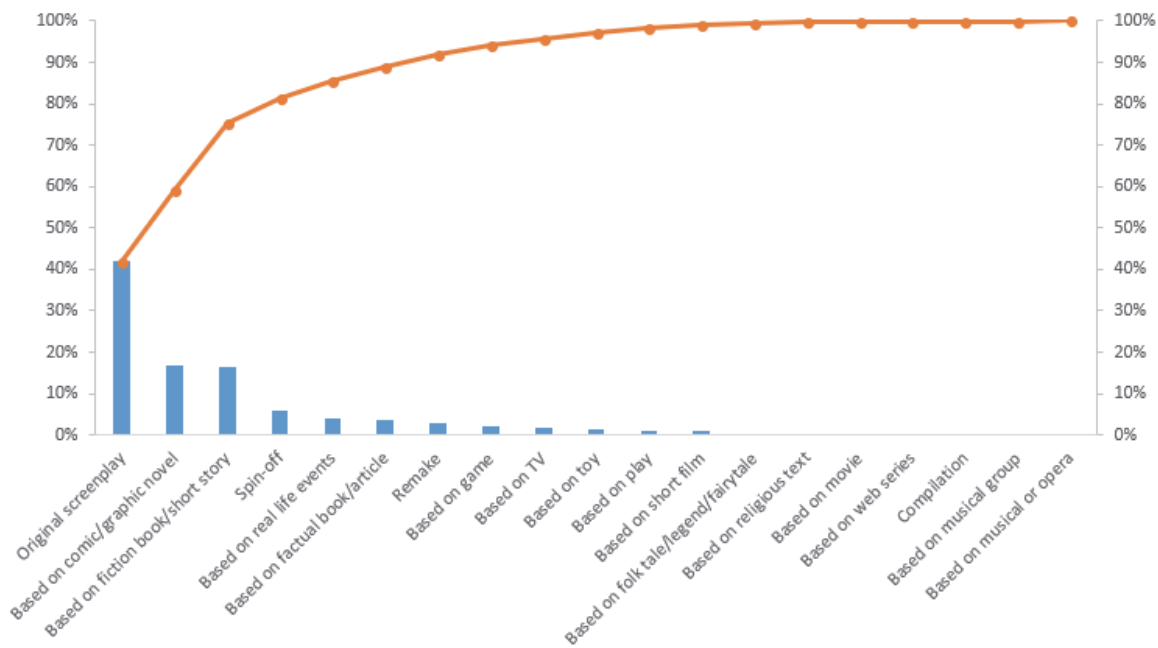


2.89 (a) Tickets Sold
cont.

Bar Chart of Number of Tickets sold in millions by type

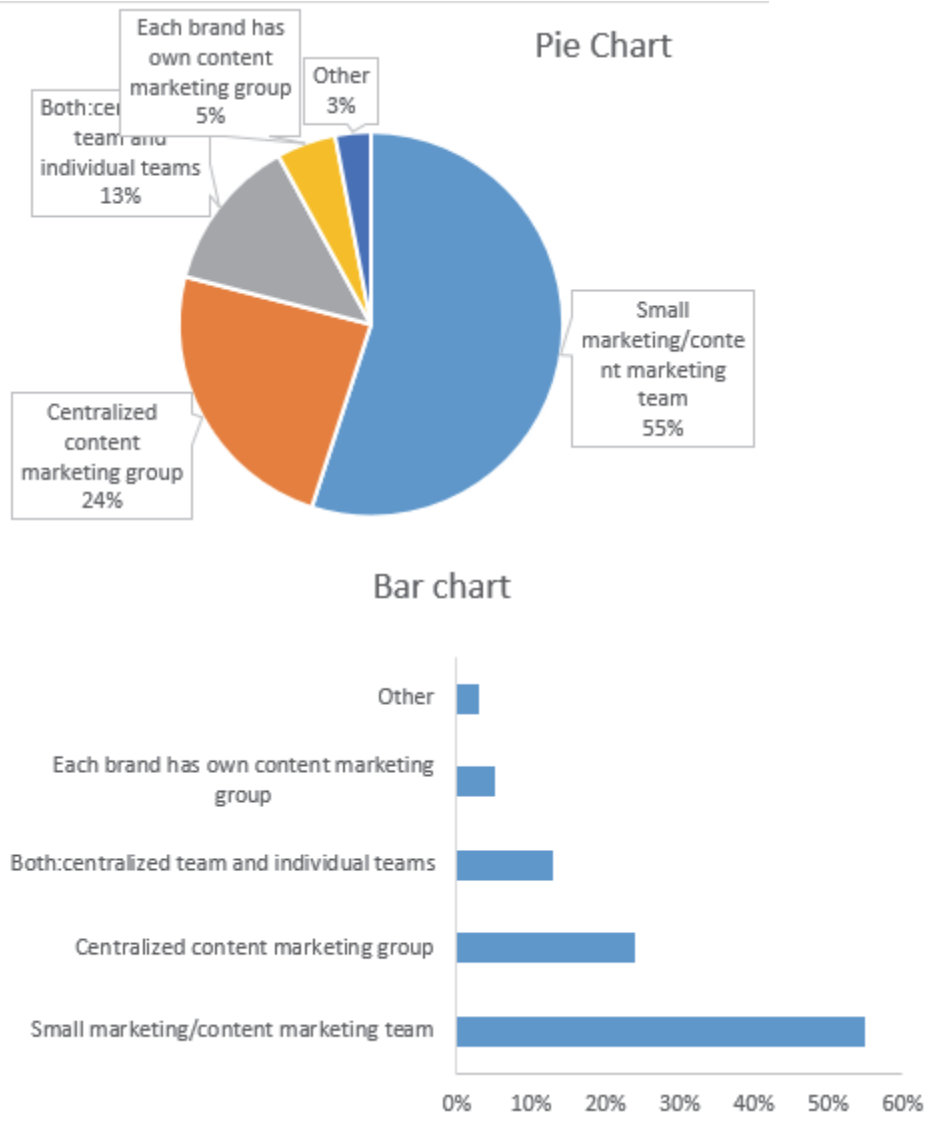


Pareto Chart of Ticket Sales

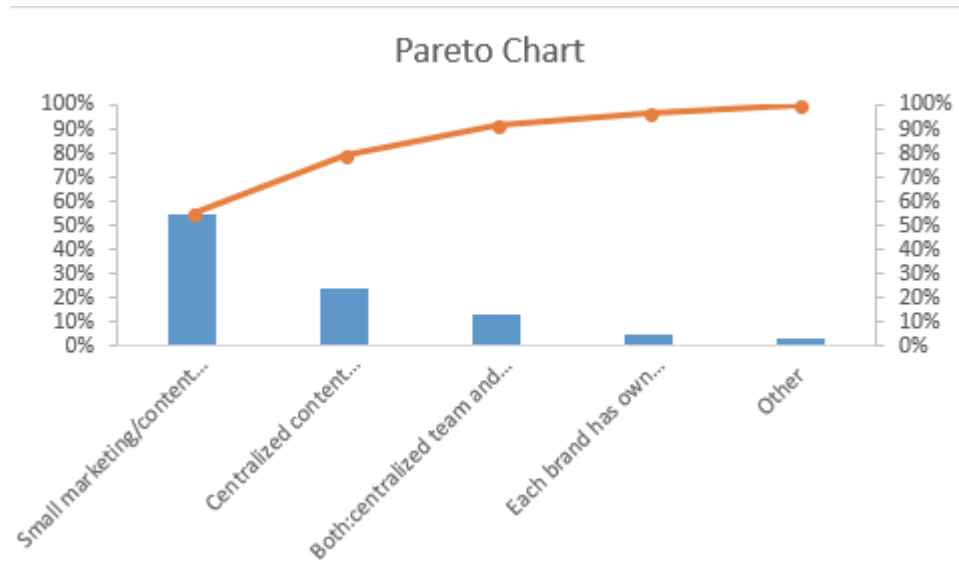


2.89 (b) cont. Based on the Pareto chart for the number of movies, “Original screenplay”, “Based on real life events” and “Based on fiction/short story” are the “vital few” and capture about 88% of the market share. According to the Pareto chart for gross (in \$millions), “Original screenplay”, “Based on fiction book/short story” and “Based on comic/graphic novel” are the “vital few” and capture about 74% of the market share. According to the Pareto chart for number of tickets sold (in millions), “Original screenplay”, “Based on fiction book/short story” and “Based on comic/graphic novel” are the “vital few” and capture about 75% of the market share.

2.90 (a)

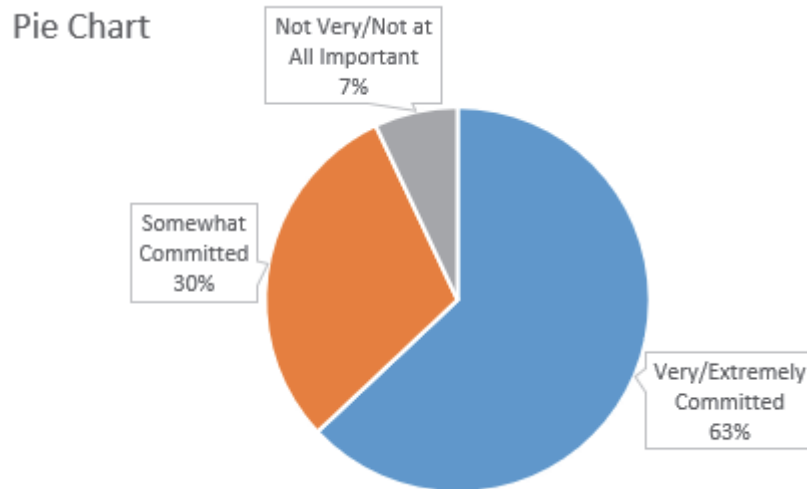


2.90 (a)
cont.

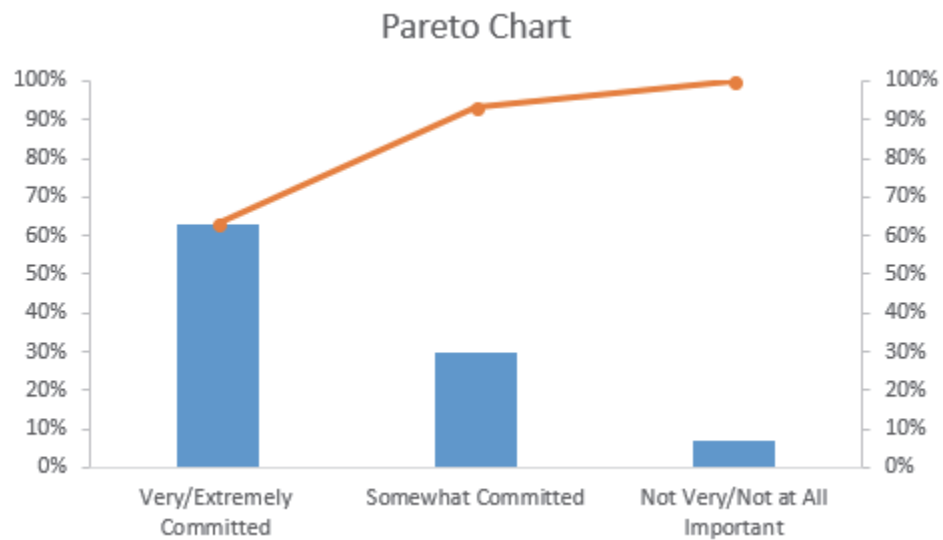
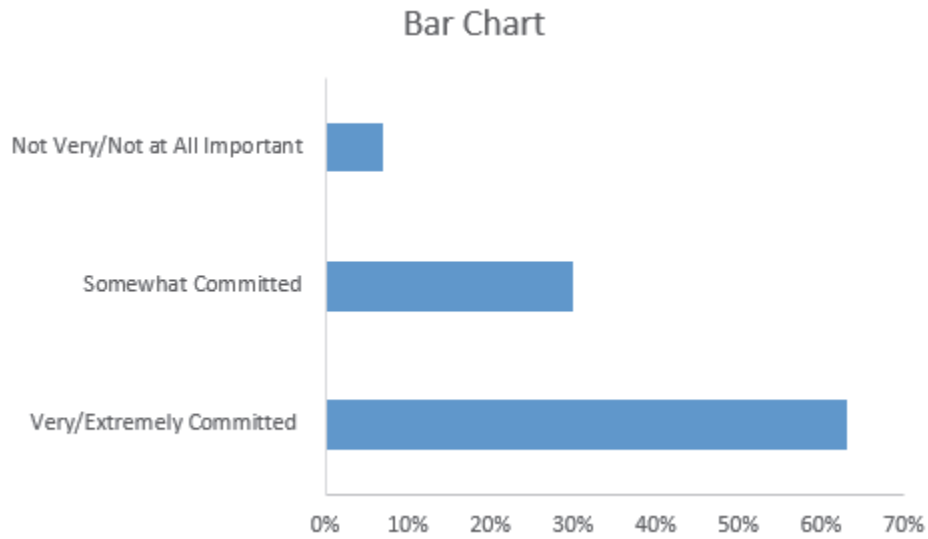


(b) The pie chart or the Pareto chart would be best. The pie chart would allow you to see each category as part of the whole, while the Pareto chart would enable you to see that Small marketing/content marketing team is the dominant category.

(c)



2.90 (c)
cont.



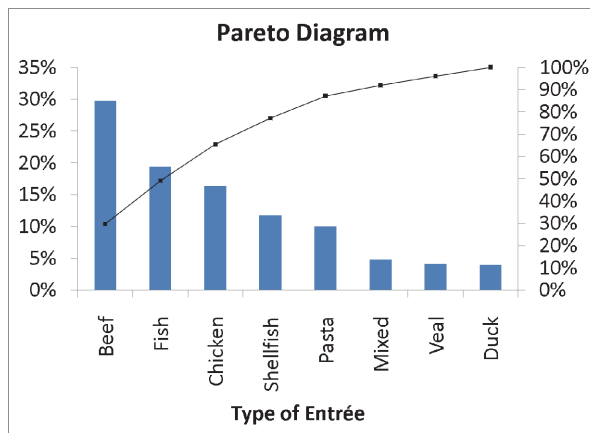
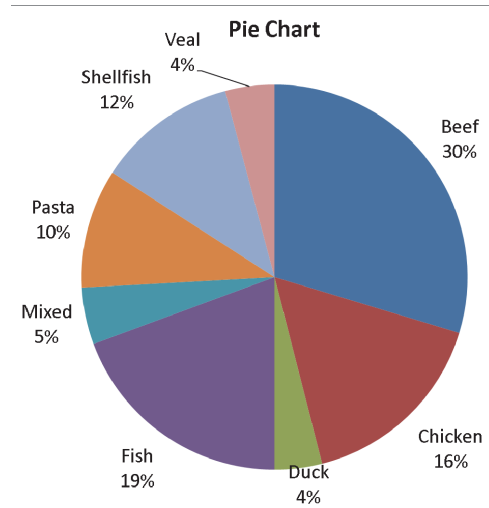
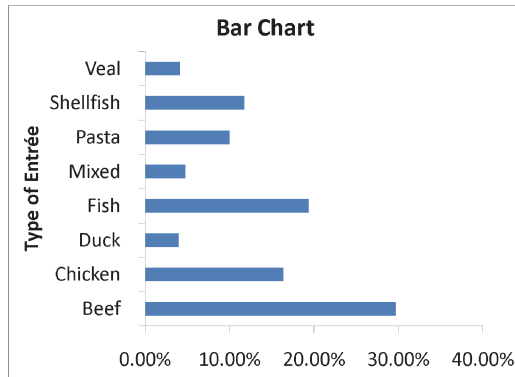
- (d) The pie chart or the Pareto chart would be best. The pie chart would allow you to see each category as part of the whole while the Pareto chart would enable you to see that very committed to content marketing is the dominant category.
- (e) Most organizations have a small marketing/content marketing team and are very committed to content marketing.

110 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)



- (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 entrées ordered by weekend patrons of a continental restaurant. When chicken is included, nearly two-thirds of the entrées are accounted for.

2.92 (a)

Dessert Ordered	Gender		Total
	Male	Female	
Yes	66%	34%	100%
No	48%	52%	100%
Total	52%	48%	100%

Dessert Ordered	Gender		Total
	Male	Female	
Yes	29%	34%	100%
No	71%	52%	100%
Total	100%	48%	100%

Dessert Ordered	Gender		Total
	Male	Female	
Yes	15%	8%	23%
No	37%	40%	77%
Total	52%	48%	100%

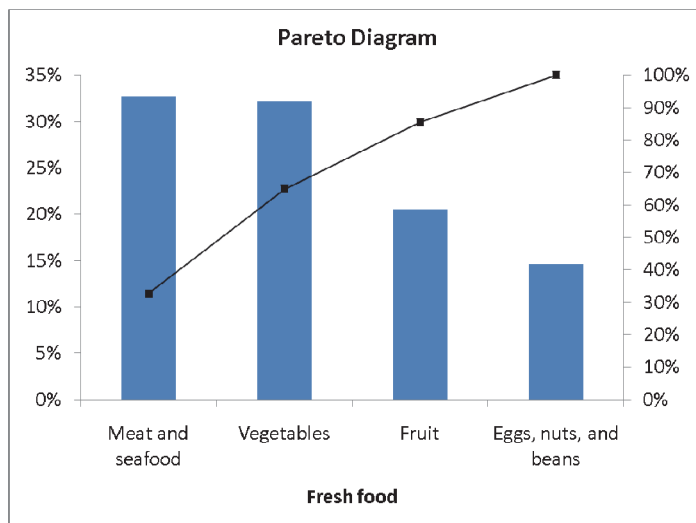
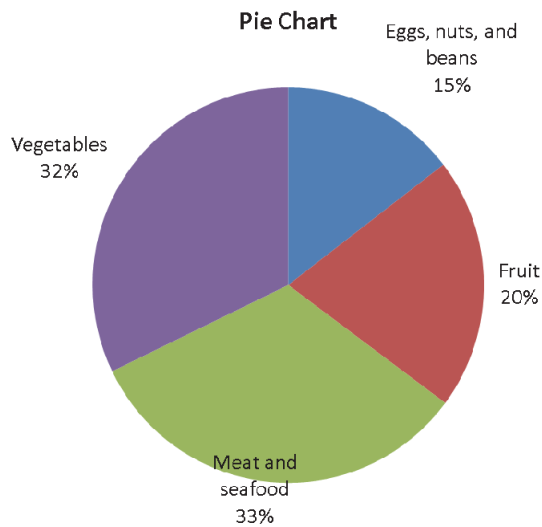
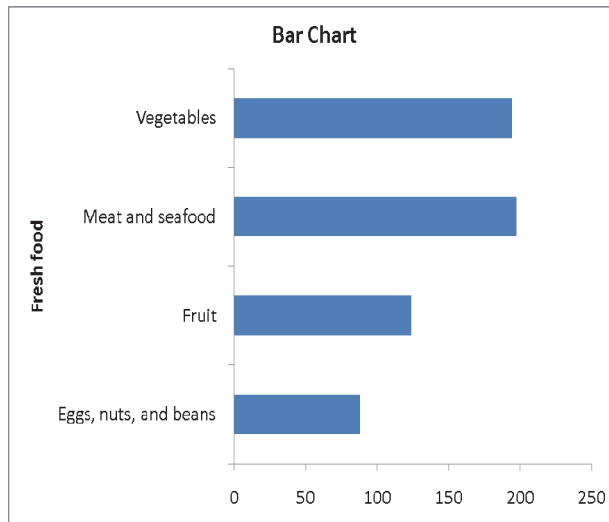
Dessert Ordered	Gender		Total
	Male	Female	
Yes	52%	48%	100%
No	25%	75%	100%
Total	31%	69%	100%

Dessert Ordered	Gender		Total
	Male	Female	
Yes	38%	16%	23%
No	62%	84%	77%
Total	100%	100%	100%

Dessert Ordered	Gender		Total
	Male	Female	
Yes	11.75%	10.79%	22.54%
No	19.52%	57.94%	77.46%
Total	31.27%	68.73%	100%

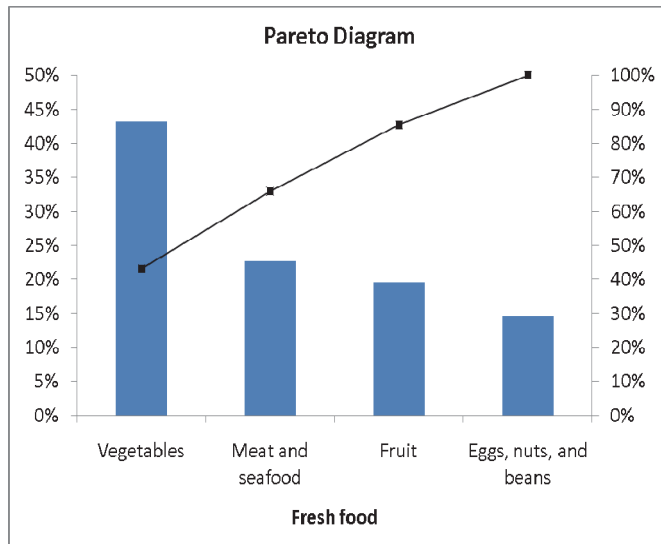
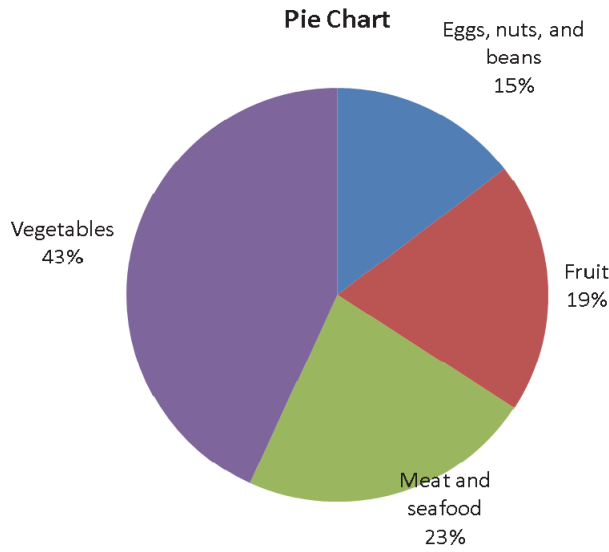
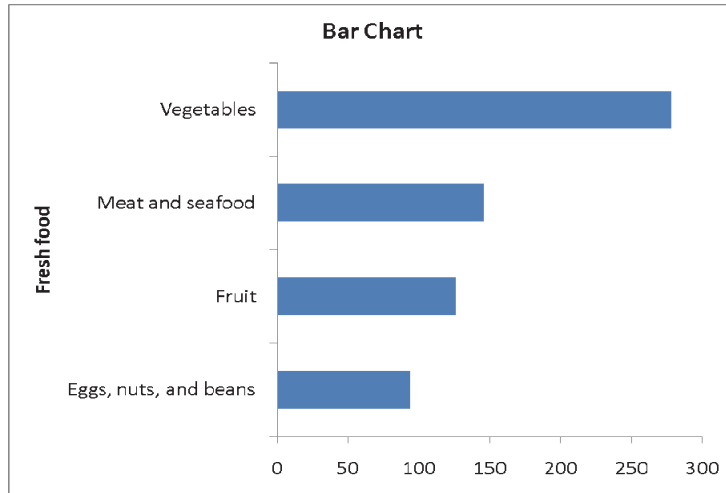
- (b) If the owner is interested in finding out the percentage of males and females who order dessert or the percentage of those who order a beef entrée and a dessert among all patrons, the table of total percentages is most informative. If the owner is interested in the effect of gender on ordering of dessert or the effect of ordering a beef entrée on the ordering of dessert, the table of column percentages will be most informative. Because dessert is usually ordered after the main entrée, and the owner has no direct control over the gender of patrons, the table of row percentages is not very useful here.
- (c) 29% of the men ordered desserts, compared to 17 of the women; men are almost twice as likely to order dessert as women. Almost 38% of the patrons ordering a beef entrée ordered dessert, compared to 16% of patrons ordering all other entrées. Patrons ordering beef are more than 2.3 times as likely to order dessert as patrons ordering any other entrée.

2.93 (a) United States Fresh Food Consumed:

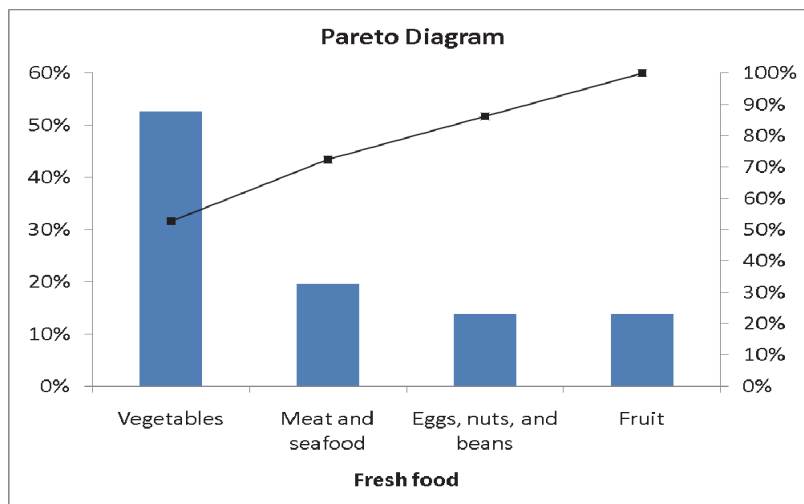
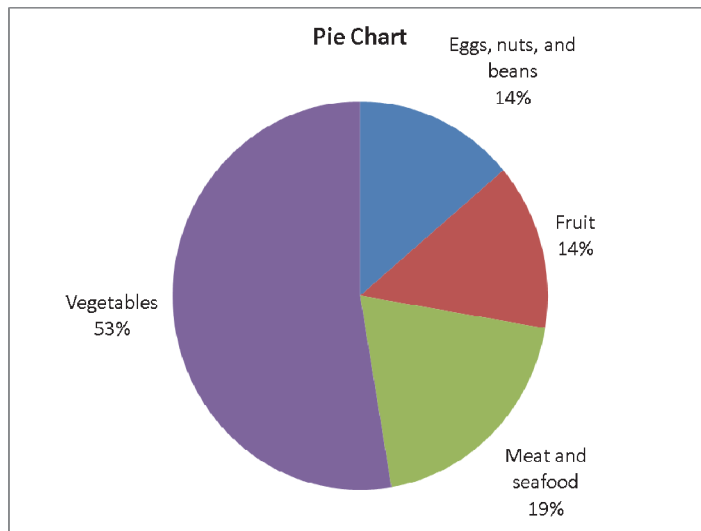
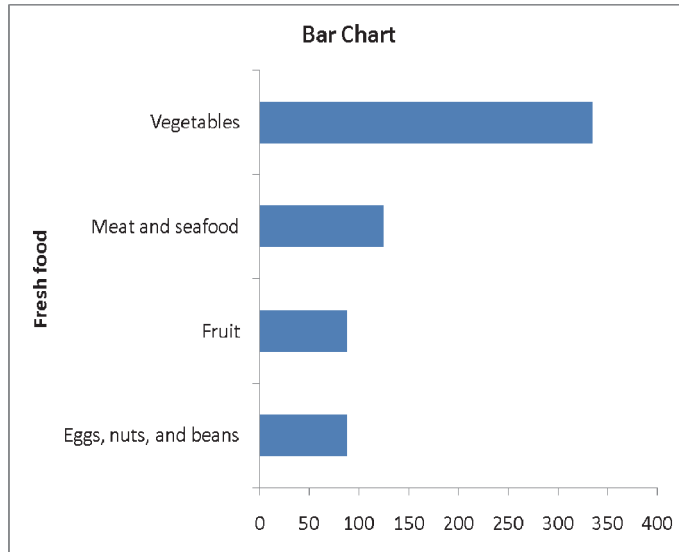


2.93 (a)
cont.

Japan Fresh Food Consumed:

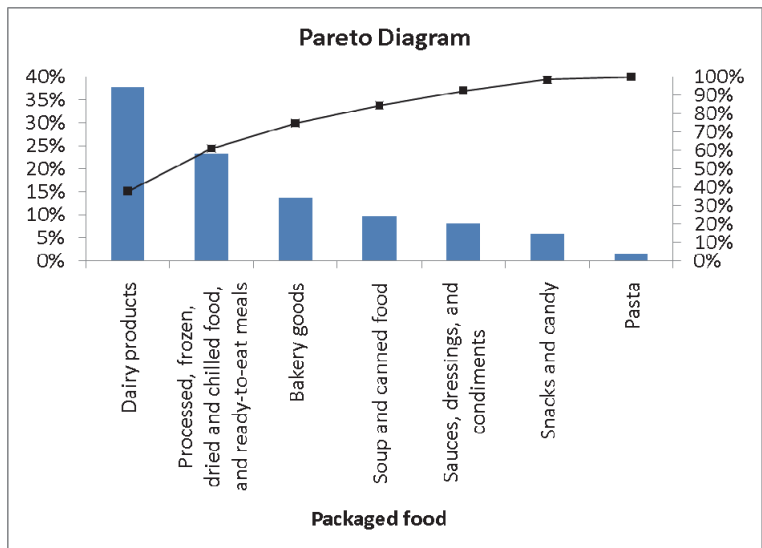
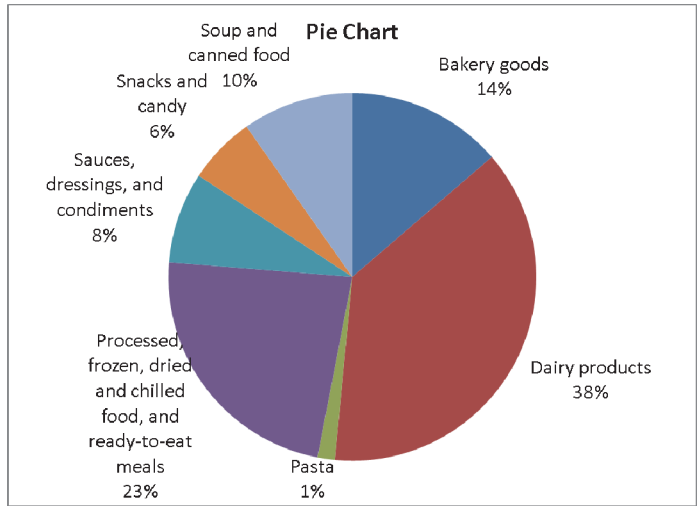
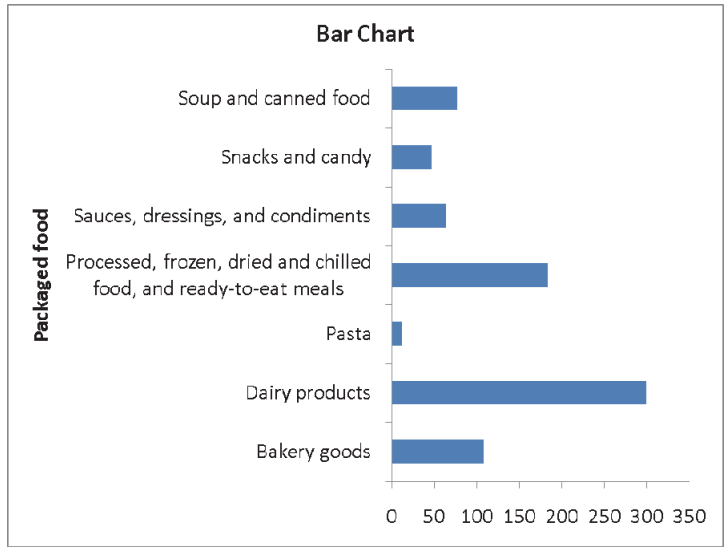


2.93 (a) **Russia Fresh Food Consumed:**
cont.

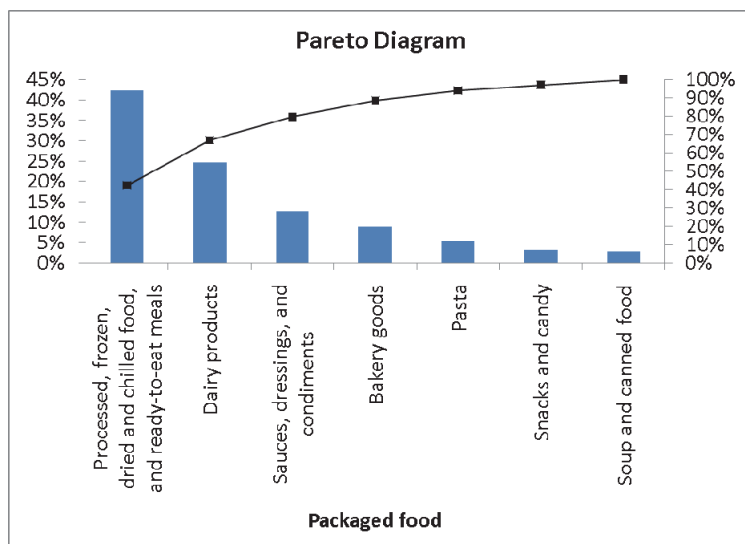
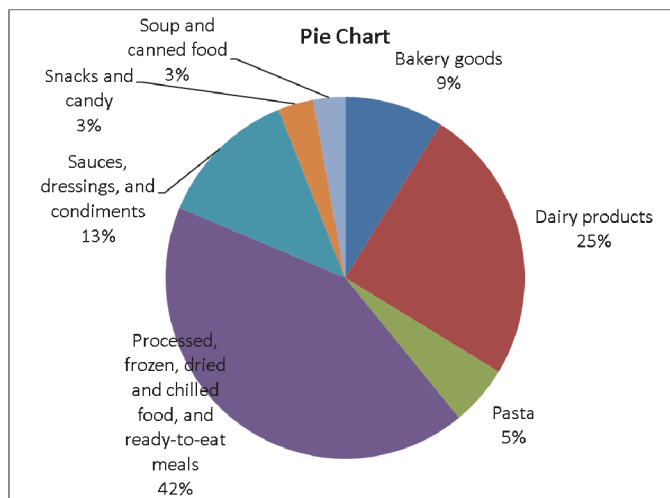
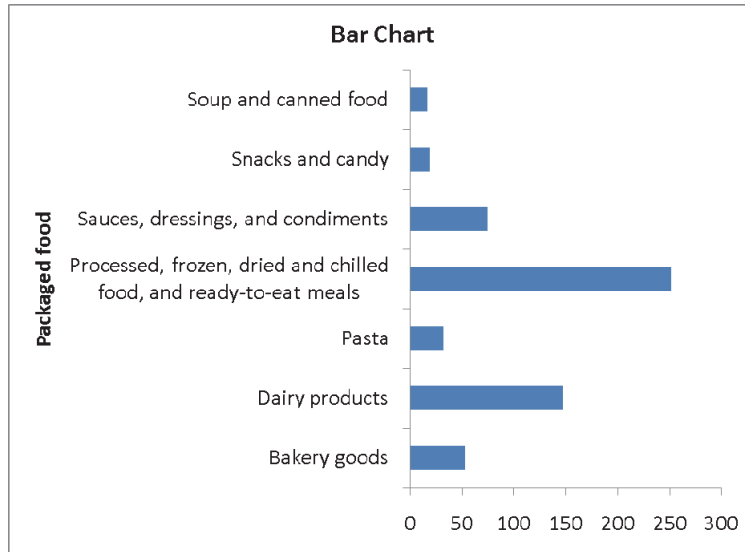


2.93 (b) cont.

United States Packaged Food Consumed:

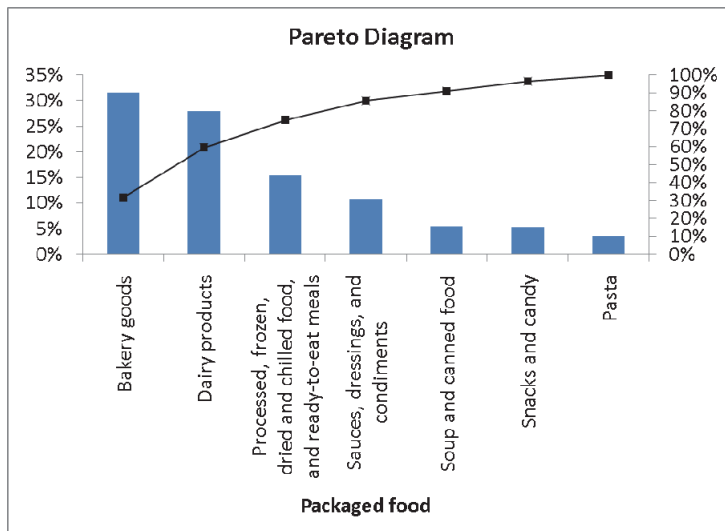
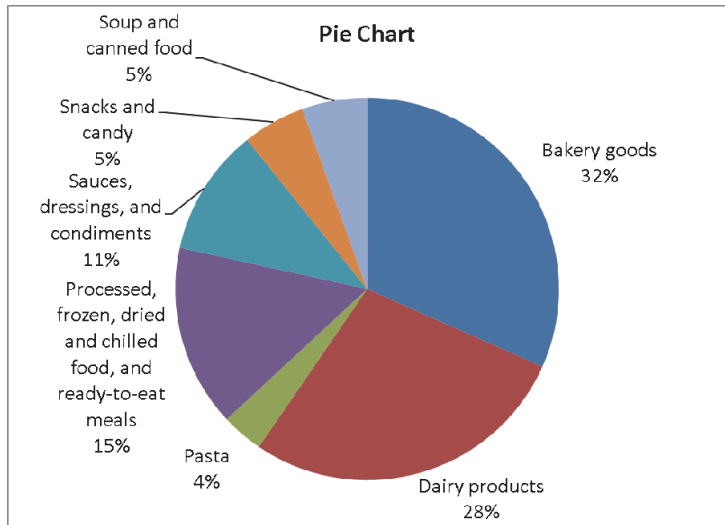
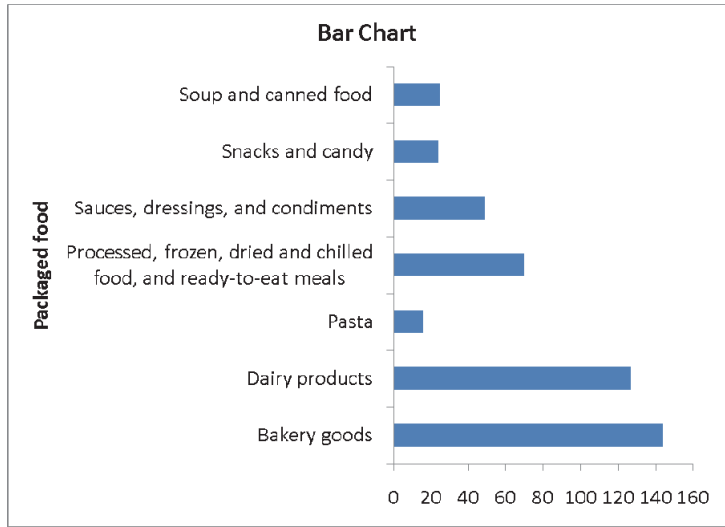


2.93 (b) Japan Packaged Food Consumed:
cont.



2.93 (b)
cont.

Russian Packaged Food Consumed:

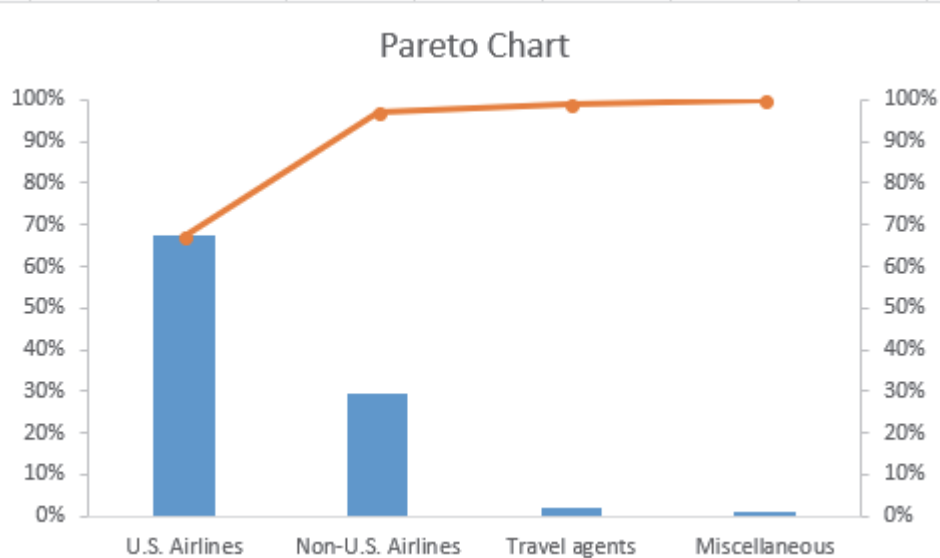


118 Chapter 2: Organizing and Visualizing Variables

2.93 (c) cont. 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.

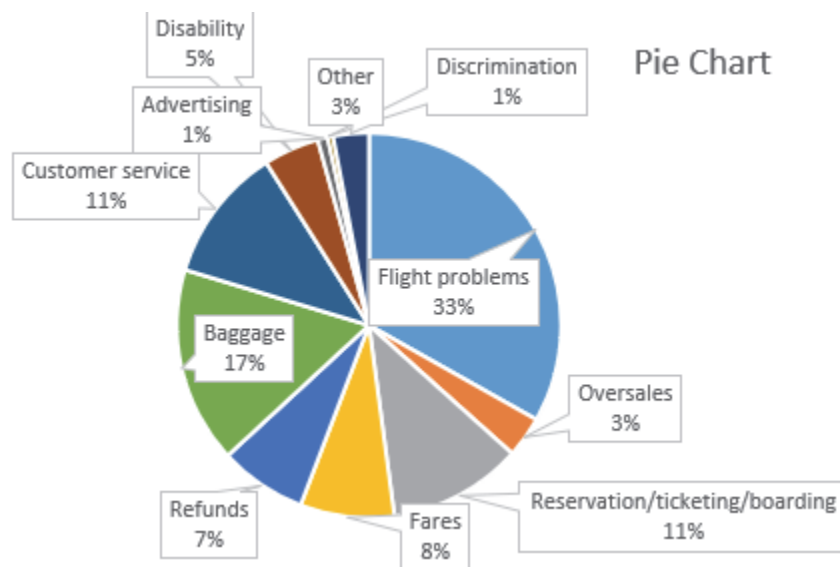
For Americans, dairy products, and processed, frozen, dried and chilled food and ready-to-eat meals make up slightly more than 60% of the packaged food consumption. For Japanese, processed, frozen, dried and chilled food, and ready-to-eat meals, and dairy products constitute more than 60% of their packaged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

2.94 (a)



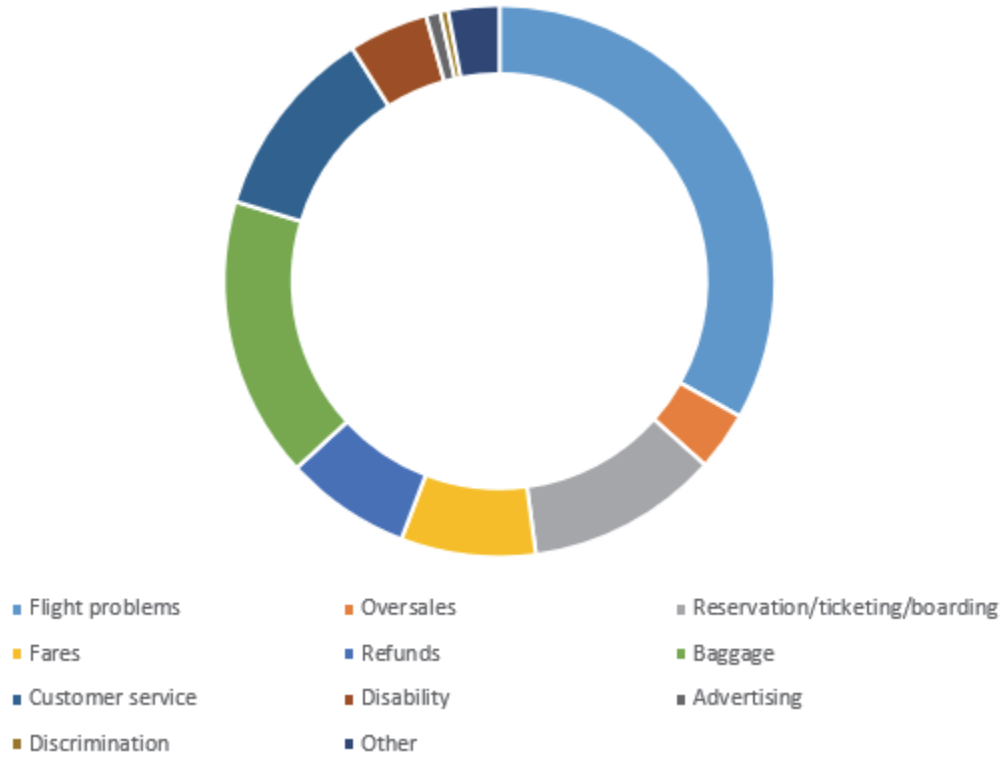
Most of the complaints were against U.S. airlines.

(b)



2.94 (b)
cont.

Doughnut Charts

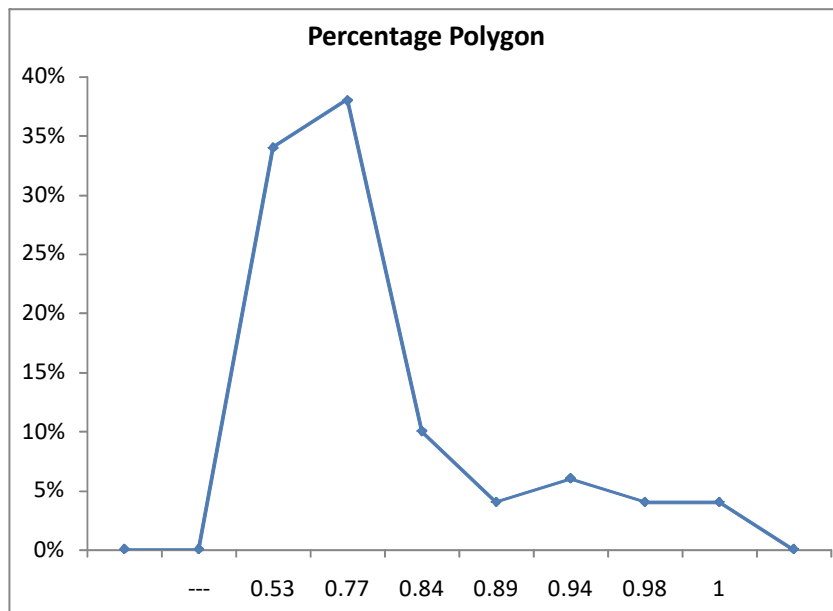
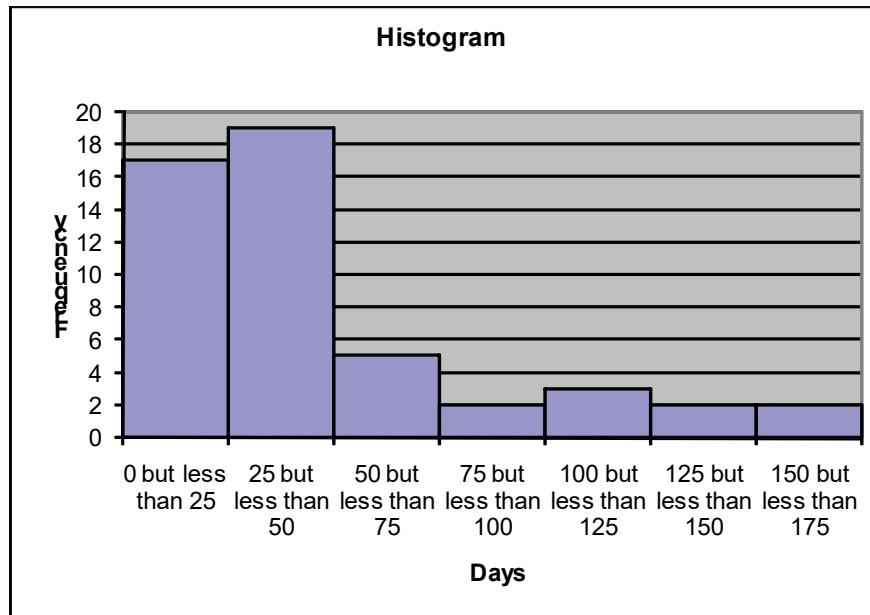


More of the complaints were due to flight problems.

2.95 (a)

<i>Range</i>	<i>Frequency</i>	<i>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%

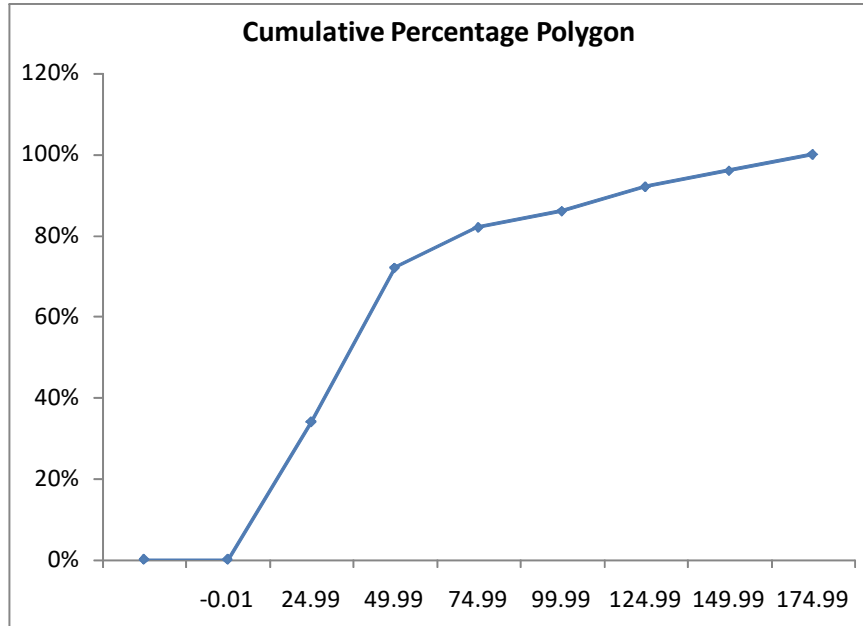
2.95 (b)
cont.



(c)

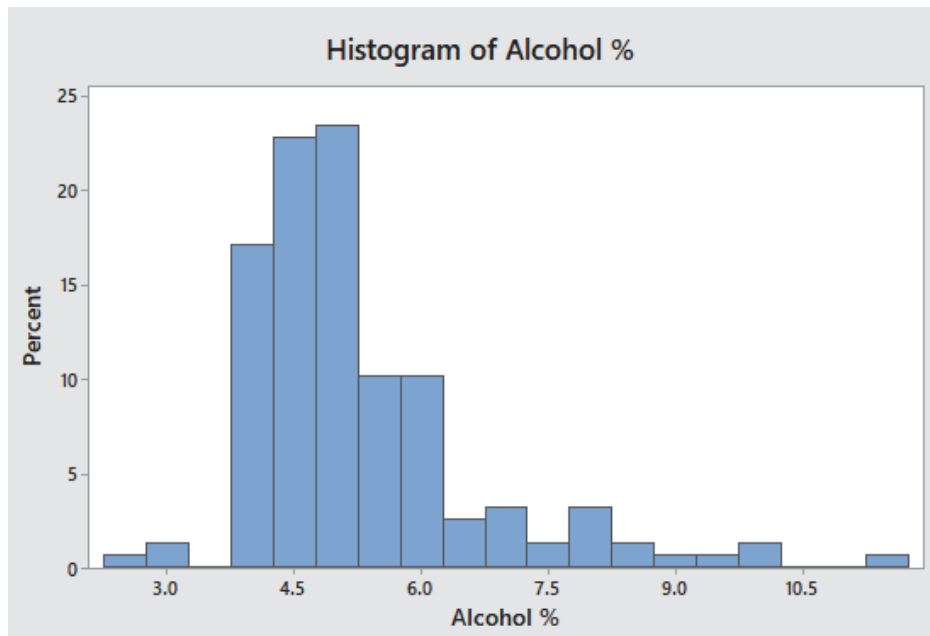
Range	Cumulative %
0 but less than 25	34%
25 but less than 50	72%
50 but less than 75	82%
75 but less than 100	86%
100 but less than 125	92%
125 but less than 150	96%
150 but less than 175	100%

2.95 (c)
cont.

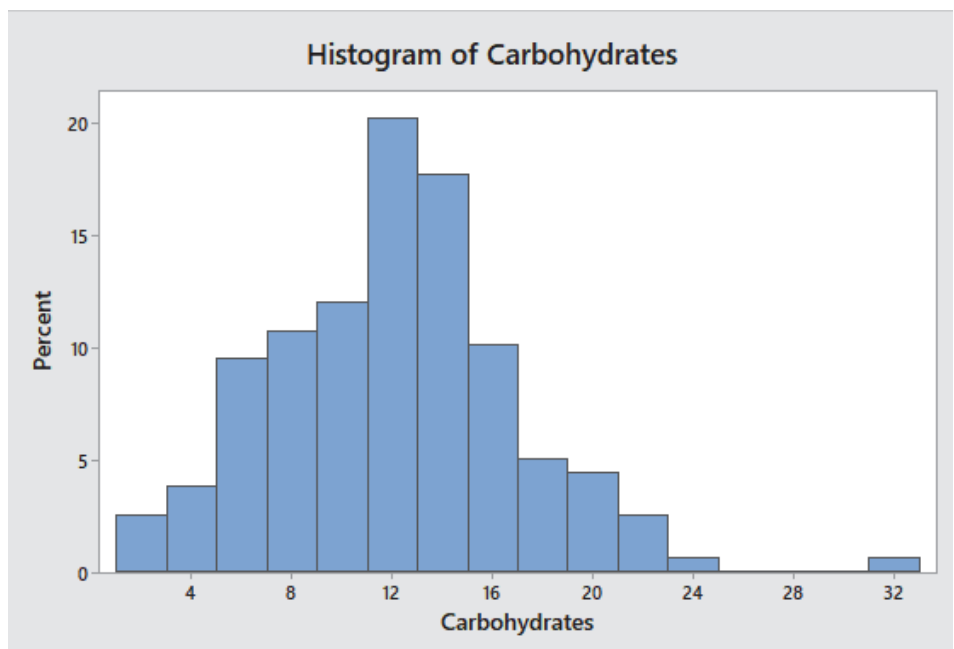
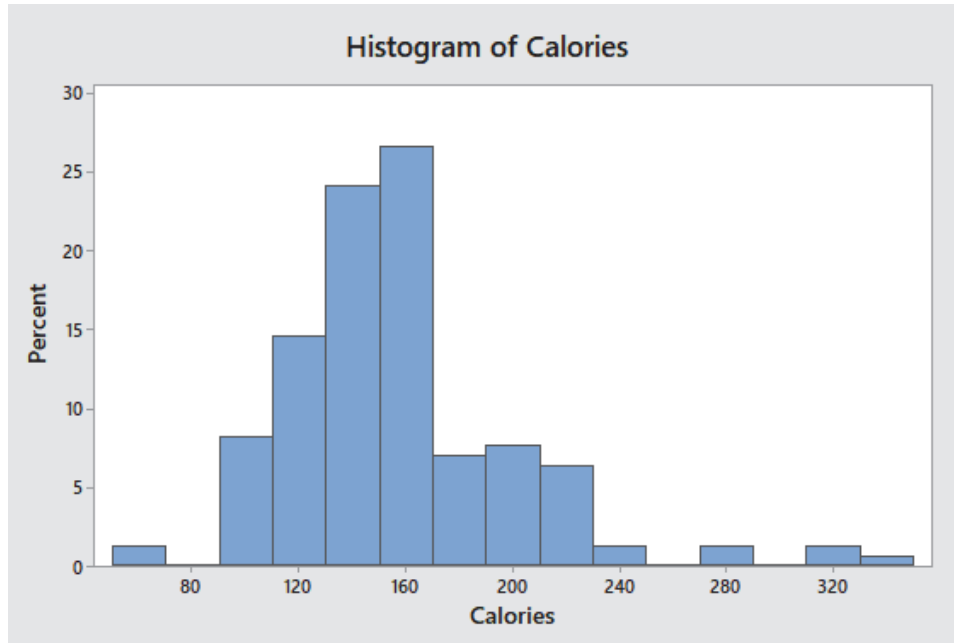


(d) You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

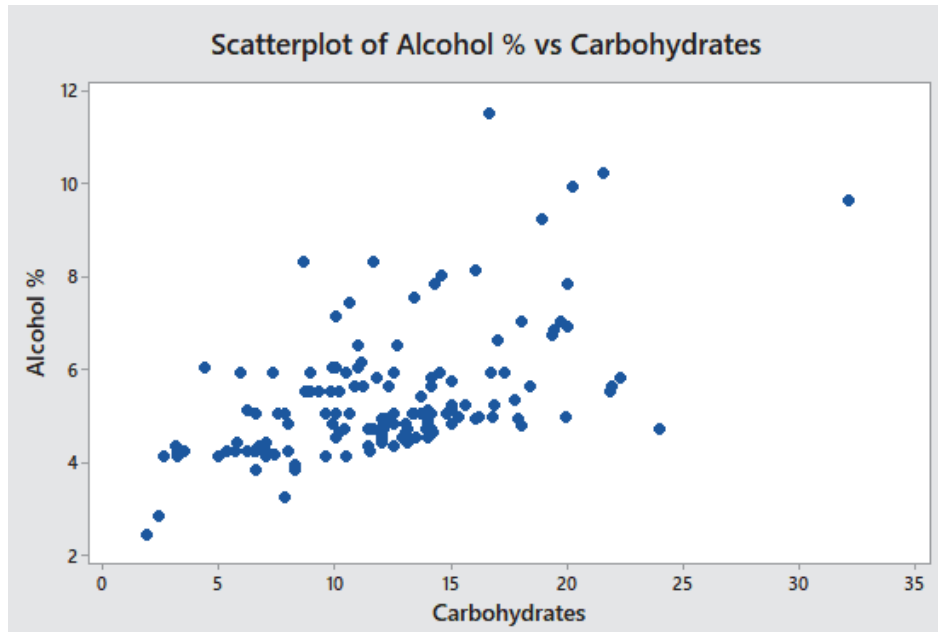
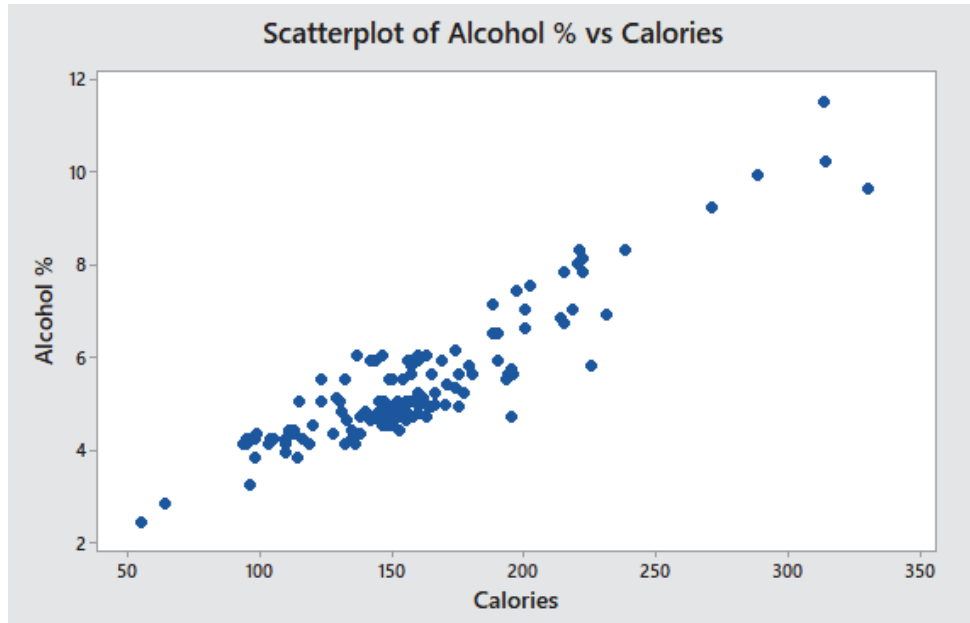
2.96 (a)



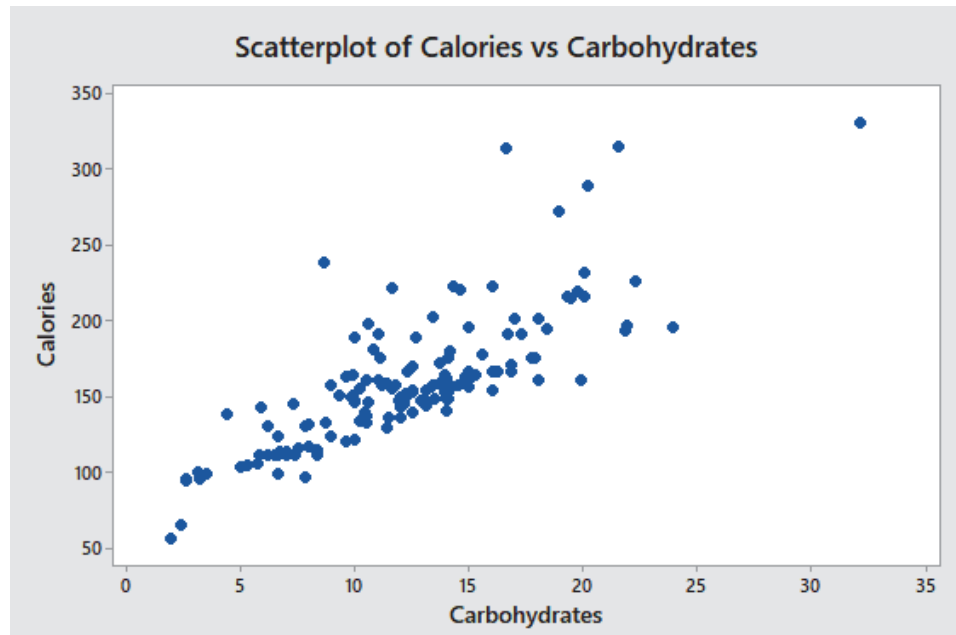
2.96 (a)
cont.



2.96 (b)
cont.



2.96 (b)
cont.

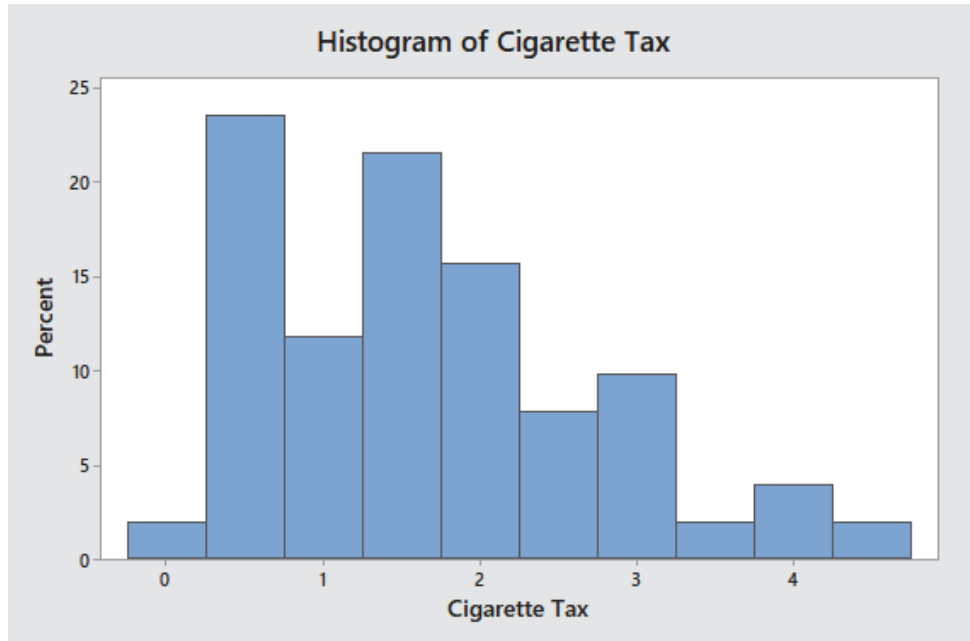


(c) The alcohol percentage is concentrated between 4% and 6%, with more between 4% and 5%. The calories are concentrated between 140 and 160. The carbohydrates are concentrated between 12 and 15. There are outliers in the percentage of alcohol in both tails. There are a few beers with alcohol content as high as around 11.5%. There are a few beers with calorie content as high as around 313 and carbohydrates as high as 32.1. There is a strong positive relationship between percentage of alcohol and calories and between calories and carbohydrates, and there is a moderately positive relationship between percentage alcohol and carbohydrates.

2.97 (a)

0.17	0.3	0.37	0.44	0.45	0.57	0.57	0.6	0.6	0.62	0.64
0.675	0.68	0.84	0.995	1.03	1.08	1.15	1.2	1.29	1.32	1.339
1.36	1.41	1.53	1.6	1.6	1.66	1.7	1.7	1.78	1.8	1.98
2	2	2	2	2	2.5	2.52	2.6	2.7	2.87	3.025
3.04	3.08	3.2	3.51	3.75	3.9	4.35				

2.97 (b)
cont.



(c) Most of the states have a Cigarette Tax between \$0.50 and \$2. Five percent of the states have a cigarette tax of at least \$3.5.

2.98 (a)

Stem-and-leaf of One-Year N = 39

```

2  0  55
6  1  0055
11 2  05558
13 3  05
16 4  055
19 5  000
19 6
(1) 7  5
19 8  0
18 9  55
16 10 00567
11 11  55
9  12 015688
3  13  015
    
```

Leaf Unit = 0.01

2.98 (a)
cont.

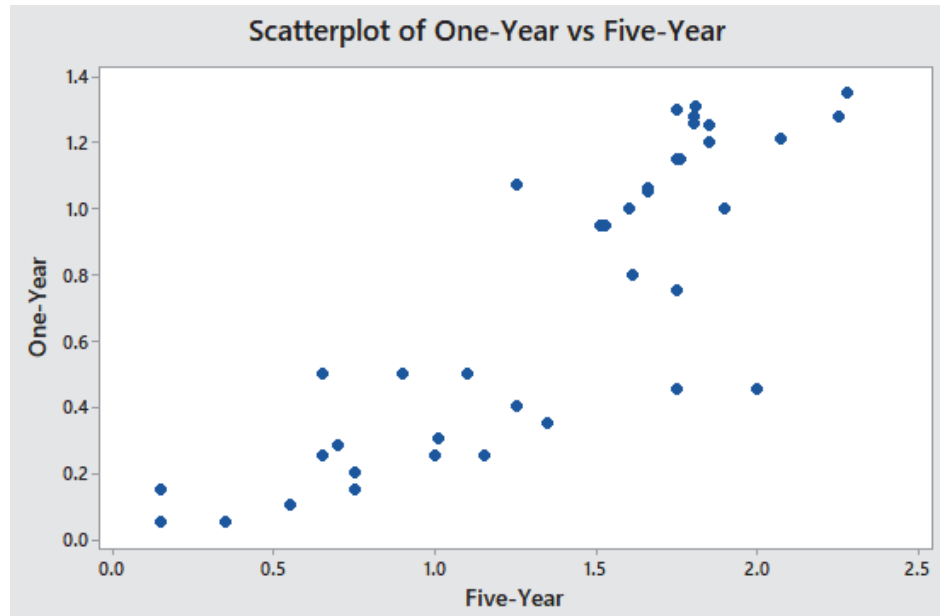
Stem-and-leaf of Five-Year N = 39

```

2  0  11
3  0  3
5  0  55
10 0  66777
11 0  9
15 1  0011
18 1  223
(2) 1  55
19 1  666677777
10 1  888889
4  2  00
2  2  22
    
```

Leaf Unit = 0.1

(b)



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

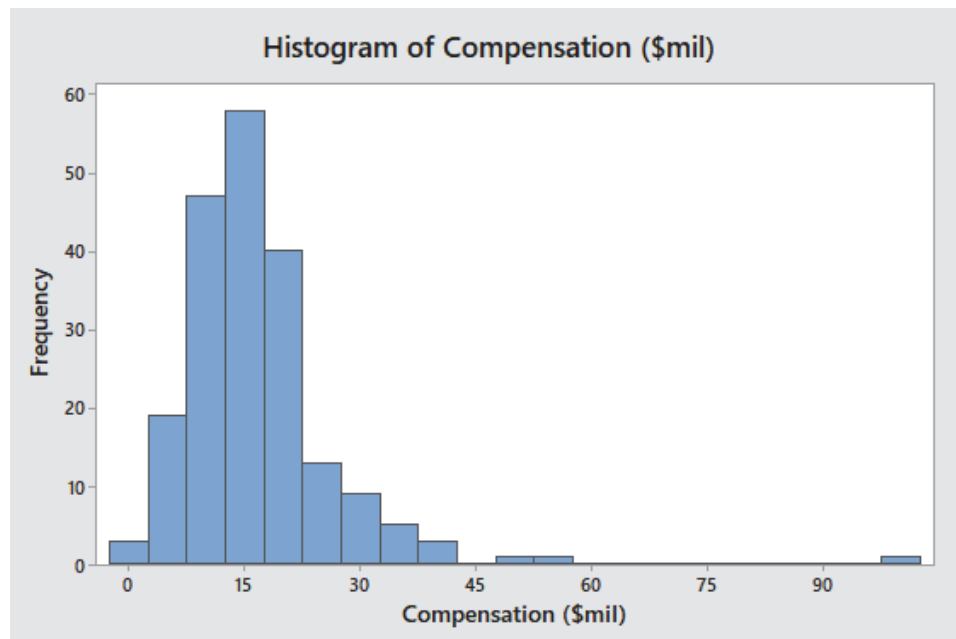
2.99 (a) Frequency Distribution

	A	B
1	Total Compensation (millions \$)	Frequency
2	0 but less than 10	41
3	10 but less than 20	108
4	20 but less than 30	38
5	30 but less than 40	8
6	40 but less than 50	3
7	50 but less than 60	1
8	60 but less than 70	0
9	70 but less than 80	0
10	80 but less than 90	0
11	90 but less than 100	0
12	100 but less than 110	1
13	Total	200
14		

Percentage distribution

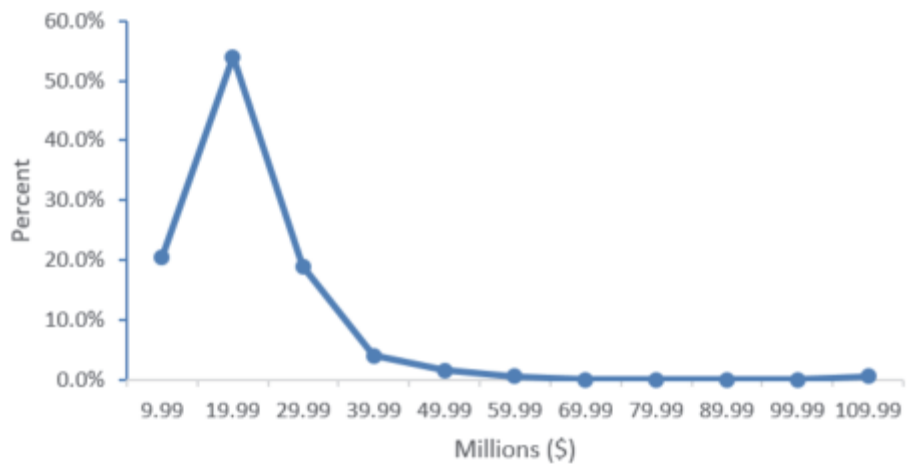
D	F
Total Compensation (millions \$)	Percentage
0 but less than 10	20.5%
10 but less than 20	54.0%
20 but less than 30	19.0%
30 but less than 40	4.0%
40 but less than 50	1.5%
50 but less than 60	0.5%
60 but less than 70	0.0%
70 but less than 80	0.0%
80 but less than 90	0.0%
90 but less than 100	0.0%
100 but less than 110	0.5%
Total	100%

(b)



2.99 (b)
cont.

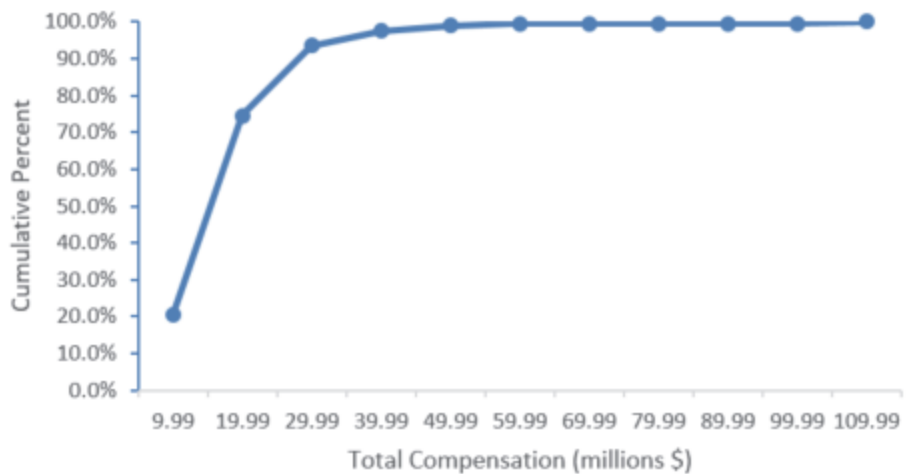
Percentage Polygon for CEO Total Compensation



(c)

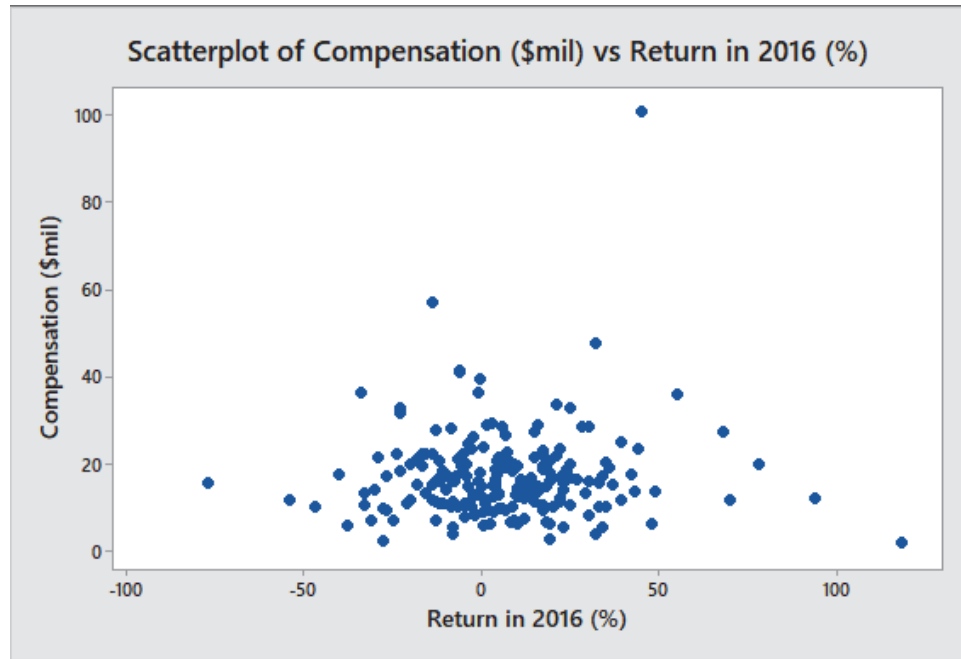
D	F	G
Total Compensation (millions \$)	Percentage	Cumulative Percentage
0 but less than 10	20.5%	20.5%
10 but less than 20	54.0%	74.5%
20 but less than 30	19.0%	93.5%
30 but less than 40	4.0%	97.5%
40 but less than 50	1.5%	99.0%
50 but less than 60	0.5%	99.5%
60 but less than 70	0.0%	99.5%
70 but less than 80	0.0%	99.5%
80 but less than 90	0.0%	99.5%
90 but less than 100	0.0%	99.5%
100 but less than 110	0.5%	100.0%
Total	100%	

Cumulative Percentage Polygon for CEO Total Compensation



2.99 cont. (d) The majority of CEOs (54%) have a total compensation between 10 and 20 million dollars. Total compensation ranges between 1.7 and 100.1 million dollars. Total compensation for 94% of the CEOs is between 1.7 and 30 million dollars.

(e)



(f) There is no relationship between the total CEO compensation and shareholder return.

2.100 (a)

Frequencies (Boston)

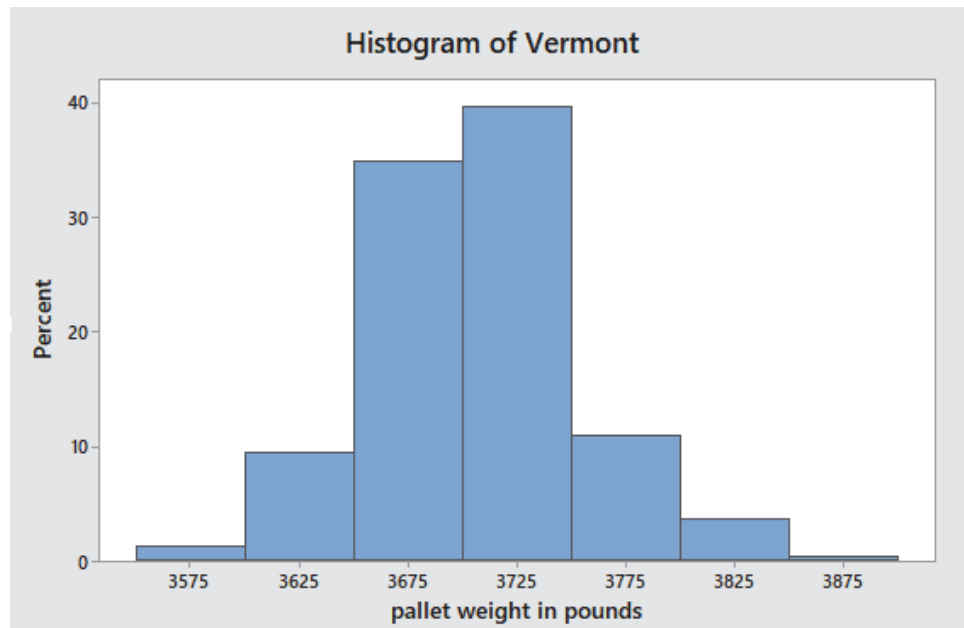
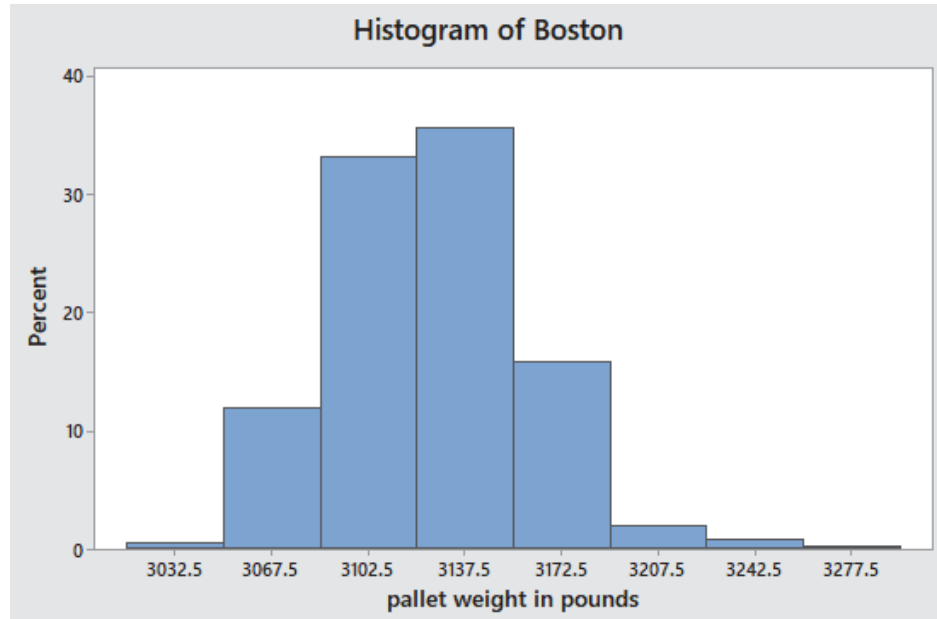
<i>Weight (Boston)</i>	<i>Frequency</i>	<i>Percentage</i>
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but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

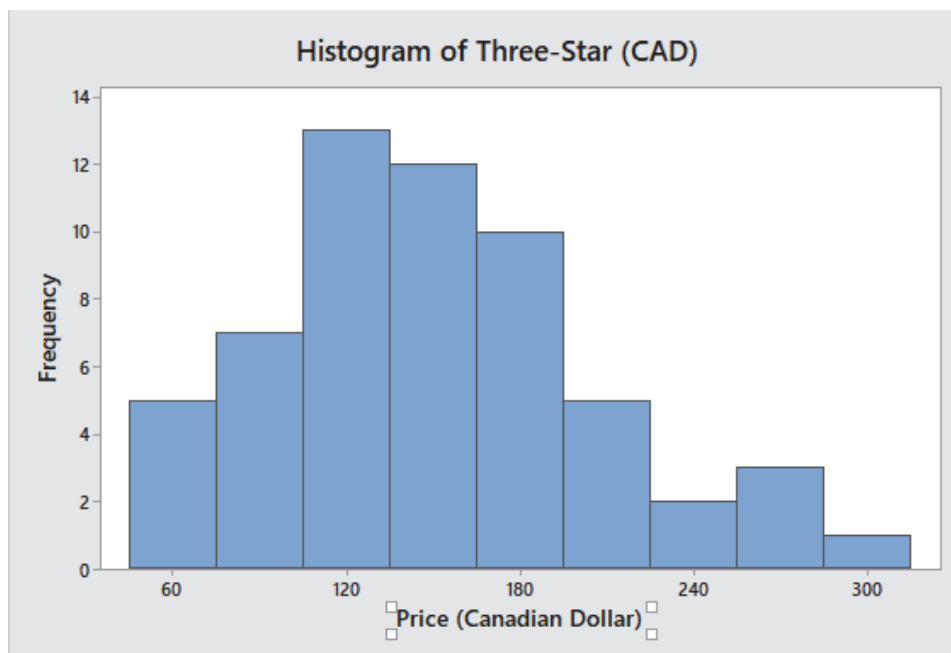
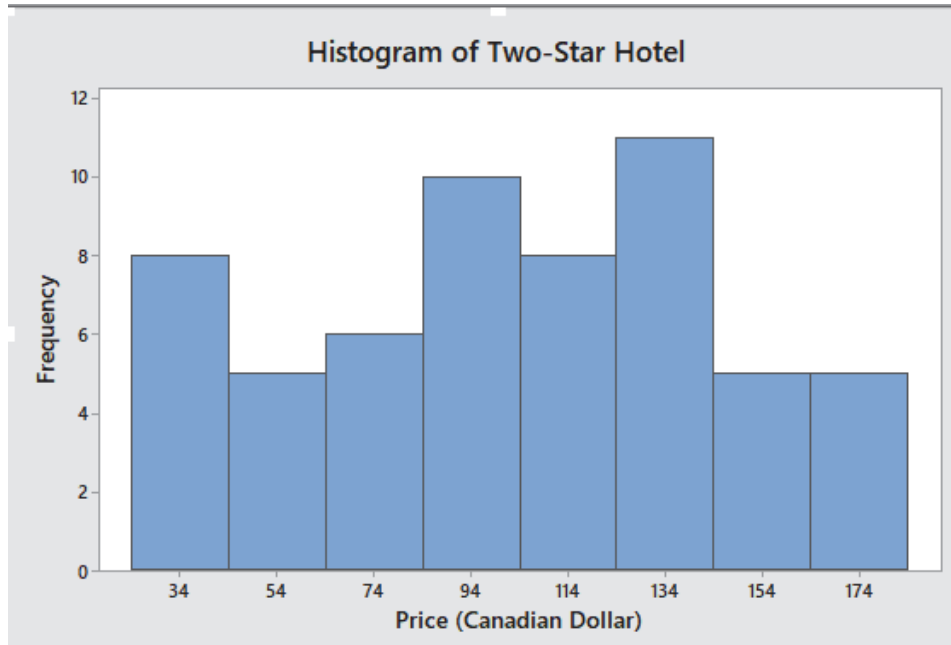
2.101 (a)

Two-Star Hotel		
Average Room Price (Canadian Dollar)	frequency	percentage
24 but less than 44	8	13.8%
44 but less than 64	5	8.6%
64 but less than 84	6	10.3%
84 but less than 104	10	17.2%
104 but less than 124	8	13.8%
124 but less than 144	11	19.0%
144 but less than 164	5	8.6%
164 but less than 184	5	8.6%
Total	58	100%

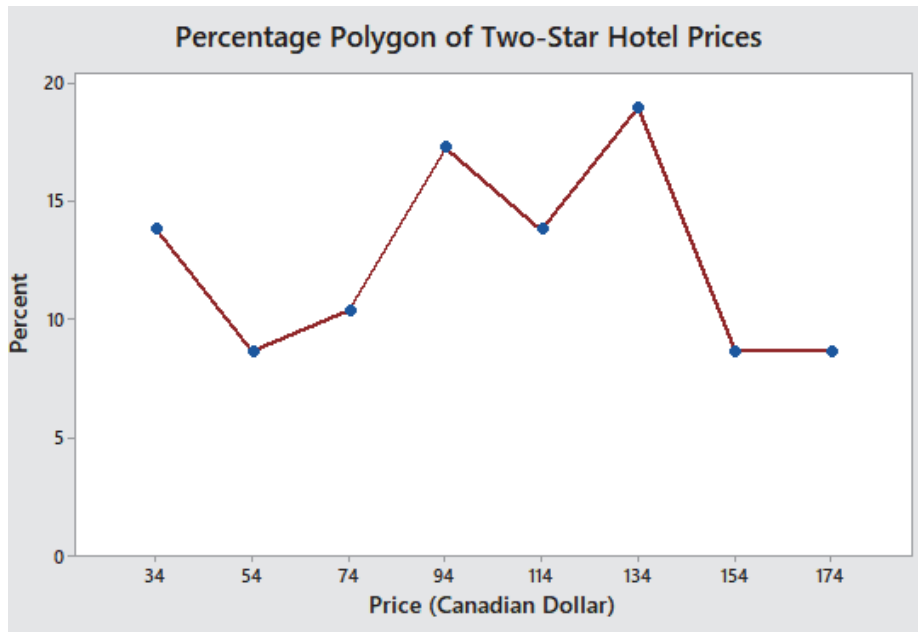
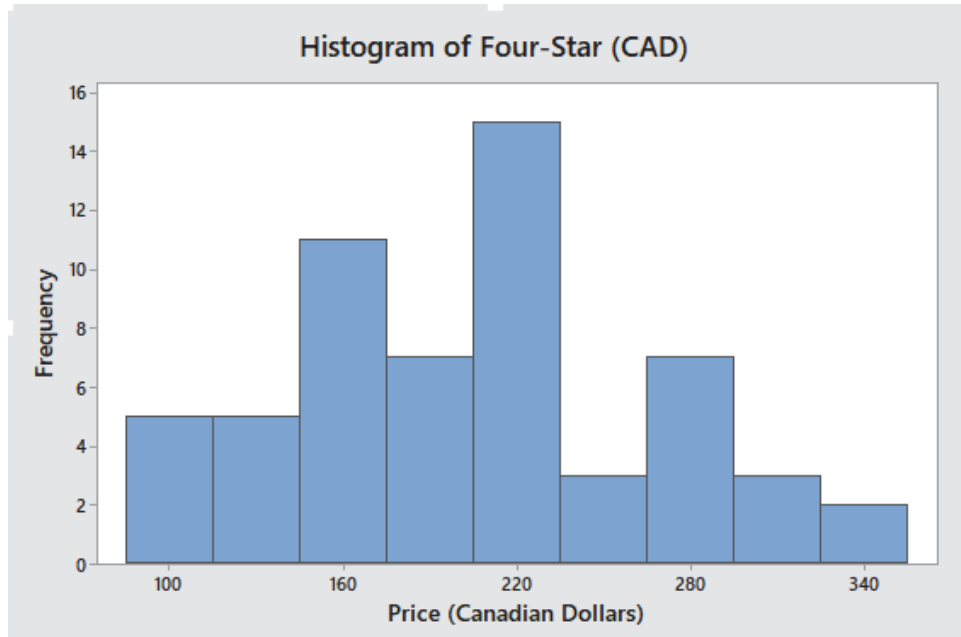
Three-Star Hotel		
Average Room Price (Canadian Dollar)	frequency	percentage
45 but less than 75	5	8.6%
75 but less than 105	7	12.1%
105 but less than 135	13	22.4%
135 but less than 165	12	20.7%
165 but less than 195	10	17.2%
195 but less than 225	5	8.6%
225 but less than 255	2	3.4%
255 but less than 285	3	5.2%
285 but less than 315	1	1.7%
Total	58	100%

Four-Star Hotel		
Average Room Price (Canadian Dollar)	frequency	percentage
85 but less than 115	5	8.6%
115 but less than 145	5	8.6%
145 but less than 175	11	19.0%
175 but less than 205	7	12.1%
205 but less than 235	15	25.9%
235 but less than 265	3	5.2%
265 but less than 295	7	12.1%
295 but less than 325	3	5.2%
325 but less than 355	2	3.4%
Total	58	100.0%

2.101 (b)
cont.

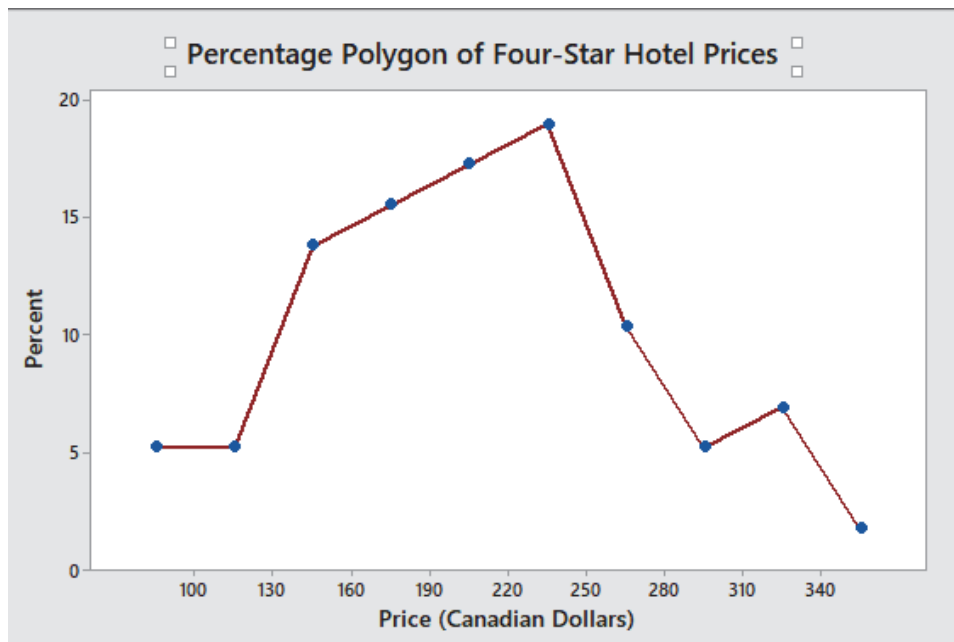
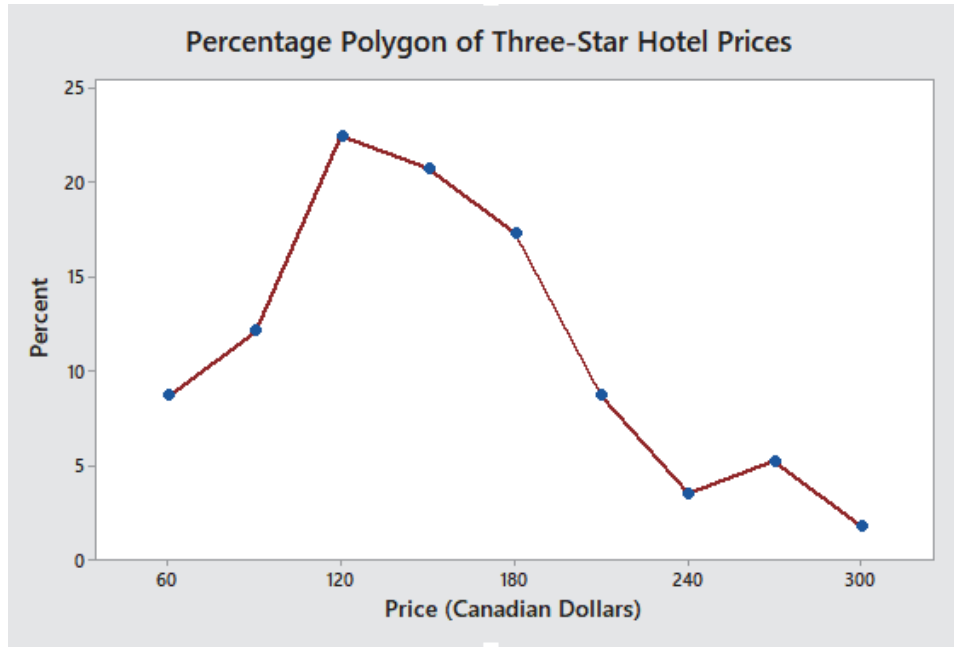


2.101 (b)
cont.



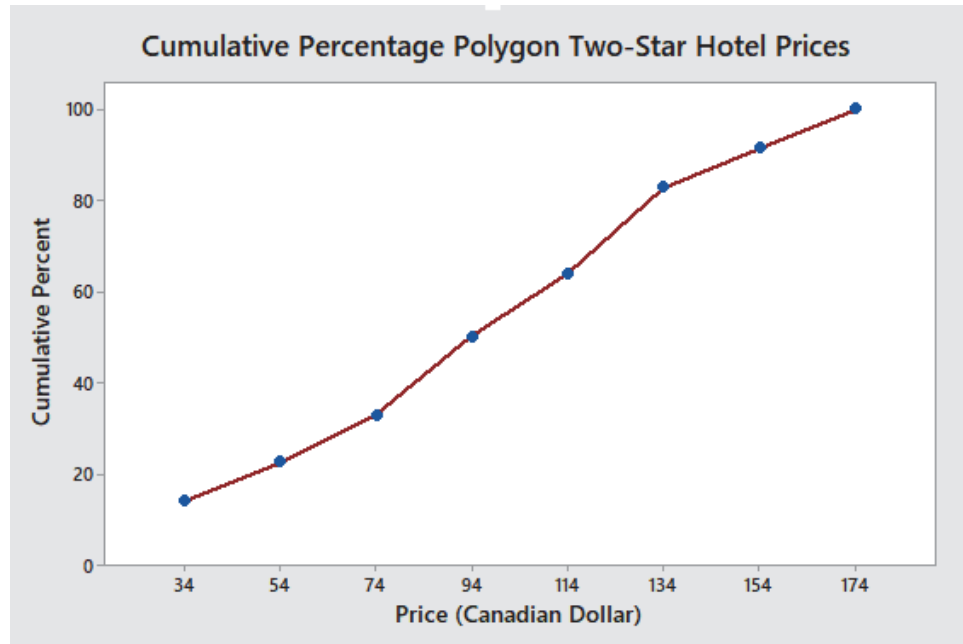
134 Chapter 2: Organizing and Visualizing Variables

2.101 (b)
cont.



2.101 (c)
cont.

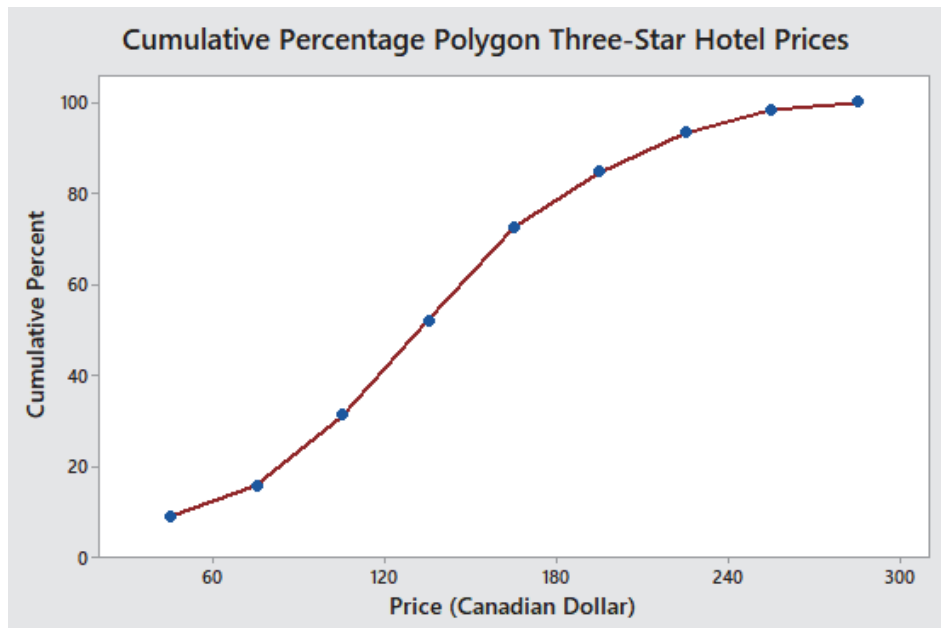
Two-Star Hotel	
Average Room Price (Canadian Dollar)	cumulative percentage
24 but less than 44	13.8%
44 but less than 64	22.4%
64 but less than 84	32.8%
84 but less than 104	50.0%
104 but less than 124	63.8%
124 but less than 144	82.8%
144 but less than 164	91.4%
164 but less than 184	100.0%



136 Chapter 2: Organizing and Visualizing Variables

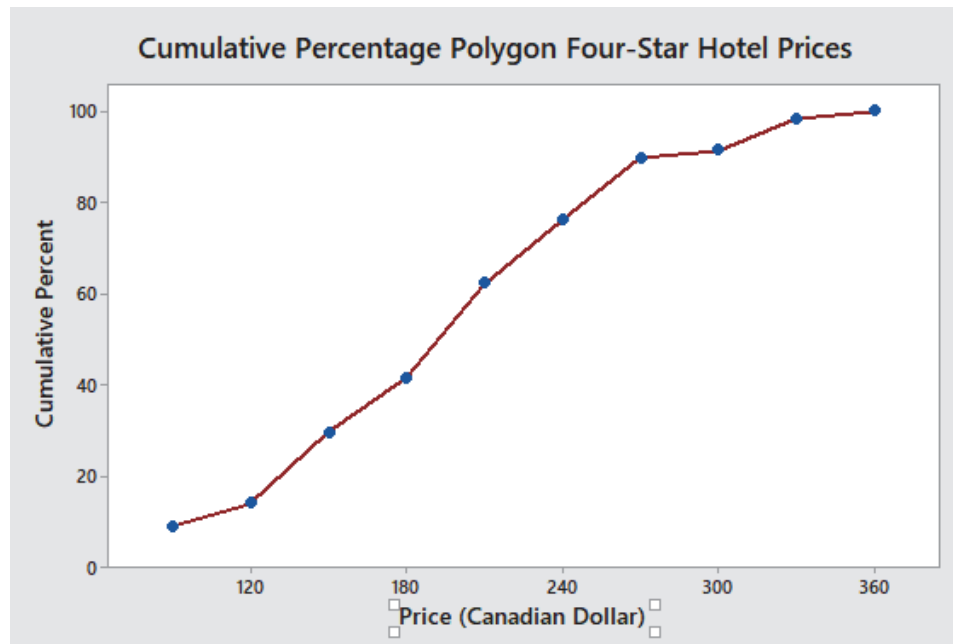
2.101 (c)
cont.

Three-Star Hotel	
Average Room Price (Canadian Dollar)	cumulative percentage
45 but less than 75	8.6%
75 but less than 105	20.7%
105 but less than 135	43.1%
135 but less than 165	63.8%
165 but less than 195	81.0%
195 but less than 225	89.7%
225 but less than 255	93.1%
255 but less than 285	98.3%
285 but less than 315	100.0%



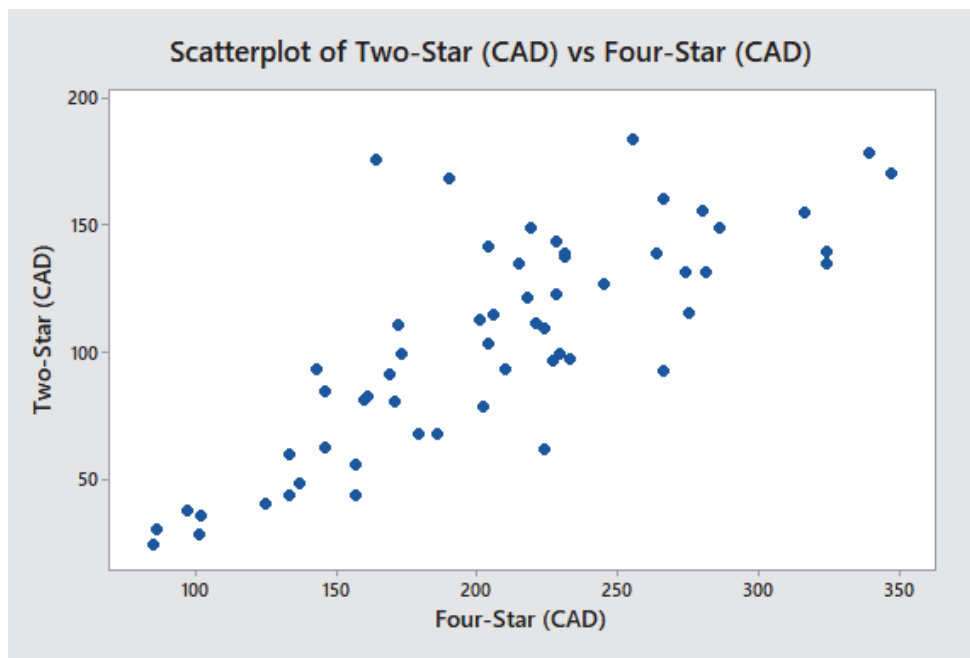
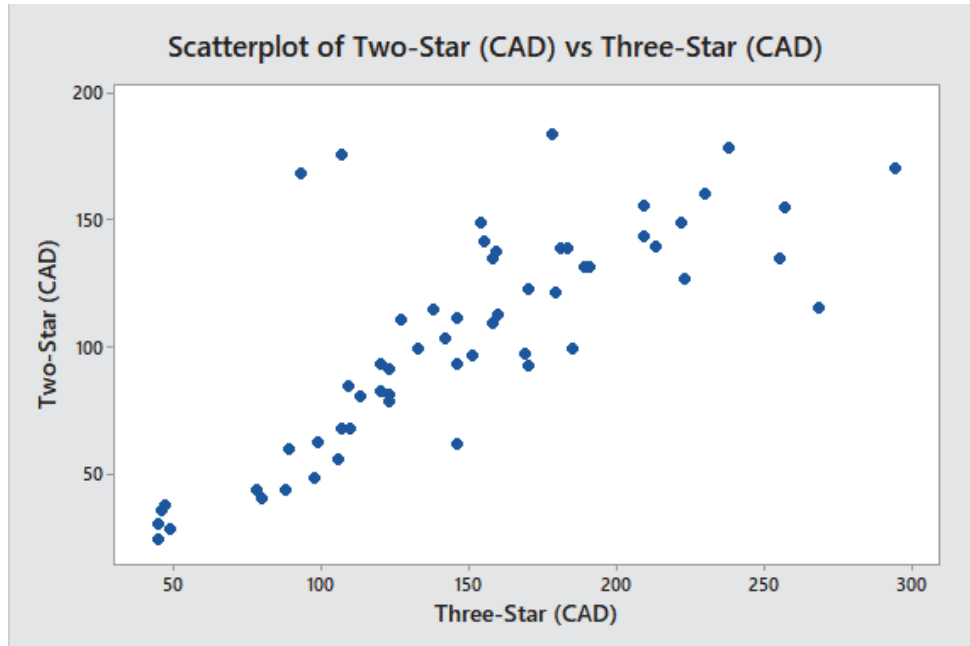
2.101 (c)
cont.

Four-Star Hotel	
Average Room Price (Canadian Dollar)	cumulative percentage
85 but less than 115	8.6%
115 but less than 145	17.2%
145 but less than 175	36.2%
175 but less than 205	48.3%
205 but less than 235	74.1%
235 but less than 265	79.3%
265 but less than 295	91.4%
295 but less than 325	96.6%
325 but less than 355	100.0%

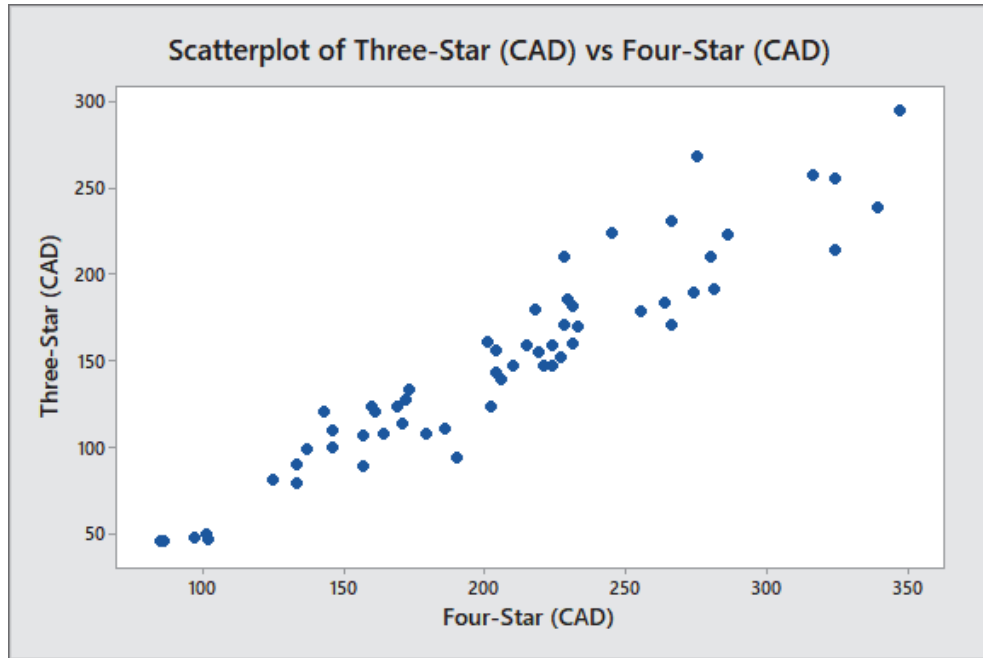


- (d) About 90% of two-star hotels have an average room price below 164 Canadian dollars. The majority of two-star hotels have room prices between 84 and 144 Canadian dollars. Approximately 93% of three-star hotels have an average room price below 255 Canadian dollars and 90% of four-star hotels have an average room price below 295 Canadian dollars. The majority of three-star hotels have room prices between 105 and 195 Canadian dollars. The majority of four-star hotels have room prices from 145 to 235 Canadian dollars.

2.101 (e)
cont.



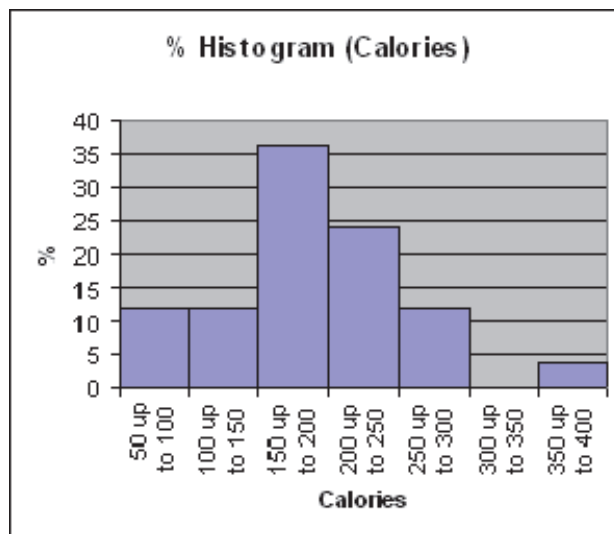
2.101 (e)
cont.



(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)

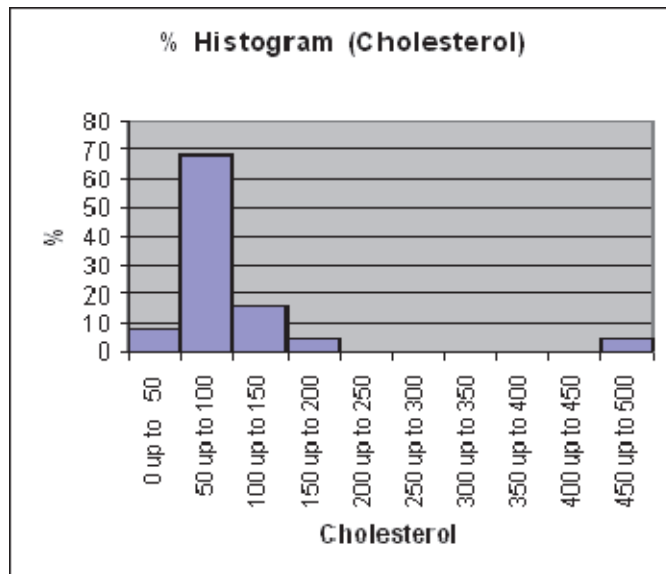
Calories	Frequency	Percentage	Percentage Less Than
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100



140 Chapter 2: Organizing and Visualizing Variables

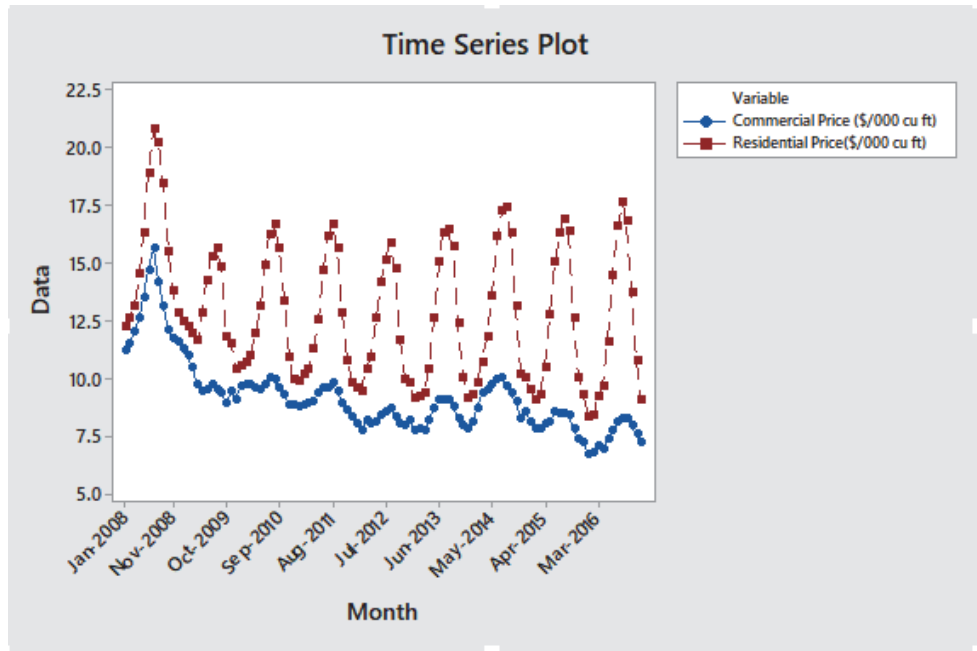
2.102 (b)
cont.

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



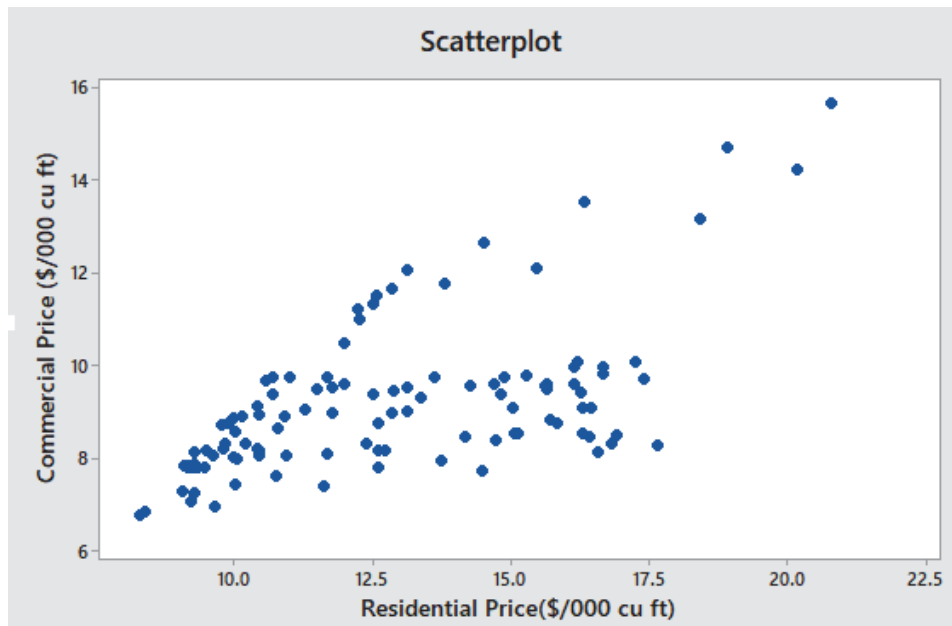
(c) The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled 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 commercial average price was highest in the summer of 2008 and has since declined. The residential average price of gasoline in the United States is higher in the summer in general.

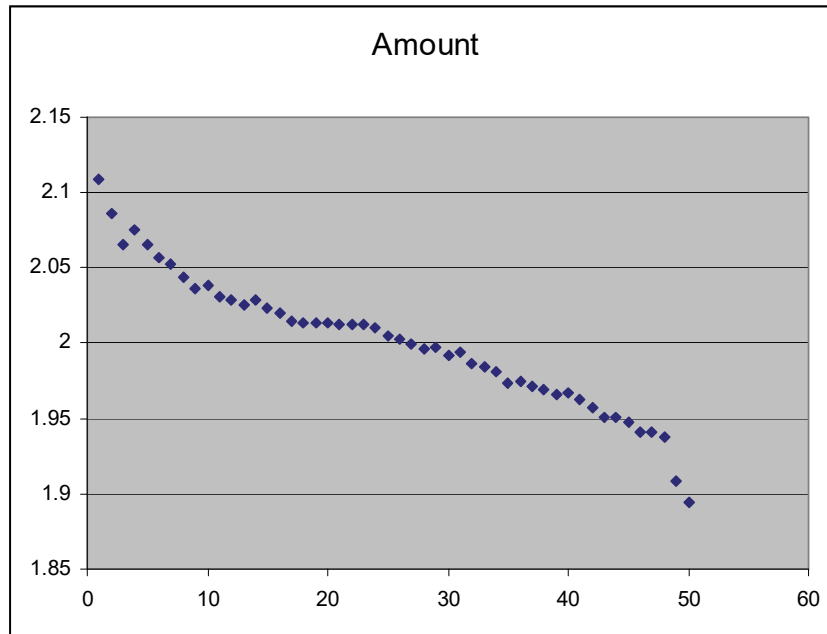
(c)



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

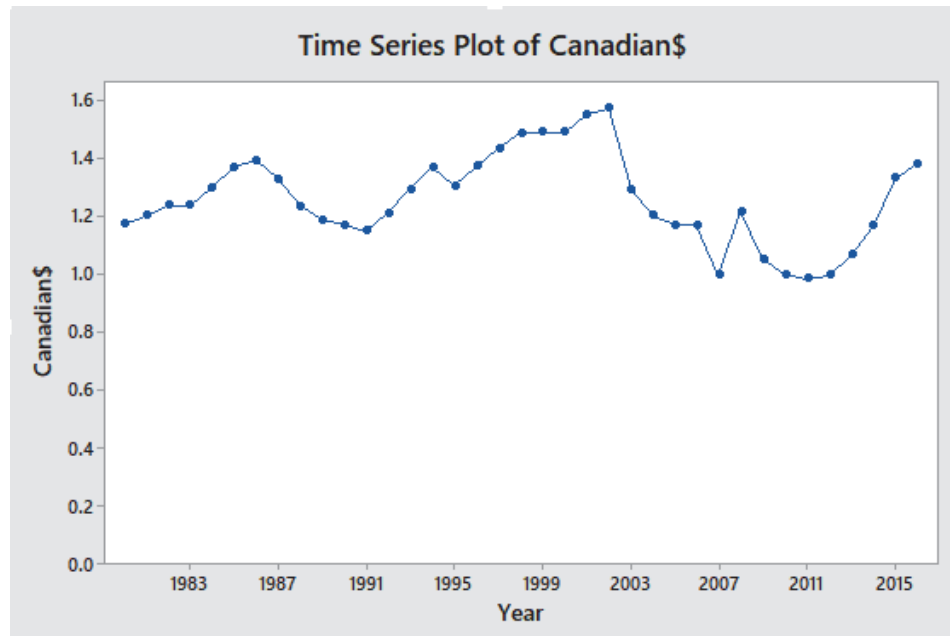
142 Chapter 2: Organizing and Visualizing Variables

2.104 (a)

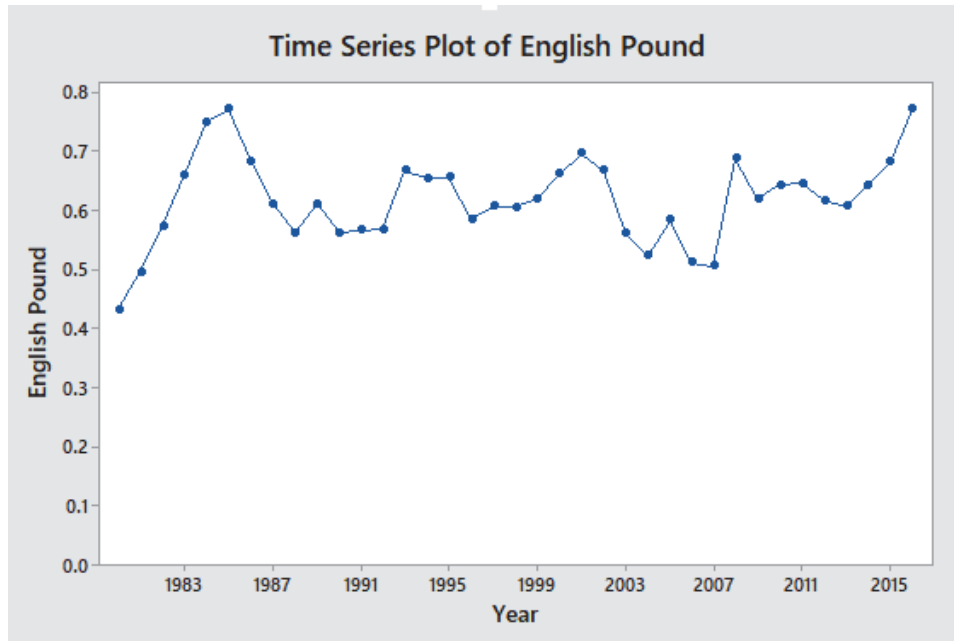
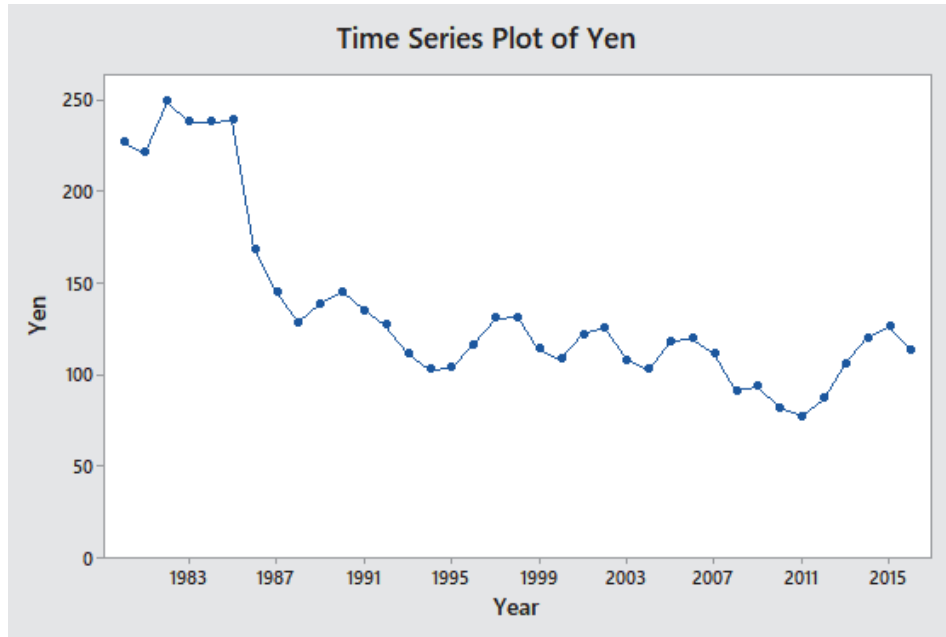


- (b) There is a downward trend in the amount filled.
- (c) The amount filled in the next bottle will most likely be below 1.894 liter.
- (d) The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution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

2.105 (a)

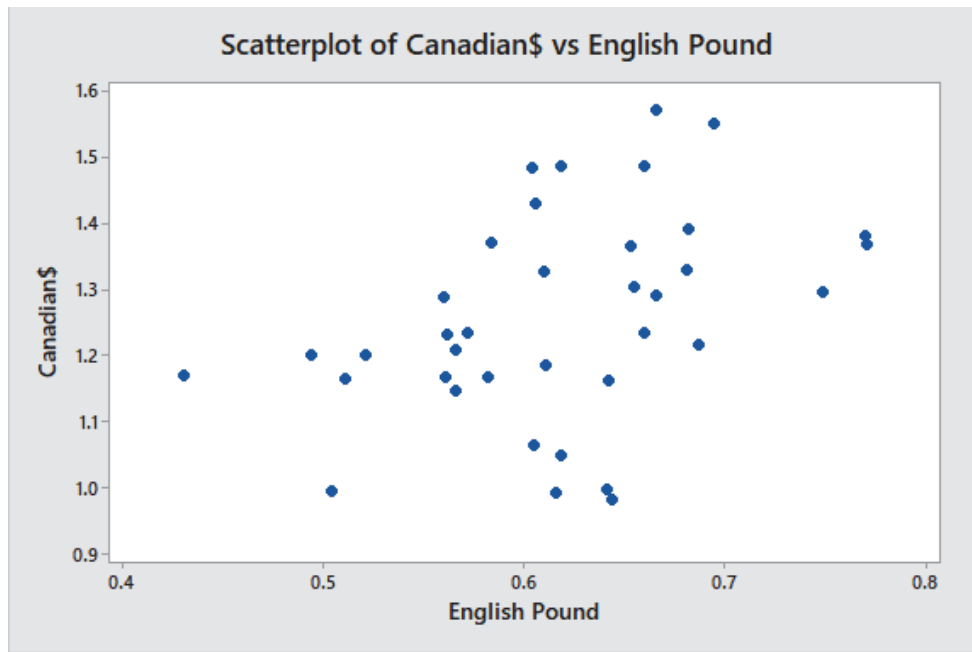
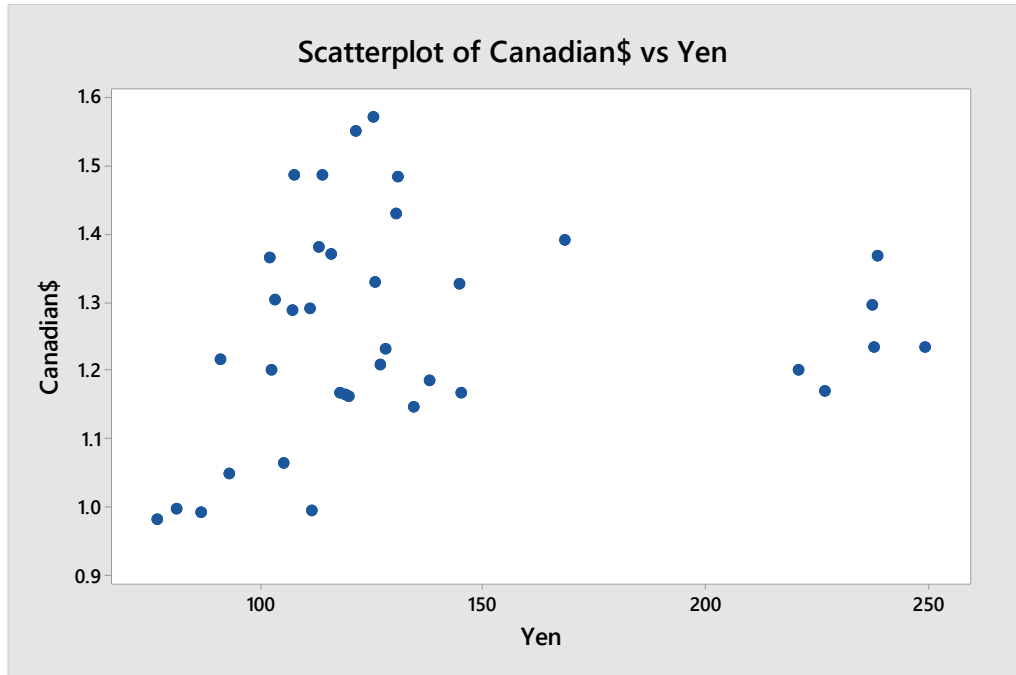


2.105 (a)
cont.

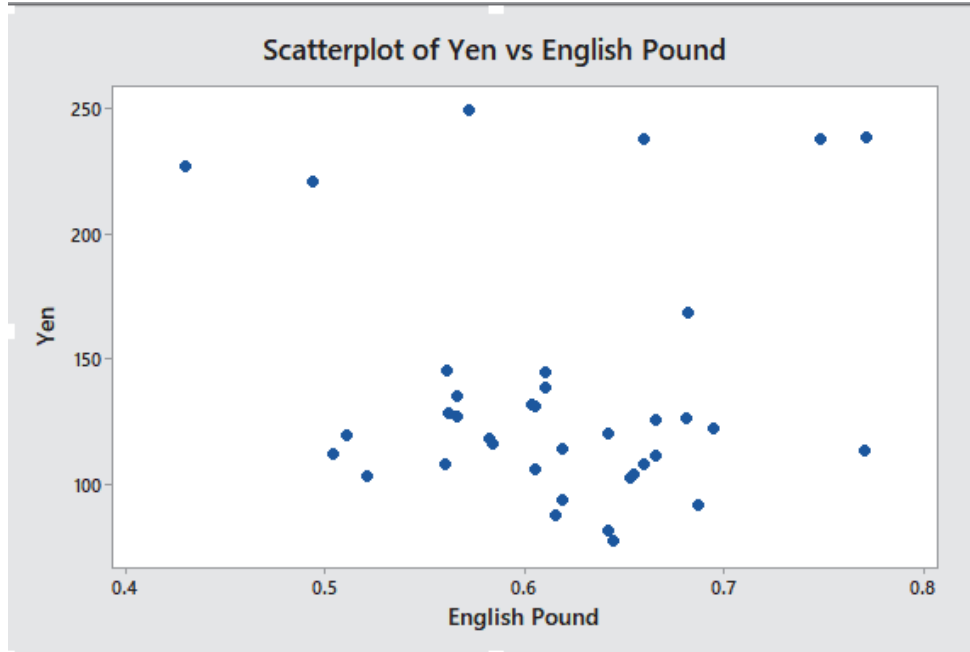


- (b) The Japanese yen had depreciated against the U.S. dollar since 1982 while the Canadian dollar appreciated gradually from 1980 to 1987 and from 1991 to 2002 and then started to depreciate until 2011. The English pound to U.S. dollar's exchange rate has been quite stable since 1983.
- (c) The U.S. dollar has appreciated against the Japanese yen since 1980 and appreciated against the Canadian dollar since 2002 in general while the exchange rate against the English bound has been stable in general.

2.105 (c)
cont.



2.105 (c)
cont.

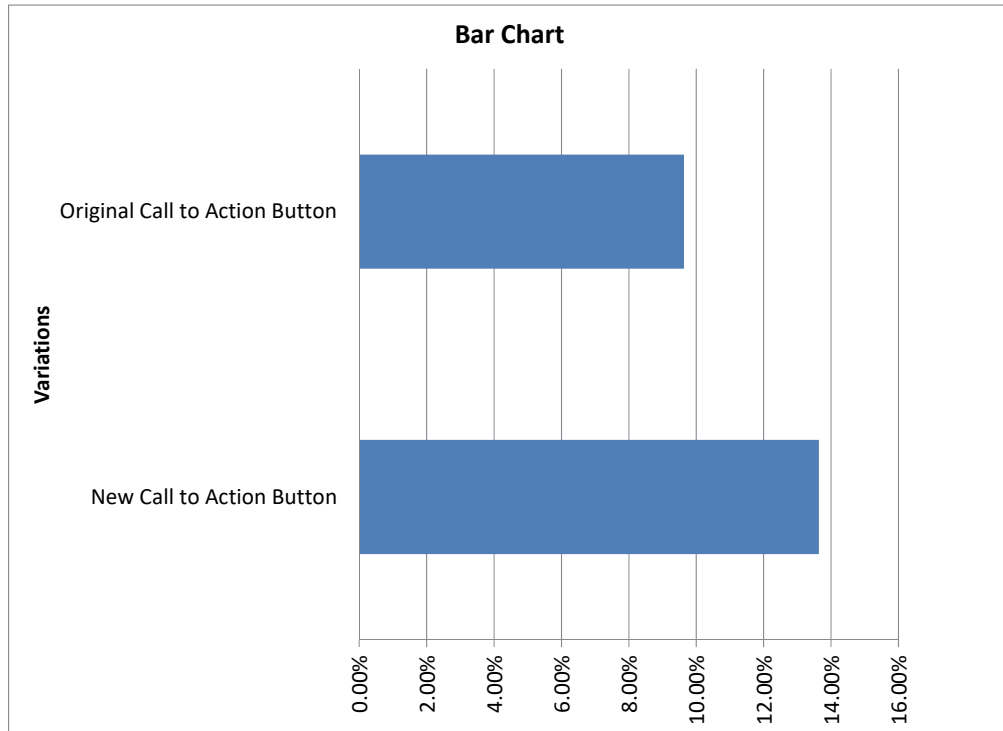


(e) There is not any obvious relationship between the Canadian dollar and Japanese yen in terms of the U.S. dollar nor any relationship between the Japanese yen and English pound. There is a slightly positive relationship between the Canadian dollar and English pound which reflects the fact that when the Canadian dollar appreciated against the U.S. dollar, so did the English pound.

2.106 (a)

Variations	Percentage of Download
Original Call to Action Button	9.64%
New Call to Action Button	13.64%

2.106 (b)
cont.

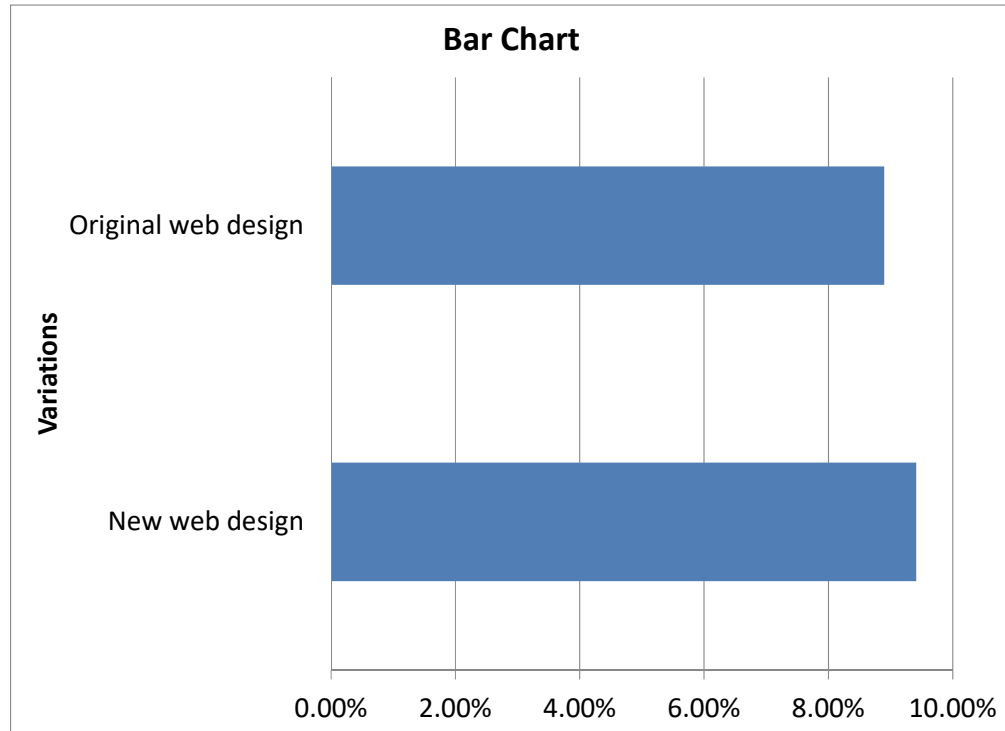


(c) The New Call to Action Button has a higher percentage of downloads at 13.64% when compared to the Original Call to Action Button with a 9.64% of downloads.

(d)

Variations	Percentage of Downloads
Original web design	8.90%
New web design	9.41%

2.106 (e)
cont.



- (f) The New web design has only a slightly higher percentage of downloads at 9.41% when compared to the Original web design with an 8.90% of downloads.
- (g) The New web design is only slightly more successful than the Original web design while the New Call to Action Button is much more successful than the Original Call to Action Button with about 41% higher percentage of downloads.

(h)

Call to Action Button	Web Design	Percentage of Downloads
Old	Old	8.30%
New	Old	13.70%
Old	New	9.50%
New	New	17.00%

- (i) The combination of the New Call to Action Button and the New web design results in slightly more than twice as high a percentage of downloads than the combination of the Old Call to Action Button and Old web design.
- (j) The New web design is only slightly more successful than the Original web design while the New Call to Action Button is much more successful than the Original Call to Action Button with about 41% higher percentage of downloads. However, the combination of the New Call to Action Button and New web design results in more than twice as high a percentage of downloads than the combination of the Old Call to Action Button and Old web design.

2.107 Class project – answers will vary depending on student responses.

2.108 Class project – answers will vary depending on student responses.

- 2.109 A descriptive analysis of the weight of the pallets of the Boston shingles revealed that the average weight was 3124.2 pounds with a standard deviation of 34.7. The average weight of 3124.2 pounds was 74.2 pounds above the expected minimum weight of 3,050 pounds. An analysis of the Vermont shingles revealed that the average weight was 3704.0 pounds with a standard deviation of 46.7. The average weight of 3704.0 pounds was 104 pounds above the expected minimum weight of 3,600 pounds. The below table includes a number of descriptive statistics for the two shingle types.

Descriptive Statistics: Boston, Vermont

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
Boston	368	0	3124.2	1.81	34.7	3044.0	3098.0	3122.0	3146.0	3266.0	0.53
Vermont	330	38	3704.0	2.57	46.7	3566.0	3670.0	3704.0	3732.0	3856.0	0.29

A frequency distribution of the Boston shingles revealed that 0.54% of the pallets were underweight and 0.27% were overweight. A frequency distribution of the Vermont shingles revealed that 1.21% of the shingles were underweight and 3.94% were overweight. The complete results are provided in the below frequency distributions.

Frequencies (Boston)

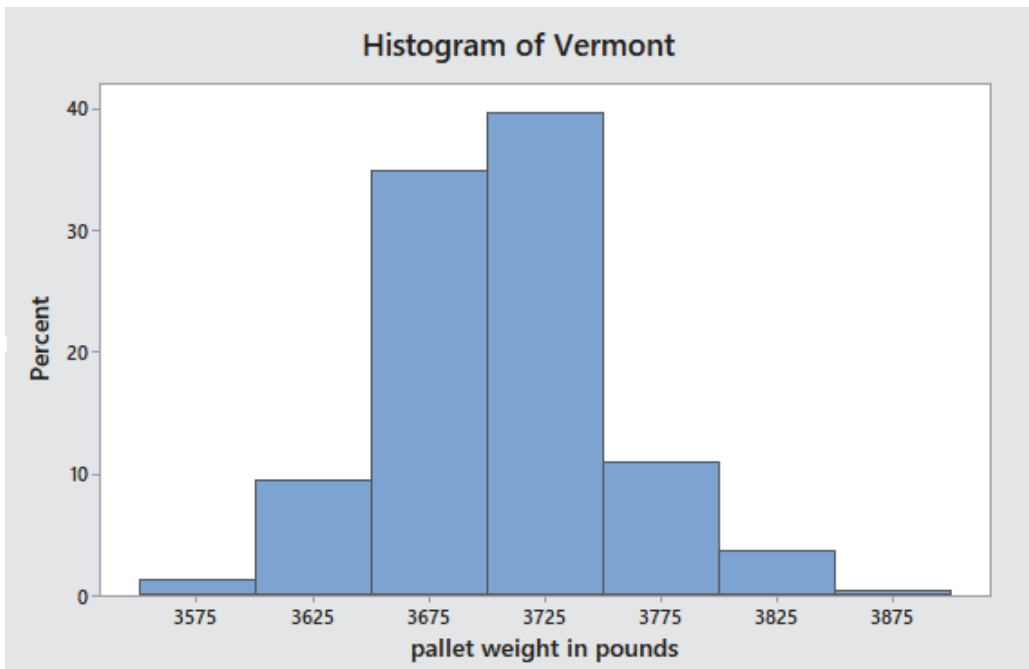
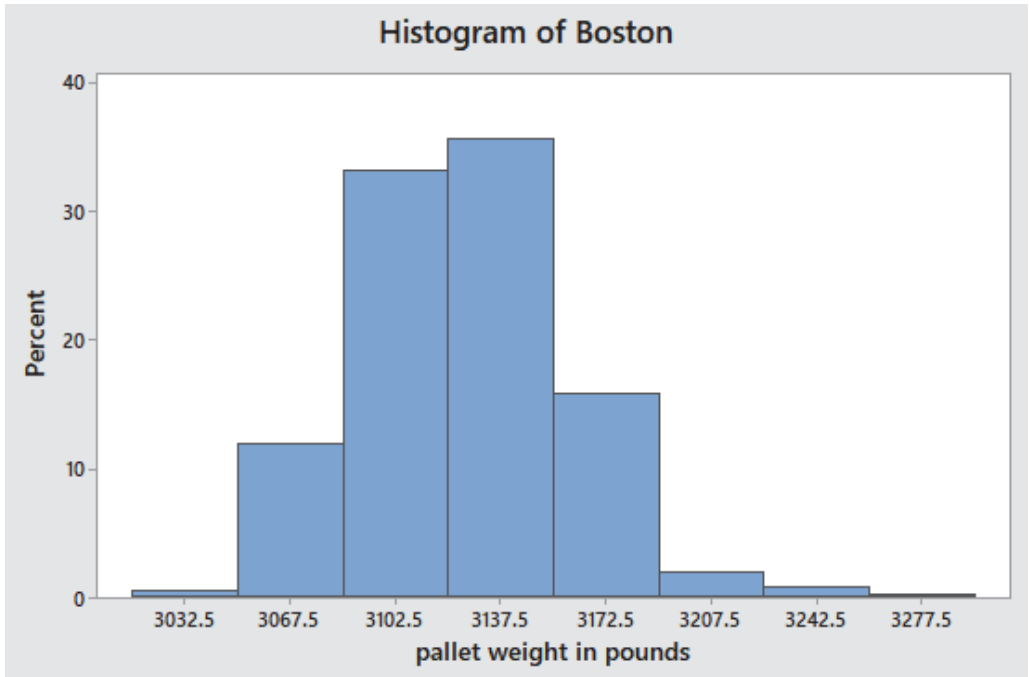
Weight (Boston)	Frequency	Percentage
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1	0.27%

Frequencies (Vermont)

Weight (Vermont)	Frequency	Percentage
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%

Histogram graphs of the Boston shingles and the Vermont shingles, shown below, revealed that the weights of the pallets appeared to be consistent with a normal distribution. In both cases, there was slight right skewness with the Boston shingles having slightly more right skewness than the Vermont shingles.

2.109
cont.



The results of the above analyses reveal that both shingle types generally met pallet weight expectations with less than 1% of the Boston shingles weighing outside of the expected parameters and just over 5% of the Vermont shingles weighing outside of the expected parameters. The results suggest that the manufacturer should consider implementation of parameter compliance strategies for the Vermont shingles.

