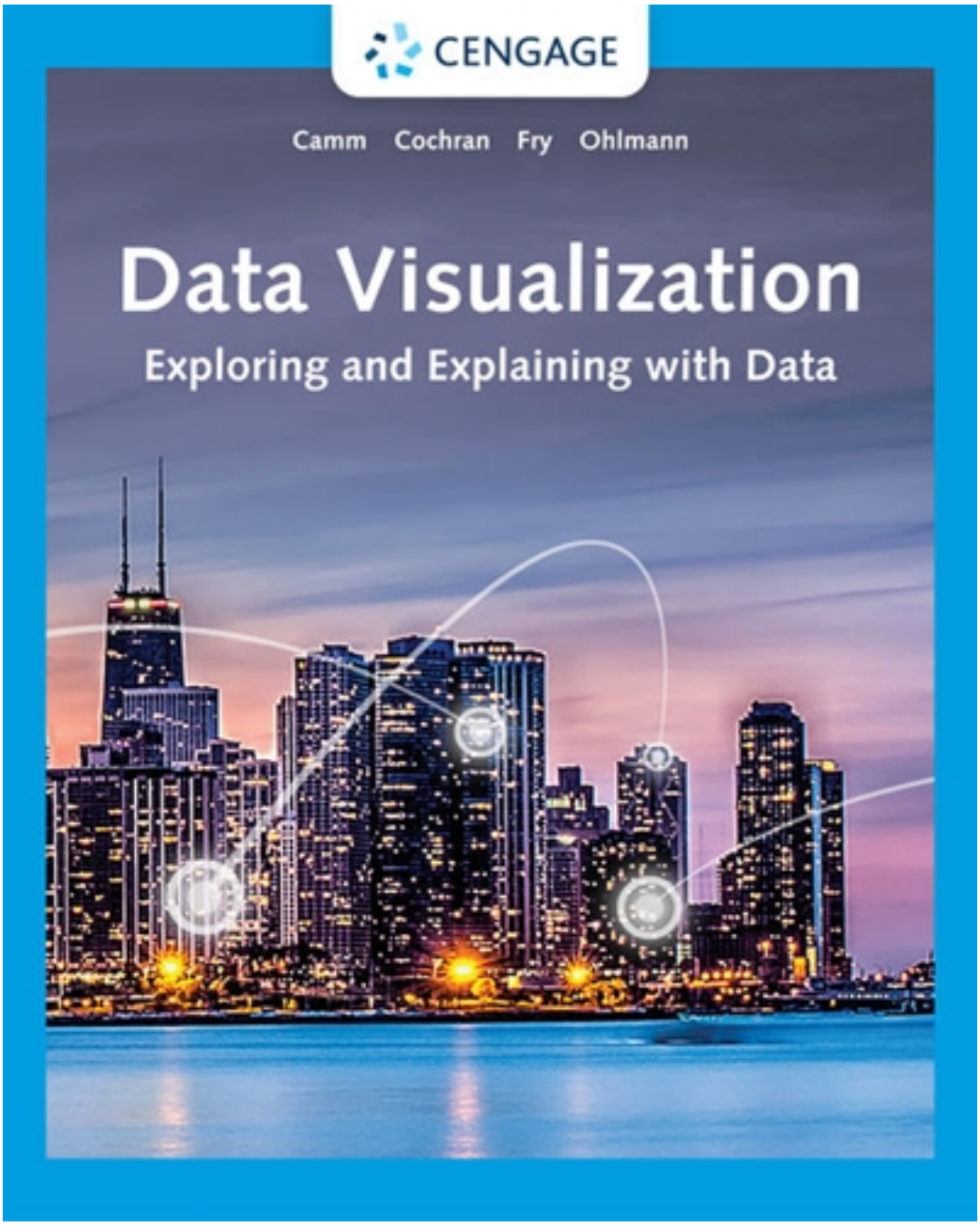
Solutions for Data Visualization 1st Edition by Camm

CLICK HERE TO ACCESS COMPLETE Solutions



Solutions

Solution and Answer Guide

Camm/Cochran/Fry/Ohlmann, Data Visualization 2022, 9780357631348; Chapter 1: Introduction

Table of Contents

End of Section Problem Solutions

1. **Types of Analytics.** Indicate which type of analytics (descriptive, predictive, or prescriptive analytics) each of the following represents. LO 1
    a. a data dashboard
    b. a model that finds the production schedule that minimizes overtime
    c. a model that forecasts sales for the next quarter
    d. a bar chart
    e. a model that allocates your financial investments to achieve your financial goal

    Answer:
    a. descriptive
    b. prescriptive
    c. predictive
    d. descriptive
    e. prescriptive

2. **Transportation Planning**. An analytics professional is asked to plan the shipment of a product for the next quarter. She employs the following process:
    a. For each of the 12 distribution centers, she plots the quarterly demand for the product over the last three years.
    b. Based on the plot for each distribution center, she develops a forecasting model to forecast demand for next quarter for each distribution center.
    c. She takes the forecast for next quarter for each distribution center and inputs those forecasts, along with the capacities of the company's four factories and transportation rates from each factory to each distribution center, into an optimization model. The optimization model suggests a shipping plan that minimizes the cost of how to satisfy the forecasted demand from the company's four different factories to the distribution centers.

    Describe the type of analytics being utilized in each of the three steps outlined above. LO 1

    Answer:
    a. Data visualization is descriptive analytics.
    b. A forecasting model is predictive analytics.
    c. An optimization mode is prescriptive analytics.

    Feedback: Text.

3. ***Wall Street Journal* Subscriber Characteristics**. A *Wall Street Journal* subscriber survey asked a series of questions about subscriber characteristics and interests. State whether each of the following questions provides categorical or quantitative data. LO 2
    a. What is your age?
    b. Are you male or female?
    c. When did you first start reading the *WSJ*? High school, college, early career, midcareer, late career, or retirement?
    d. How long have you been in your present job or position?
    e. What type of vehicle are you considering for your next purchase? Nine response categories for this question include sedan, sports car, SUV, minivan, and so on.

    Answer:
    a. quantitative
    b. categorical
    c. categorical
    d. quantitative
    e. categorial

    Feedback: Text.

4. **Comparing Smartwatches.** *Consumer Reports* provides product evaluations for its subscribers. The following table shows data from *Consumer Reports* for five smartwatches on the following characteristics:

    Overall Score—a score awarded for a variety of performance factors

    Price—the retail price

    Recommended—does *Consumer Reports* recommend purchasing the smartwatch based on performance and strengths?

    Best Buy—if *Consumer Reports* recommends purchasing the smartwatch, does it also consider it a "best buy" based on a blend of performance and value?

    For each of the four pieces of data, indicate whether the data are quantitative or categorical and whether the data are cross-sectional or time series. LO 2

Make	Overall Score	Recommended	Best Buy	Price
Apple Watch Series 5	84	Yes	No	$395
Fitbit Versa 2	78	Yes	Yes	$200
Garmin Venu	77	Yes	No	$350
Fitbit Versa Lite	65	No	No	$100

    Answer:
    Overall Score and Price are quantitative. Recommended and Best Buy are categorical (both are yes or no). All of these data are cross-sectional.

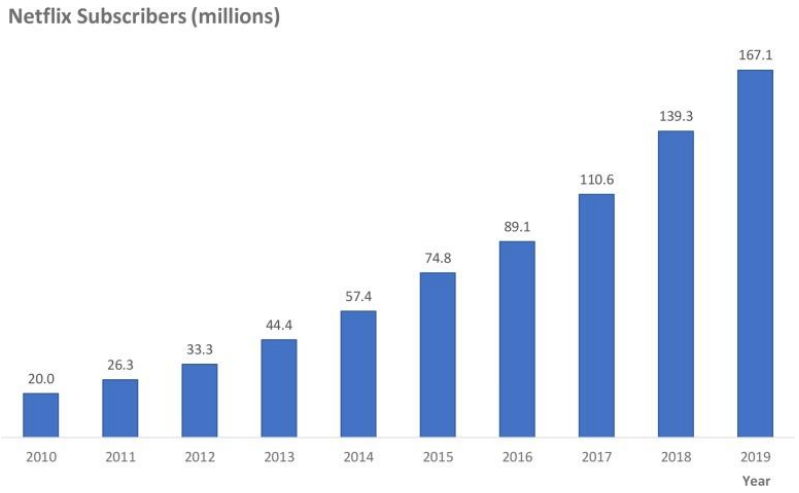
    Feedback: Text.

5. **House Price and Square Footage.** Suppose we want to better understand the relationship between house price and square footage of the house, and we have collected house price and square footage for 75 houses in a particular neighborhood of Cincinnati, Ohio, from the Zillow website on January 3, 2021. LO 2, 3
    a. Are these data quantitative or categorical?
    b. Are these data cross-sectional or times series?
    c. Which of the following type of chart would provide the best display of these data? Explain your answer.
        i. Bar chart
        ii. Column chart
        iii. Scatter chart

    Answer:
    a. House price and square footage are quantitative.
    b. The data were collected at a single point in time, so these data are cross-sectional.
    c. iii Bar charts and column charts are better for categories. Here we do not have categories, we have pairs of numerical data for each of 75 houses (price, square footage). Therefore, a scatter chart with square footage as the horizontal axis and price as the vertical axis would be appropriate.

    Feedback: Text.

6. **Netflix Subscribers.** The following chart displays the total number of Netflix subscribers from 2010 to 2019. LO 1, 2, 3



    a. Are these data quantitative or categorical?
    b. Are these data cross-sectional or time series?
    c. What type of chart is this?

Answer:

a. Number of Netflix subscribers is quantitative.
b. Since these data are collected at several points in time, they are time series data.
c. This is a column chart.

Feedback: Text.

7. **U.S. Netflix Subscribers.** Refer to the previous problem. Suppose that in addition to the total number of Netflix subscribers, we have the number of those subscribers by year for the years 2010–2019 who live in the United States. Our message is to emphasize how much of the growth is coming from the United States. Which of the following types of charts would best display the data? Explain your answer. LO 2, 3

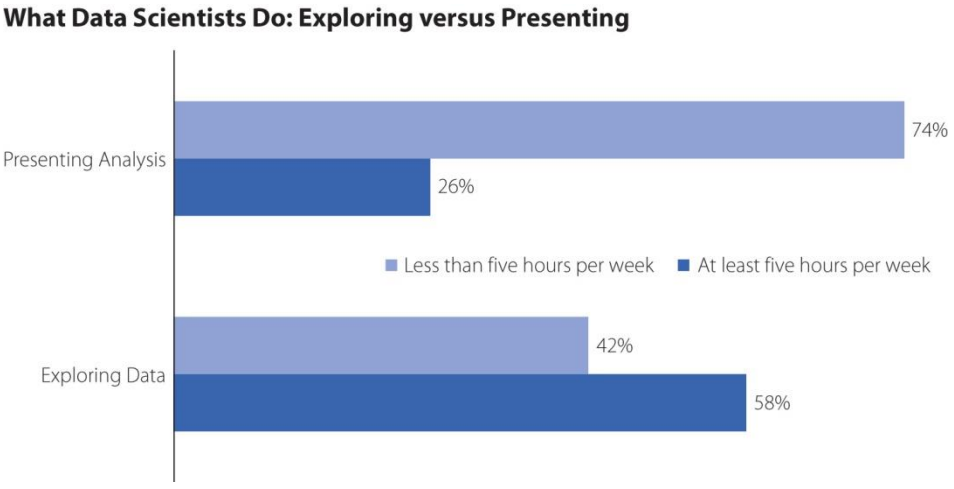
   i. Bar chart
   ii. Clustered column chart
   iii. Stacked column chart
   iv. Stock chart

   Answer: iii

   A stacked column chart is most appropriate here as it will show the total number of subscribers (height of the column) but also the same for U.S. subscribers (denoted by color within a column).

   Feedback: Text.

8. **How Data Scientists Spend Their Day.** *The Wall Street Journal* reported the results of a survey of data scientists. The survey asked the data scientists how they spend their time. The following chart shows the percentage of respondents who answered less than five hours per week or at least five hours per week for the amount of time they spend on exploring data and on presenting analyses. LO 2, 3, 4



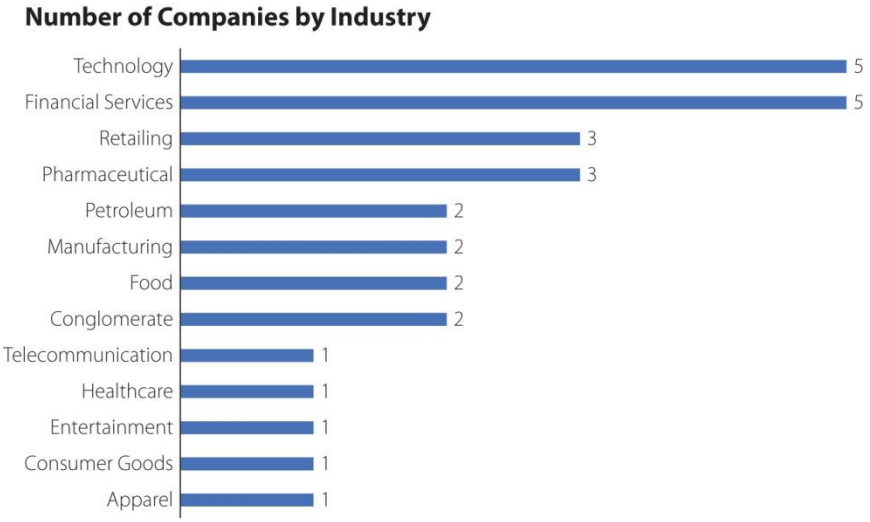
   a. Are these data quantitative or categorical?
   b. Are these data cross-sectional or time series?
   c. What type of chart is this?
   d. What conclusions can you make based on this chart?

Answer:

a. These data are categorical: exploring data and presenting data; less than or equal to four hours per week or at least five hours per week.
b. These data are cross sectional since the survey was administered to a person at one point in time.
c. The chart is a clustered bar chart (also known as a side-by-side bar chart).
d. Data scientists spend more time exploring data than presenting results.

Feedback: Text.

9. **Industries in the Dow Jones Industrial Index.** Refer to the data on the Dow Jones Industrial Index given in Table 1.3. The following chart displays the number of companies in each industry that make up this index. LO 3



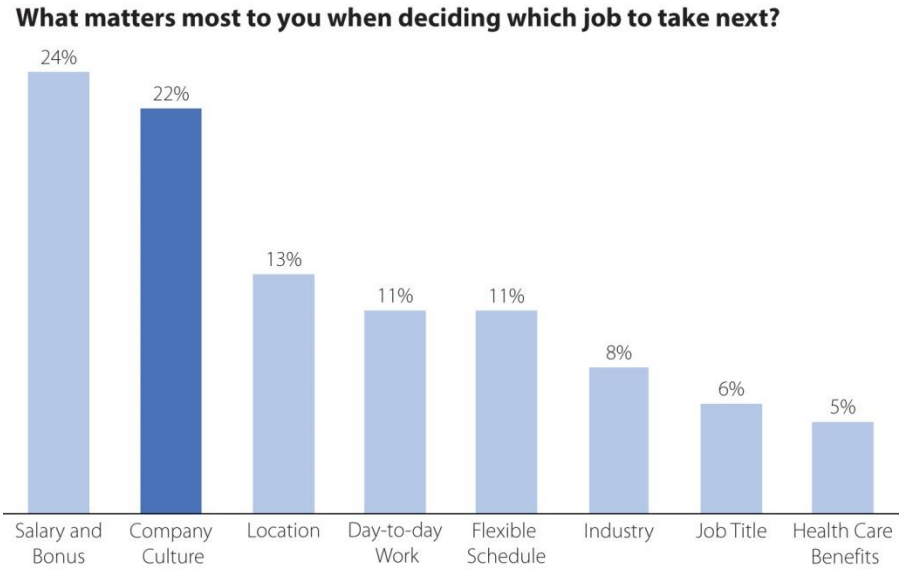
a. What type of chart is this?
b. Which industry has the highest number of companies in the Dow Jones Industrial Index?

Answer:

a. This is a bar chart.
b. Financial Services and Technology each have five companies in the Dow Jones Industrial Index.

Feedback: Text.

10. **Job Factors.** The following chart is based on the same data used to construct Figure 1.3. The data are percentages of respondents to a survey who listed various factors as most important when making a job decision. LO 3, 4

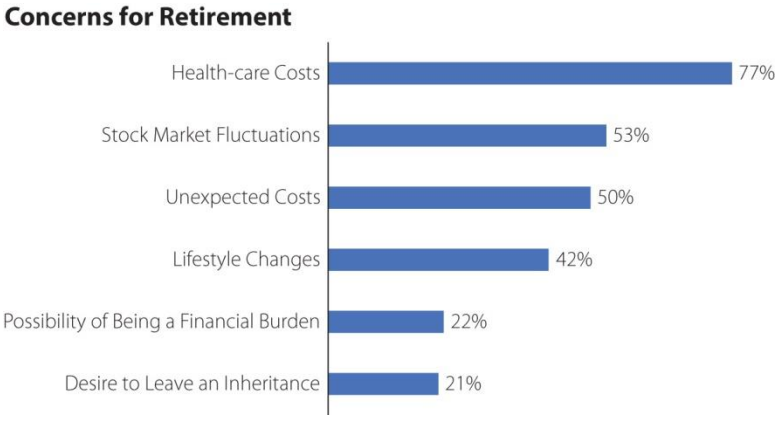


a. What type of chart is this?
b. What is the fifth most-cited factor?

Answer:
a. This is a column chart.
b. The fifth most-cited factor is Industry (Day-to-day Work and Flexible Schedule tie for fourth).

Feedback: Text.

11. **Retirement Financial Concerns.** The results of the American Institute of Certified Public Accountants' *Personal Financial Planning Trends Survey* indicated 48% of the clients had concerns about outliving their money. The top reasons for these concerns and the percentage of respondents who cited the reason were as follows. LO 3, 4

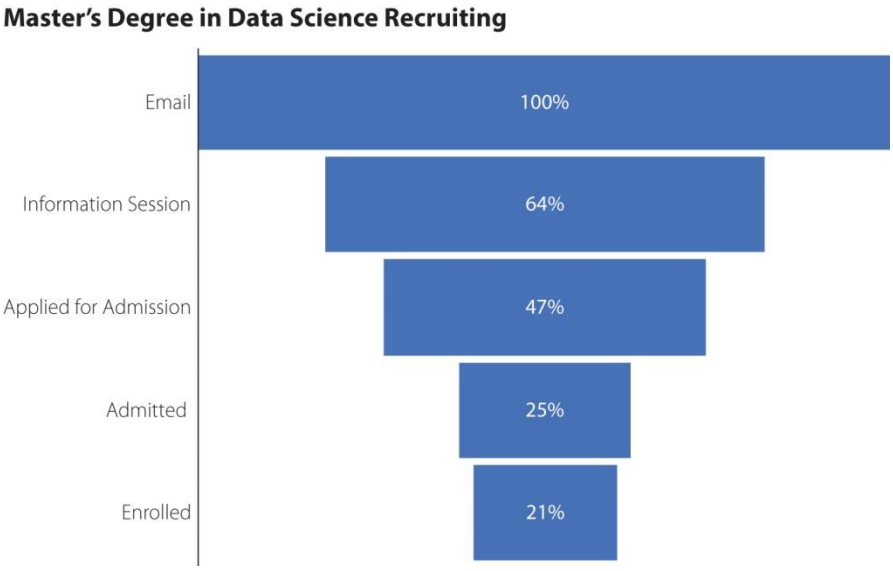


a. What type of chart is this?
b. Only 48% of the survey respondents had financial concerns about retirement (outliving their money). What percentage of the total people surveyed had retirement health-care cost concerns?

Answer:
a. This is a bar chart.
b. 77% of 48% is .77x.48 = .3696 or 36.96%.

Feedback: Text.

12. **Master's Degree Program Recruiting.** The recruiting process for a full-time master's program in data science consists of the following steps. The program director obtains email addresses of undergraduate seniors who have taken the Graduate Record Exam (GRE) and expressed an interest in data science. An email inviting the students to an online information session is sent. At the information session, faculty discuss the program and answer questions. Students apply through a web portal. An admissions committee makes an offer of admission (or not) along with any financial aid. If the person is admitted, the person either accepts or rejects the offer. Consider the following chart. LO 3, 4



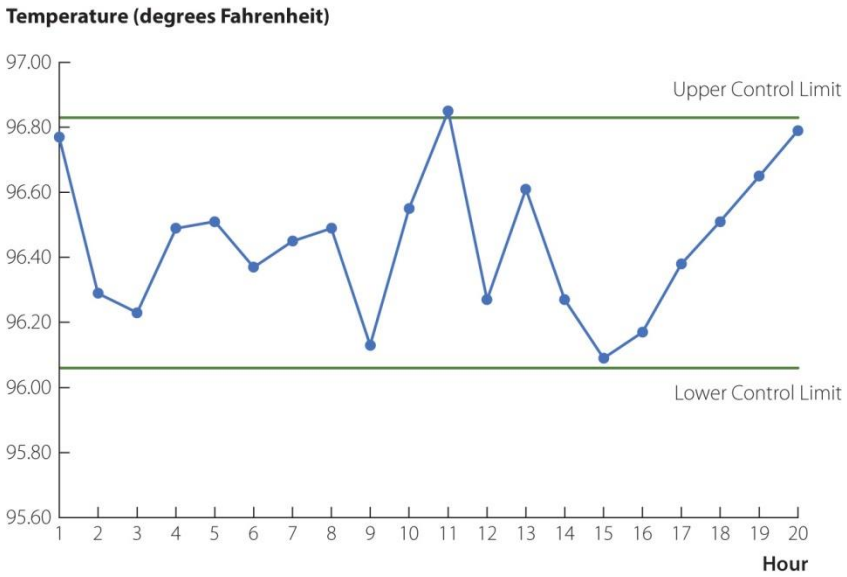
a. What type of chart is this?
b. Which of the following is the correct interpretation of the 21% for Enrolled?
   i. Of those who were sent an email, 21% enrolled.
   ii. Of those who were admitted, 21% enrolled.
   iii. Of those who applied for admission, 21% enrolled.
   iv. None of the above

Answer:

a. This is a funnel chart.
b. i.

Feedback: Text.

13. **Chemical Process Control.** The following chart is a quality control chart of the temperature of a chemical manufacturing process. What observations can you make about the process? LO 3

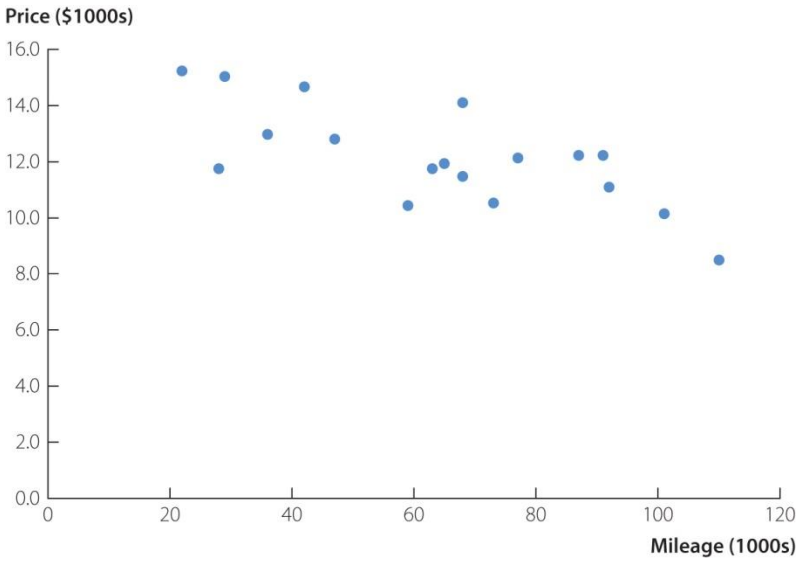


Answer:

The process appears to be in control. Only one point is outside the control limits (hour 11), and it barely exceeds the upper limit. However, the last five data points exhibit an upward trend and might be cause for concern. The temperature should be monitored closely moving forward.

Feedback: Text.

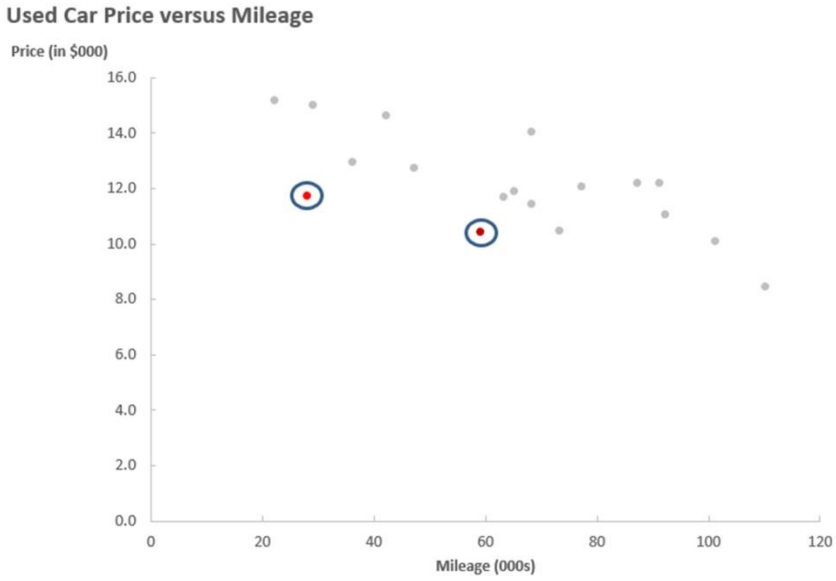
14. **Buying a Used Car.** The following chart shows data for a sample of 18 used cars of the same brand, model, and year. LO 2, 3, 4



a. Are these data quantitative or categorical?
b. What type of chart is this?
c. How might you use this chart to find a used car to purchase?

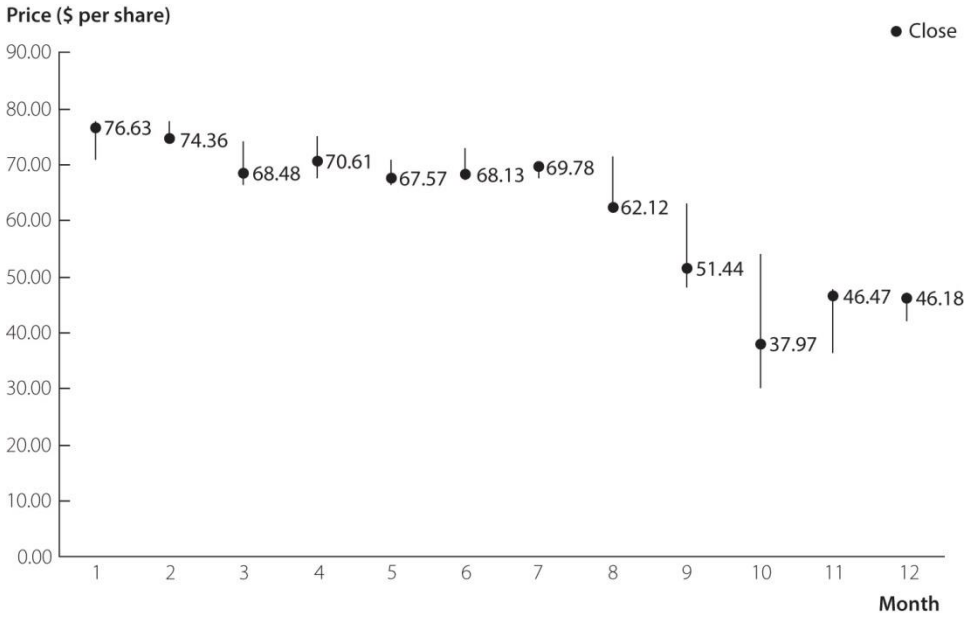
Answer:

a. Price and mileage are both quantitative.
b. This is a scatter chart.
c. You could identify a potentially good deal by looking for low mileage (to the left on the horizontal axis) and low price (lower on the vertical axis). Two potential good deals are shown in the following chart (circled and in red).



Feedback: Text.

15. **Tracking Stock Prices.** The following high-low-close stock chart gives the stock price for Exxon Mobile Corporation over a 12-month period. The data are the low, high, and closing price per share on the first trading day of the month. What can you say about the stock price and volatility of the stock price over this 12-month period? LO 3



Answer:

As the closing stock price dropped in months 8–10, the price became more volatile as evidenced by the longer bars (the range from low price to high price). The volatility of the price appears to be decreasing in months 11and 12.

Feedback: Text.

Instructor Manual

Camm, Cochran, Fry & Ohlmann, Data Visualization - Exploring and Explaining with Data, 1st Edition, © 2021 Cengage, 978-035-763-1348; Chapter 1: Introduction

Table of Contents

Purpose and Perspective of the Chapter

This introductory chapter provides the students with a perspective of how data visualization fits within analytics. First, we break down data visualization into its two major applications to explore and explain data. Then, we consider the preferred data visualizations for different types of data. Finally, we show practical applications of data visualizations to some of the major functional areas in business.

Cengage Supplements

List of Instructor Downloads

The following product-level supplements provide additional information that may help you prepare for your course. They are available in the Instructor Resource Center.

- A PowerPoint of the book chapter is available for download from the Instructor's Resource Center.
- A Solutions and Answers Guide for the book chapter is available for download from the Instructor's Resource Center.

List of Student Downloads

Students should download the following items from the Student Companion Center to complete the activities and assignments related to this chapter:

On the book's Companion Website, students can download the Data Files that accompany this text.

Chapter Objectives

This chapter addresses the following objectives:

LO 1.1 Define analytics and describe the different types of analytics

LO 1.2 Describe the different types of data and give an example of each

LO 1.3 Describe various examples of data visualization used in practice

LO 1.4 Identify the various charts defined in this chapter

Complete List of Chapter Activities and Assessments

For additional guidance, refer to the Teaching Online Guide.

Chapter Objective	PPT slides	Activity/Assessment	Duration
1.1	4-10	Activity in PPT	1 hour
1.2	11-14, 23	Activity in PPT	45 minutes
1.3	15-22, 24	Activity in PPT	1.5 hours
1.4	8-10, 12, 14-22	Activity in PPT	1.5 hours
1.1-1.4		Chapter 1 Problems	45 minutes
1.2	23	Discussion in PPT	15 minutes
1.3	24	Discussion in PPT	15 minutes
1.3, 1.4	25	Knowledge Checks in PPT	10 minutes

[return to top]

Key Terms

Analytics: the scientific process of transforming data into insights for making better decisions.

Bar chart: a chart that displays the magnitude of a quantitative variable by category or time period using horizontal bars.

Categorical data: data for which labels or names identify categories of like items.

Column chart: a chart that displays the magnitude of a quantitative variable by category or time period using vertical bars.

Cross-sectional data: data collected from several entities over the same time frame.

Data dashboard: a data visualization tool that includes multiple outputs and may update in real-time.

Data visualization: the graphical representation of data and information using charts, graphs, and maps.

Funnel chart: a chart that shows the progression of a numerical variable in a process from large to smaller values.

High-low-close stock chart: a chart that shows three numerical values: high value, low value, and closing value for the price of a share of stock over time.

Quantitative data: data for which numerical values indicate magnitude.

Scatter chart: a graphical presentation of the relationship between two quantitative variables.

Time series data: data collected over several points in time (minutes, hours, days, months, years, etc.).

[return to top]

Chapter Outline

I. Analytics (1-01, PPT Slides 4-6)
   a. Analytics is the scientific process of transforming data into insights for making better decisions.
   b. Three developments have spurred the explosive growth in the use of analytics for decision making:
      1. Technological advances – incredible amounts of data from scanner technology, e-commerce, social networks, sensors, and personal electronic devices such as cell phones.
      2. Methodological developments – faster algorithms can handle and explore massive amounts of data for data visualization, machine learning, optimization, and simulation.
      3. Explosion in computing power and storage capability – better computational hardware, parallel computing, and cloud computing, enable businesses to solve larger problems faster and with greater accuracy.
   c. A categorization of analytical methods:
      1. Descriptive analytics – encompasses a set of analytical tools that describes what has happened in the past.
         (1) Examples include data queries, standard reports, descriptive statistics, data visualization, cluster analysis, and what-if spreadsheet models.

2. Predictive analytics – consists of analytical tools that use models constructed from past data to predict the future or ascertain the impact of one variable on another.
    (1) Examples include linear regression, time series analysis, and predictive data mining.
3. Prescriptive Analytics – includes mathematical or logical models that indicate the best course of action to take in decision making.
    (1) Examples include optimization, simulation, and decision analysis.

II. Why visualize data? (1-02, PPT Slides 7-10)
   a. We create data visualizations for two reasons:
      1. Exploring data
      2. Communicating/explaining a message
   b. Data visualization for exploration: identify patterns.
      1. Data visualization is a powerful tool to identify patterns, such as creating a column chart that displays the seasonal pattern existing in the attendance by month at a zoo.
   c. Data visualization for exploration: understand relationships.
      1. Data visualization also helps to understand the relationship between two variables better, whether it is linear or not, beyond the information that regression statistics may provide.
   d. Data visualization is also essential for explaining relationships found in data and explaining the results of predictive and prescriptive models.

III. Types of data (01.03, PPT Slides 11-14)
   a. Quantitative vs. categorical data
      1. Quantitative data
         (1) Data for which numerical values indicate magnitude. Arithmetic operations, such as addition, subtraction, multiplication, and division, can be performed on quantitative data.
            (a) Examples: Share Price ($), and Volume.
      2. Categorical data
         (1) Data for which labels or names identify categories of like items. Arithmetic operations cannot be performed on categorical data.
            (a) Examples: Company, Symbol, and Industry
   b. Cross-sectional vs. time series data
      1. Cross-sectional data
         (1) Data collected from several entities over the same time frame.
      2. Time series data
         (1) Data collected over several time periods.

(a) Business and economic publications frequently include graphs of time series data.
(b) Graphs help analysts understand what happened in the past, identify trends over time, and project future levels for the time series.

c. Big data
    1. There is no universally accepted definition of big data.
    2. A working definition of big data:
        (1) any set of data too large or too complex to be handled by a desktop computer.
    3. The four Vs of big data:
        (1) Volume - the amount of data generated
        (2) Velocity - the speed at which the data are generated
        (3) Variety - the diversity in types and structures of data generated
        (4) Veracity - the reliability of the data generated

d. Word cloud
    1. A word cloud is a visual display that contains the key terms of a document. The size of the word is proportional to the frequency with which the word appears in the document.
    2. A word cloud is a frequently used chart to summarize words used in large sets of text data.

IV. Data visualization in practice (01.04, PPT Slides 15-22)

a. Accounting. A clustered column chart showing Benford's Law versus Tucker Software's Accounts Payable Entries
    1. Benford's Law (the First-Digit Law) states that the proportion of observations in which the first digit is 1 through 9, respectively, follows given probabilities.
    2. Benford's Law may help detect fraud. If the first digits of numbers in a data set do not conform to Benford's Law, fraud investigation may be warranted.

b. Finance. A high-low-close stock chart for Verizon Wireless
    1. A High-Low-Close Stock chart shows the high value, low value, and closing value of the price of a share of stock over time.
    2. This chart shows how the closing price changes over time and the price volatility each day.

c. Human Resource Management. A stacked column chart of employee turnover by month
    1. A stacked column chart shows part-to-whole comparisons, either over time or across categories, using different colors or shades of color.

2. This chart shows how January and July-October are the months in which most employees left the company, and April through June the months with most new hires.

d. Marketing. A funnel chart of website conversions for a software company
   1. A funnel chart shows the progression of a numerical variable for various categories from larger to smaller values.
   2. This chart helps to compare the conversion effectiveness of different website configurations, the use of bots, or changes in support services.

e. Operations. Time series data for unit sales of a product
   1. A line chart shows a variable of interest plotted over several time periods. A line chart helps to understand what happened in the past, identify trends over time, and predict future levels.
   2. This chart helps identify a repeating pattern and shows how units sold might also be increasing slightly over time.

f. Engineering. A quality control chart for dog food production
   1. A control chart is a graphical display of a variable of interest plotted over time relative to lower and upper control limits. It helps determine if a production process is in or out of control.
   2. Points beginning to appear outside the control limits are signals to inspect the process and make any necessary corrections.

g. Science. A spaghetti chart of hurricane tracks from multiple predictive models
   1. Geographic maps help display the results of complicated predictive models, such as predicting the path of a hurricane.
   2. A spaghetti chart owes its name to the fact that the depiction of multiple flows through a system using a line for each possible path resemble spaghetti

h. Sports. A shot chart for NBA player Chris Paul
   1. A shot chart displays the location of the shots attempted by a player during a basketball game with different symbols or colors indicating the outcome of a shot.
   2. This chart shows shot attempts by NBA player Chris Paul, with a blue dot indicating a successful shot and an orange x a missed shot.

V. Problems.
   a. #1 (LO 1.1, PPT Slide 6) Types of analytics.
   b. #5 (LOs 1.1 & 1.2, PPT Slides 11-12) House price and square footage.
   c. #12 (LOs 1.3 & 1.4, PPT Slide 18) Master's degree program recruiting.
   d. #14 (LOs 1.2, 1.3 & 1.4, PPT Slides 9, 11) Buying a used car.
   e. #15 (LO 1.3, PPT Slide 16) Tracking stock prices.

[return to top]

Discussion Activities

You can assign these activities several ways: in a discussion forum in your LMS, as whole-class discussions in person, or as a partner or group activity in class.

1. Exploration of *Anscombe*'s Data Sets (1-02 Why Visualize Data?, LO 1.2, PPT Slide 9) Duration 15 minutes.
    a. Consider the two scatter charts and related trendline and regression statistics shown for *Anscombe*'s data sets in Figure 1.2 (page 7.) The estimated regression equations and related R-squares for both data sets are identical. Does fitting a line to the data appear to be a wise choice for both data sets? Explain your answer.
        i. Answer: The fact the estimated regression equations and R-squares are identical for both data sets does not necessarily imply that the decision of fitting a line to both is wise. Linear regression should only be used when the x and y variable values form a linear shape on the scatter chart. Figure 1.2 tells us how a linear regression provides a good fit only for Data Set 1 because the dispersion of the residuals above and below the regression line is random and uniform. However, we can see how this not be true for Data Set 2, in which the regression line severely overestimates the data points on the left and the right of the chart, whereas it underestimates the data points in the middle of the chart.
    b. What would be a more appropriate regression equation to fit *Anscombe*'s Data Set 2?
        i. Answer: Even though a discussion on non-linear regression is beyond the scope of this textbook, a curvilinear relationship that better fits the pattern shown by Data Set 2 would be appropriate. When considering non-linear fitting, one should not over-fit the data set. Thus, it would be recommendable to begin the non-linear fit exploration by incremental degrees, such as by fitting a parabola ($\hat{y} = b_0 + b_1 x + b_2 x^2$) first, and only increase the polynomial degree if the pattern in the scatter chart justifies it. Other non-linear fittings such as logarithmic or exponential are also possible.
2. Tucker Software's Account Payable Entries Fraud Detection (1-04 Data Visualization in Practice, LO 1.3, PPT Slide 15) Duration 15 minutes.
    a. Consider the first digit of Tucker software's accounts payable entries in the clustered column chart in Figure 1.6 (page 12.) Does it appear that the data follow Benford's Law? Explain your answer.

i. Answer: The first digits for Tucker's accounts payable entries do not appear to follow Benford's Law for expected probabilities of first digits 1 through 9. It seems that there is an excessive number of first digits of 5 and 9. Because the sum of the proportions for all the first digits must add to 1 (the sample space), it follows that the other digits are underrepresented for what Benford's Law would dictate.

b. Which first digits from Tucker's accounts payable entries stand out as underrepresented in terms of absolute and relative proportional difference for the corresponding expected probabilities as dictated by Benford's Law?
   i. Answer: First digits 3 and 4, although smaller than first digit 1 in absolute value, appear to be underrepresented by the same relative proportion as first digit 1 is. However, we should further investigate the relative proportional difference between Tucker's accounts payable entries and its expected probability for first digit 7, as it appears far more pronounced than the others.

[return to top]

Appendix

Generic Rubrics

Providing students with rubrics helps them understand expectations and components of assignments. Rubrics help students become more aware of their learning process and progress, and they improve students' work through timely and detailed feedback. Customize these rubrics as you wish. The writing rubric indicates 40 points, and the discussion rubric indicates 30 points.

Standard Writing Rubric

Criteria	Meets Requirements	Needs Improvement	Incomplete
Content	The assignment clearly and comprehensively addresses all questions in the assignment. 15 points	The assignment partially addresses some or all questions in the assignment. 8 points	The assignment does not address the questions in the assignment. 0 points
Organization and Clarity	The assignment presents ideas in a clear manner and with a strong organizational structure. The assignment includes an appropriate introduction, content, and conclusion. The coverage of facts, arguments, and conclusions is logically related and consistent. 10 points	The assignment presents ideas in a mostly clear manner and with a mostly strong organizational structure. The assignment includes an appropriate introduction, content, and conclusion. The coverage of facts, arguments, and conclusions are most logically related and consistent. 7 points	The assignment does not present ideas in a clear manner and with a strong organizational structure. The assignment includes an introduction, content, and conclusion, but coverage of facts, arguments, and conclusions are not logically related and consistent. 0 points
Research	The assignment is based upon appropriate and adequate academic literature, including peer-reviewed journals and other scholarly work. 5 points	The assignment is based upon adequate academic literature but does not include peer-reviewed journals and other scholarly work. 3 points	The assignment is not based upon appropriate and adequate academic literature and does not include peer-reviewed journals and other scholarly work. 0 points
Research	The assignment follows the required citation guidelines. 5 points	The assignment follows some of the required citation guidelines. 3 points	The assignment does not follow the required citation guidelines. 0 points
Grammar and Spelling	The assignment has two or fewer grammatical and spelling errors. 5 points	The assignment has three to five grammatical and spelling errors. 3 points	The assignment is incomplete or unintelligible. 0 points

Standard Discussion Rubric

Criteria	Meets Requirements	Needs Improvement	Incomplete
Participation	Submits or participates in the discussion by the posted deadlines. Follows all the assignment's instructions for the initial post and responses. 5 points	Does not participate or submit discussion by the posted deadlines. Does not follow instructions for initial post and responses. 3 points	Does not participate in the discussion. 0 points
Contribution Quality	Comments stay on task. Comments add value to the discussion topic. Comments motivate other students to respond. 20 points	Comments may not stay on task. Comments may not add value to the discussion topic. Comments may not motivate other students to respond. 10 points	Does not participate in the discussion. 0 points
Etiquette	Maintains appropriate language. Constructively offers criticism. Provides both positive and negative feedback. 5 points	Does not always maintain the appropriate language. Offers criticism offensively. Provides only negative feedback. 3 points	Does not participate in the discussion. 0 points