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 2](#_Toc72181772)

[Cengage Supplements 2](#_Toc72181773)

[List of Instructor Downloads 2](#_Toc72181774)

[List of Student Downloads 2](#_Toc72181775)

[Chapter Objectives 3](#_Toc72181776)

[Complete List of Chapter Activities and Assessments 3](#_Toc72181777)

[Key Terms 3](#_Toc72181778)

[Chapter Outline 4](#_Toc72181779)

[Discussion Activities 7](#_Toc72181780)

[Appendix 10](#_Toc72181781)

[Generic Rubrics 10](#_Toc72181782)

[Standard Writing Rubric 10](#_Toc72181783)

[Standard Discussion Rubric 11](#_Toc72181784)

# 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]](#_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 tosmaller values.

**High-low-close stock chart:** a chart that shows three numerical values: high value, lowvalue, 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]](#_top)

# Chapter Outline

1. Analytics (1-01, PPT Slides 4-6)
   1. Analytics is the scientific process of transforming data into insights for making better decisions.
   2. 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.
   3. 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.
2. Why visualize data? (1-02, PPT Slides 7-10)
   1. We create data visualizations for two reasons:
      1. Exploring data
      2. Communicating/explaining a message
   2. 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.
   3. 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.
   4. Data visualization is also essential for explaining relationships found in data and explaining the results of predictive and prescriptive models.
3. Types of data (01.03, PPT Slides 11-14)
   1. 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.
            1. 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.
            1. Examples: Company, Symbol, and Industry
   2. 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.
            1. Business and economic publications frequently include graphs of time series data.
            2. Graphs help analysts understand what happened in the past, identify trends over time, and project future levels for the time series.
   3. 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
   4. 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.
4. Data visualization in practice (01.04, PPT Slides 15-22)
   1. 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.
   2. 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.
   3. 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.
   4. 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.
   5. 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.
   6. 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.
   7. 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
   8. 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.
5. Problems.
   1. #1 (LO 1.1, PPT Slide 6) Types of analytics.
   2. #5 (LOs 1.1 & 1.2, PPT Slides 11-12) House price and square footage.
   3. #12 (LOs 1.3 & 1.4, PPT Slide 18) Master’s degree program recruiting.
   4. #14 (LOs 1.2, 1.3 & 1.4, PPT Slides 9, 11) Buying a used car.
   5. #15 (LO 1.3, PPT Slide 16) Tracking stock prices.

[[return to top]](#_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.
   1. 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.
      1. 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.
   2. What would be a more appropriate regression equation to fit *Anscombe*'s Data Set 2?
      1. 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 () 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.
   1. 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.
      1. 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.
   2. 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?
      1. 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]](#_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 |