

Chapter 2

Presentation of Quantitative Data: Data Visualization



Contents

2.1	Introduction	21
2.2	Data Classification	22
2.3	Data Context and Data Orientation	23
2.3.1	Data Preparation Advice	26
2.4	Types of Charts and Graphs	29
2.4.1	Ribbons and the Excel Menu System	29
2.4.2	Some Frequently Used Charts	31
2.4.3	Specific Steps for Creating a Chart	35
2.5	An Example of Graphical Data Analysis and Presentation	38
2.5.1	Example—Tere’s Budget for the 2nd Semester of College	39
2.5.2	Collecting Data	40
2.5.3	Summarizing Data	40
2.5.4	Analyzing Data	43
2.5.5	Presenting Data	48
2.6	Some Final Practical Graphical Presentation Advice	51
2.7	Summary	55
	Key Terms	56
	Problems and Exercises	56

2.1 Introduction

We often think of data as being numerical values, and in business, those values are often stated in terms of units of currency (dollars, pesos, dinars, etc.). Although data in the form of currency are ubiquitous, it is quite easy to imagine other numerical units: percentages, counts in categories, units of sales, etc. This chapter, in conjunction with Chap. 3, discusses how we can best use Excel’s graphic capabilities to effectively present quantitative data (**ratio** and **interval**) to inform and influence an audience, whether it is in euros or some other quantitative measure. In Chaps. 4 and 5, we will acknowledge that not all data are numerical by focusing on qualitative (**categorical/nominal** or **ordinal**) data. The process of data gathering often produces a combination of data types, and throughout our discussions, it will be impossible to

ignore this fact: quantitative and qualitative data often occur together. Let us begin our study of **data visualization**.

Unfortunately, the scope of this book does not permit in depth coverage of the data *collection* process, so I strongly suggest you consult a reference on data research methods before you begin a significant data collection project. I will make some brief remarks about the planning and collection of data, but we will generally assume that data has been collected in an efficient manner, and that the data accurately represents what is intended to be measured. Now, let us consider the essential ingredients of good data visualization and the issues that can make it either easy or difficult to succeed. We will begin with a general discussion of data: how to classify it, and the context within which it exists.

2.2 Data Classification

Skilled data analysts spend a great deal of time and effort in planning a data collection effort. They begin by considering the type of data they *can* and *will* collect to meet their goals for the use of the data. Just as carpenters are careful in selecting their tools, so are analysts in their choice of data. You cannot expect a *low precision* tool to perform *high precision* work. The same is true for data. A good analyst is cognizant of the types of analyses they can perform on various categories of data. This is particularly true in statistical analysis, where there are often strict rules for the types of analyses that can be performed on various types of data.

The standard characteristics that help us categorize data are presented in Table 2.1. Each successive category, nominal to ratio, permits greater *measurement* precision and more extensive statistical analysis. Thus, we can see from Table 2.1 that ratio data measurement is more precise than nominal data measurement. It is important to remember that all these forms of data, regardless of their classification, are valuable, and we collect data in different forms by considering availability and our analysis goals. For example, nominal data are used in many marketing studies, while ratio data are more often the tools of finance, operations, and economics; yet, all business functions collect data in each of these categories.

For nominal and ordinal data, we use non-metric measurement scales in the form of categorical properties or attributes. Interval and ratio data are based on metric measurement scales allowing a wide variety of mathematical operations to be performed on the data. The major difference between interval and ratio measurement scales is the existence of an absolute zero for ratio scales, and arbitrary zero points for interval scales. For example, consider a comparison of the Fahrenheit and Celsius temperature scales. The zero points for these scales are arbitrarily set and do not indicate an “absolute absence” of temperature. Similarly, it is incorrect to suggest that 40° C is half as hot as 80° C. By contrast, it can be said that 16 ounces of coffee *are*, in fact, twice as heavy as 8 ounces. Ultimately, the ratio scale has the highest information content of any of the measurement scales.

Table 2.1 Data categorization

Data	Description	Properties	Examples
Nominal or categorical data	Data that can be placed into mutually exclusive categories	Quantitative relationships among and between data are meaningless and descriptive statistics are meaningless	Country in which you were born, a geographic region, your gender—These are either/or categories
Ordinal data	Data are ordered or ranked according to some characteristic	Categories can be compared to one another, but the difference in categories is generally meaningless and calculating averages is suspect	Ranking breakfast cereals—Preferring cereal X <i>more</i> than Y implies nothing about <i>how much more</i> you like one versus the other
Interval data	Data characterized and ordered by a specific distance between each observation, but having no natural zero	Ratios are meaningless, thus 15° C is not half as warm as 30° C	The Fahrenheit (or Celsius) temperature scale or consumer survey scales that are <i>specified</i> to be interval scales
Ratio data	Data that have a natural zero	These data have both ratios and differences that are meaningful	Sales revenue, time to perform a task, length, or weight

Just as thorough problem definition is essential to problem-solving, careful selection of appropriate data categories is essential in a data collection effort. Data collection is an arduous and often costly task, so why not carefully plan for the use of the data prior to its collection? Additionally, remember that there are few things that will anger a cost conscious superior more than the news that you have to repeat a data collection effort.

2.3 Data Context and Data Orientation

The data that we collect and assemble for presentation purposes exists in a unique **data context**: a set of conditions (an environment) related to the data. This context is important to our understanding of the data. We relate data to time (e.g. daily, quarterly, yearly, etc.), to categorical treatments (e.g. an economic downturn, sales in Europe, etc.), and to events (e.g. sales promotions, demographic changes, etc.). Just as we record the values of quantitative data, we also record the context of data, such as revenue generated by product A, in quarter B, due to salesperson C, in sales territory D. Thus, associated with the quantitative data element that we record are numerous other important data elements that may, or may not, be quantitative.

Sometimes the context is obvious, sometimes the context is complex and difficult to identify, and oftentimes, there is more than a single context that is essential to consider. Without an understanding of the data context, important insights related to the data can be lost. To make matters worse, the context related to the data may

change or reveal itself only after substantial time has passed. For example, consider data that indicates a substantial loss of value in your stock portfolio, recorded from 1990 to 2008. If the only context that is considered is time, it is possible to ignore a host of important contextual issues—e.g. the bursting of the dot-com *bubble* of the late 1990s. Without knowledge of this contextual event, you may simply conclude that you are a poor stock picker.

It is impossible to anticipate all the elements of data context that should be collected, but whatever data we collect should be sufficient to provide a context that suits our needs and goals. If I am interested in promoting the idea that the revenues of my business are growing over time and growing only in selected product categories, I will assemble time-oriented revenue data for the various products of interest. Thus, the related dimensions of my revenue data are time and product. There may also be an economic context, such as demographic conditions that may influence types of sales. Determining the contextual dimensions that are important will influence what data we collect and how we present it. Additionally, you can save a great deal of effort and *after-the-fact* data adjustment by carefully considering, in advance, the various dimensions that you will need.

Consider the owner of a small business that is interested in recording expenses in a variety of accounts for cash flow management, income statement preparation, and tax purposes. This is an important activity for any small business. Cash flow is the lifeblood of these businesses, and if it is not managed well, the results can be catastrophic. Each time the business owner incurs an expense, he either collects a receipt (upon final payment) or an invoice (a request for payment). Additionally, suppliers to small businesses often request a deposit that represents a form of partial payment and a commitment to the services provided by the supplier.

An example of these data is shown in the worksheet in Table 2.2. Each of the primary data entries, referred to as **records**, contain several important and diverse dimensions, referred to as fields—date, account, amount, nature of the expense, names, a comment, etc. A record represents a single observation of the collected data fields, as in item 3 (printing on 1/5/2004) of Table 2.2. This record contains 7 fields—Item 1, Printing, \$2543.21, etc.—and each record is a row in the worksheet.

Somewhere in our business owner's office is an old shoebox that is the final resting place for his primary data. It is filled with scraps of paper: invoices and receipts. At the end of each week, our businessperson empties the box and records what he believes to be the important elements of each receipt or invoice. Table 2.2 is an example of the type of data that the owner might collect from the receipts and invoices over time. The receipts and invoices can contain more data than needs to be recorded or used for analysis and decision making. The dilemma that the owner faces is the amount and type of data to record in the worksheet; recording too much data can lead to wasted effort and neglect of other important activities, and recording too little data can lead to overlooking important business issues.

What advice can we provide our businessperson that might make his efforts in collecting, assembling, and recording data more useful and efficient? Below I provide several guidelines that can make the process of planning for a data collection effort more straightforward.

Table 2.2 Payment example

Item	Account	\$ Amount	Date Rcvd.	Deposit	Days to pay	Comment
1	Office supply	\$123.45	1/2/2004	\$10.00	0	Project X
2	Office supply	\$54.40	1/5/2004	\$0.00	0	Project Y
3	Printing	\$2543.21	1/5/2004	\$350.00	45	Feb. brochure
4	Cleaning	\$78.83	1/8/2004	\$0.00	15	Monthly
Service						
5	Coffee	\$56.92	1/9/2004	\$0.00	15	Monthly
Service						
6	Office supply	\$914.22	1/12/2004	\$100.00	30	Project X
7	Printing	\$755.00	1/13/2004	\$50.00	30	Hand bills
8	Office supply	\$478.88	1/16/2004	\$50.00	30	Computer
9	Office rent	\$1632.00	1/19/2004	\$0.00	15	Monthly
10	Fire insurance	\$1254.73	1/22/2004	\$0.00	60	Quarterly
11	Cleaning	\$135.64	1/22/2004	\$0.00	15	Water damage
Service						
12	Orphan's	\$300.00	1/27/2004	\$0.00	0	Charity*
Fund						
13	Office supply	\$343.78	1/30/2004	\$100.00	15	Laser printer
14	Printing	\$2211.82	2/4/2004	\$350.00	45	Mar. brochure
15	Coffee	\$56.92	2/5/2004	\$0.00	15	Monthly
Service						
16	Cleaning	\$78.83	2/10/2004	\$0.00	15	Monthly
Service						
17	Printing	\$254.17	2/12/2004	\$50.00	15	Hand bills
18	Office supply	\$412.19	2/12/2004	\$50.00	30	Project Y
19	Office supply	\$1467.44	2/13/2004	\$150.00	30	Project W
20	Office supply	\$221.52	2/16/2004	\$50.00	15	Project X
21	Office rent	\$1632.00	2/18/2004	\$0.00	15	Monthly
22	Police fund	\$250.00	2/19/2004	\$0.00	15	Charity
23	Printing	\$87.34	2/23/2004	\$25.00	0	Posters
24	Printing	\$94.12	2/23/2004	\$25.00	0	Posters
25	Entertaining	\$298.32	2/26/2004	\$0.00	0	Project Y
26	Orphan's	\$300.00	2/27/2004	\$0.00	0	Charity
Fund						
27	Office supply	\$1669.76	3/1/2004	\$150.00	45	Project Z
28	Office supply	\$1111.02	3/2/2004	\$150.00	30	Project W
29	Office supply	\$76.21	3/4/2004	\$25.00	0	Project W
30	Coffee	\$56.92	3/5/2004	\$0.00	15	Monthly
Service						
31	Office supply	\$914.22	3/8/2004	\$100.00	30	Project X
32	Cleaning	\$78.83	3/9/2004	\$0.00	15	Monthly

(continued)

Table 2.2 (continued)

Item	Account	\$ Amount	Date Rcvd.	Deposit	Days to pay	Comment
Service						
33	Printing	\$455.10	3/12/2002	\$100.00	15	Hand bills
34	Office supply	\$1572.31	3/15/2002	\$150.00	45	Project Y
35	Office rent	\$1632.00	3/17/2002	\$0.00	15	Monthly
36	Police fund	\$250.00	3/23/2002	\$0.00	15	Charity
37	Office supply	\$642.11	3/26/2002	\$100.00	30	Project W
38	Office supply	\$712.16	3/29/2002	\$100.00	30	Project Z
39	Orphan's	\$300.00	3/29/2002	\$0.00	0	Charity
Fund						

2.3.1 Data Preparation Advice

1. *Not all data are created equal*—Spend some time and effort considering the category of data (nominal, ratio, etc.) that you will collect and how you will use it. Do you have choices in the categorical type of data you can collect? How will you use the data in analysis and presentation?
2. *More is better*—If you are uncertain of the specific dimensions of a data observation that you will need for analysis, err on the side of recording a greater number of dimensions (more information on the context). It is easier not to use collected data than it is to add the un-collected data later. Adding data later can be costly and assumes that you will be able to locate it, which may be difficult or impossible.
3. *More is **not** better*—If you can communicate what you need to communicate with less data, then by all means do so. Bloated databases can lead to distractions and misunderstanding. With new computer memory technology, the cost of data storage is declining rapidly, but there is still a cost to data entry, storage, verifying data accuracy, and achieving records for long periods of time.
4. *Keep it simple and columnar*—Select a simple, unique title for each data dimension or field (e.g. Revenue, Address, etc.) and record the data in a column, with each row representing a record, or observation, of recorded data. Each column (field) represents a different dimension of the data. Table 2.2 is a good example of columnar data entry for seven data fields.
5. *Comments are useful*—It may be wise to place a *miscellaneous* dimension or field reserved for a variety of written observations—a **comment field**. Be careful! Because of their unique nature, comments are often difficult, if not impossible, to query via structured database query languages (**SQL**). Try to pick key words for entry (*overdue*, *lost sale*, etc.) if you plan to later query the field.
6. *Consistency in category titles*—Although you may not consider that there is a significant difference between the category titles *Deposit* and *\$Deposit*, Excel will view them as completely distinct field titles. Excel is not capable of understanding that the terms may be synonymous in your mind.

Let’s examine Table 2.2 in light of the data preparation advice we have just received, but first, let’s take a look at a typical invoice and the data that it might contain. Figure 2.1 shows an invoice for office supply items purchased at Hamm Office Supply, Inc. Note the amount of data that this generic invoice (an MS Office Template) contains is quite substantial: approximately 20 fields. Of course, some of the data are only of marginal value, such as our address—we know that the invoice was intended for our firm, and we know where we are located. Yet, it is verification that the Hamm invoice is in fact intended for our firm. Notice that each line item in the invoice will require *multiple* item entries—qty (quantity), description, unit price, and total. Given the potential for large quantities of data, it would be wise to consider a **relational database**, such as MS Access, to optimize data entry effort. Of course, even if the data are stored in a relational database, that does not

Hamm Office Supply

Invoice No.
AB-1234

INVOICE

Customer

Name _____
 Address _____
 City _____ State _____ ZIP _____
 Phone _____

Misc.

Date _____
 Order No. _____
 Rep _____
 FOB _____

Qty	Description	Unit Price	TOTAL

Payment

Comments _____
 Name _____
 CC # _____
 Expires _____

Tax Rate(s)

Sub-Total	
Shipping	
TOTAL	

Office Use Only

Fig. 2.1 Generic invoice

restrict us from using Excel to analyze the data by downloading it from Access to Excel; in fact, this is a wonderful advantage of the Office suite.

Now, let's examine the data in Table 2.2, in light of our advice:

1. *Not all data are created equal*—Our businessperson has assembled a variety of data dimensions or fields to provide the central data element (\$ Amount) with ample context and orientation. The seven fields that comprise each record appear to be sufficient for the businessperson's goal of recording the expenses and describing the context associated with his business operation. This includes recording each expense to ultimately calculate annual profit or loss, tracking expenses associated with projects or other uses of funds (e.g. charity), and the timing of expenses (Date Rcvd., Days to Pay, etc.) and subsequent cash flow. If the businessperson expands his examination of the transactions, some data may be missing (for example, Order Number or Shipping Cost). Only the future will reveal if these data elements will become important, but for now, these data are not collected in the spreadsheet.
2. *More is better*—The data elements that our businessperson has selected may not all be used in our graphical presentation, but this could change in the future. It is better to collect more data initially than to perform an extensive collection of data at a later date. The invoices and scraps of paper representing primary data may be difficult to find or identify in 3 months.
3. *More is not better*—Our businessperson has carefully selected the data that he feels is necessary, without creating excessive data entry effort.
4. *Keep it simple and columnar*—Unique and simple titles for the various data dimensions (e.g. Account, Date Rcvd., etc.) have been selected and arranged in columnar fashion. Adding, inserting, or deleting a column is virtually costless for an Excel user, skilled or unskilled.
5. *Comments are useful*—The Comment field has been designated for the specific project (e.g. Project X), source item (e.g. Computer), or other important information (e.g. Monthly charge). If any criticism can be made here, it is that maybe these data elements deserve a title other than Comment. For example, entitle this data element *Project/Sources of Expense* and use the Comment title as a less structured data category. These could range from comments relating to customer service experiences to information on possible competitors that provide similar services.
6. *Consistency in category titles*—Although you may not consider there to be a significant difference between the account titles Office Supply and Office Supplies, Excel will view them as completely distinct accounts. Our businessperson appears to have been consistent in the use of account types and comment entries. It is not unusual for these entries to be converted to numerical codes, thus replacing Printing with an account code, such as 351.

2.4 Types of Charts and Graphs

There are literally hundreds of types of charts and graphs (these are synonymous terms) available in Excel. Thus, the possibilities for selecting a presentation format are both interesting and daunting. Which graph type is best for my needs? Often the answer is that more than one type of graph will perform the presentation goal required; thus, the selection is a matter of your taste or that of your audience. Therefore, it is convenient to divide the problem of selecting a presentation format into two parts: 1) the actual data presentation, and 2) the embellishment that will surround it. In certain situations, we choose to do as little embellishment as possible; in others, we find it necessary to dress the data presentation in lovely colors, backgrounds, and labeling. To determine how to blend these two parts, ask yourself few simple questions:

1. What is the purpose of the data presentation? Is it possible to show the data without embellishment, or do you want to attract attention through your presentation *style*? In a business world where people are exposed to many, many presentations, it may be necessary to do something extraordinary to gain attention or simply conform to the norm.
2. At what point does my embellishment of the data become distracting? Does the embellishment cover or conceal the data? Don't forget that from an information perspective, it is *all* about the data, so don't detract from its presentation by adding superfluous and distracting adornment.
3. Am I being true to my taste and style of presentation? This author's taste in formatting is guided by some simple principles that can be stated in several familiar laws: *less is more*, *small is beautiful*, and *keep it simple*. As long as you are able to deliver the desired information and achieve your presentation goal, there is no problem with our differences in taste.
4. Formatting should be consistent among graphs in a workbook—don't mix various formats, unless there is good reason to do so.

2.4.1 Ribbons and the Excel Menu System

So, how do we put together a graph or chart? In pre-2007 Excel, an ingenious tool called a **Chart Wizard** was available to perform these tasks. As the name implies, the Chart Wizard guided you through standardized steps (four, to be exact) that took the guesswork out of creating charts. If you followed the four steps, it was almost fool-proof, and if you read all the options available to you for each of the four steps, it would allow you to create charts very quickly. In Excel 2007, the wizard was replaced because of a major development in the Excel 2007 user interface—ribbons. Ribbons replaced the old hierarchical pull-down menu system that was the basis for user interaction with Excel. Ribbons are menus and commands organized in tabs that provide access to the functionality for specific uses. Some of these will appear

familiar to pre-Excel 2007 users and others will not—Home, Insert, Page Layout, Formulas, Data, Review, Format, and View. Within each tab, you will find **groups** of related functionality and commands. Additionally, some menus specific to an activity, for example, the creation of a graph or chart, will appear as the activity is taking place. For those just beginning to use Excel 2016 and with no previous exposure to Excel, you will probably find the menu system quite easy to use; for those with prior experience with Excel, the transition may be a bit frustrating at times. I have found the ribbon system quite useful, in spite of the occasional difficulty of finding functionality that I was accustomed to before Excel 2007. Figure 2.2 shows the Insert tab where the Charts group is found.

In this Figure, a very simple graph of six data points for two data series, *data 1* and *data 2*, is shown as two variations of the column graph. One also displays the data used to create the graph. Additionally, since the leftmost graph has been selected, indicated by the border that surrounds the graph, a group of menus appear at the top of the ribbon—**Chart Tools**. These tools contain menus for Design, Layout, and Format. This group is relevant to the creation of a chart or graph. Ultimately, ribbons lead to a flatter, or less hierarchical, menu system.

Our first step in chart creation is to organize our data in a worksheet. In Fig. 2.2 the six data points for the two series have a familiar columnar orientation and have titles, *data 1* and *data 2*. By capturing the **data range** containing the data that you intend to chart before engaging the charts group in the Insert tab, you automatically identify the data to be graphed. Note that this can, but need not, include the column title of the data specified as text. By capturing the title, the graph will assume that you want to name the data series the same as title selected. If you place alphabetic

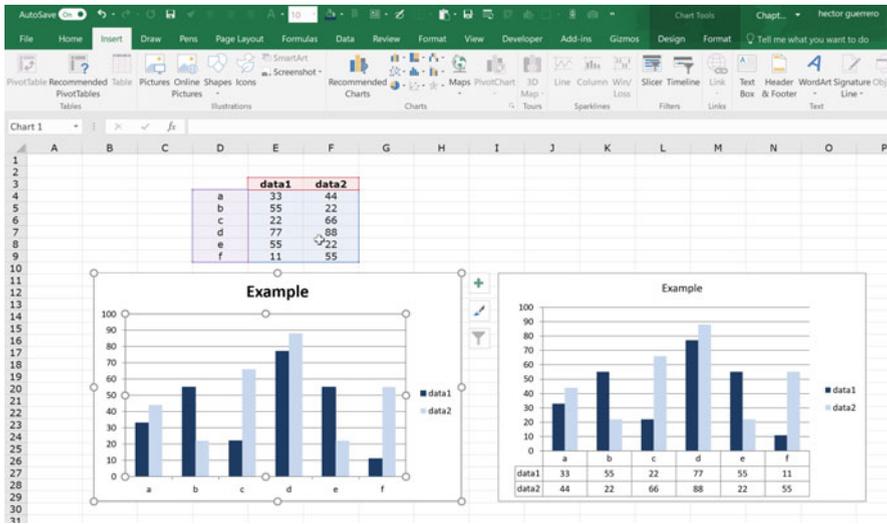


Fig. 2.2 Insert tab and excel chart group

characters, a through f in this case, in the first column of the captured data, the graph will use these characters as the x-axis of the chart.

If you prefer not to capture the data prior to engaging the charts group, you can either: (1) open and capture a blank chart type and *copy* and *paste* the data to the blank chart type, or (2) use a right click of your mouse to *select data*. Obviously, there will be numerous detailed steps to capturing data and labeling the graph appropriately. We defer to a detailed example of creating graphs using the chart group for the next section.

2.4.2 Some Frequently Used Charts

It is always dangerous to make bold assertions, but it is generally understood that the *mother of all graphs* is the **Column** or **Bar chart**. They differ only in their vertical and horizontal orientation, respectively. They easily represent the most often occurring data situation: some observed numerical variable that is measured in a single dimension (often time). Consider a simple set of data related to five products (A–E) and their sales over a 2-year period of time and measured in millions of dollars. The first four quarters represent year one, and the second four quarters represent year two. These data are shown in Table 2.3. In Quarter 1 of the second year, sales revenue for product B was \$49,000,000.

A quick visual examination of the data in Table 2.3 reveals that the product sales are relatively similar in magnitude (less than 100), but with differences in quarterly increases and decreases within the individual products. For example, product A varies substantially over the eight quarters, while product D shows relatively little variation. Additionally, it appears that when product A shows high sales in early quarters (1 and 2), product E shows low sales in early quarters—they appear to be somewhat negatively correlated, although a graph may reveal more conclusive information. **Negative correlation** implies that one data series moves in the opposite direction from another; **positive correlation** suggests that both series move in the same direction. In later chapters, we will discuss statistical correlation in greater detail.

Table 2.3 Sales data for products A–E (in millions of dollars)

Quarter	A	B	C	D	E
1	98	45	64	21	23
2	58	21	45	23	14
3	23	36	21	31	56
4	43	21	14	30	78
1	89	49	27	35	27
2	52	20	40	40	20
3	24	43	58	37	67
4	34	21	76	40	89

Let’s experiment with a few chart types to examine the data and tease out insights related to product A–E sales. The first graph, Fig. 2.3, displays a simple column chart of sales for the five-product series in each of eight quarters. The relative magnitude of the five products in a quarter is easily observed, but note that the five-product series are difficult to follow through time, despite the color coding. It is difficult to concentrate solely on a single series (e.g. Product A) through time; it is even more difficult to follow multiple series.

In Fig. 2.4, the chart type used is a **Stacked Column**. This graph provides a view not only of the individual product sales, but also of the quarterly totals for products A–E. By observing the absolute height of each stacked column, one can see that total

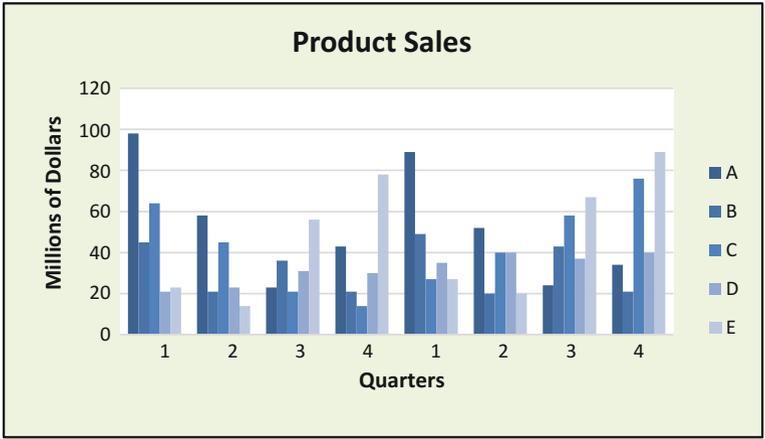


Fig. 2.3 Column chart for products A–E

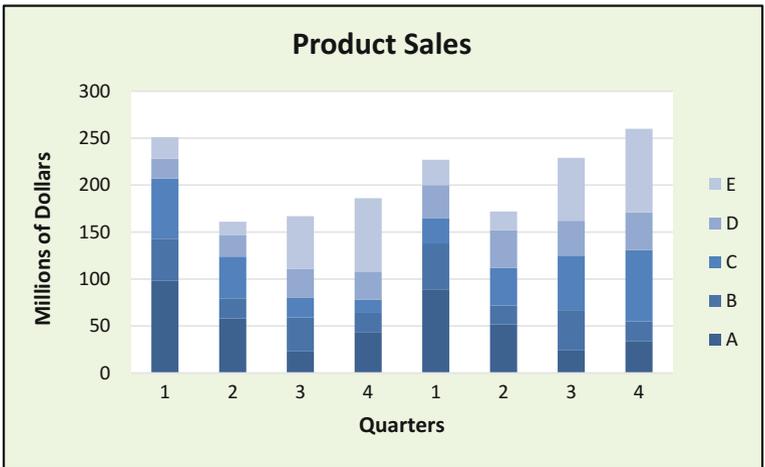


Fig. 2.4 Stacked column chart for products A–E

product sales in Quarter 1 of Year 1 (first horizontal value 1) are greater than Quarter 1 of Year 2 (second horizontal value 1). The relative size of each color within a column provides information of the sales quantities for each product in the quarter. For our data, the Stacked Column chart provides visual information about quarterly totals that is easier to discern. Yet, it remains difficult to track the quarterly changes within products and among products over time. For example, it would be difficult to determine if product D is greater or smaller in Quarter 3 or 4 of Year 1, or to determine the magnitude of each.

Next, Fig. 2.5 demonstrates a **3-D Column** (3 dimensional) chart. This is a visually impressive graph due to the 3-D effect, but much of the information relating to time-based behavior of the products is lost due to the inability to clearly view columns that are overlapped by other columns. The angle of perspective for 3-D graphs can be changed to partially remedy this problem, but if a single graph is used to chart many data series, they can still be difficult or impossible to view.

Now, let us convert the chart type to a **Line Chart** and determine if there is an improvement or difference in the visual interpretation of the data. Before we begin, we must be careful to consider what we mean by an improvement, because an improvement is only an improvement relative to a goal that we establish for data presentation. For example, consider the goal that the presentation portrays the changes in each product’s sales over quarters. We will want to use a chart that easily permits the viewer’s eye to follow the quarterly related change in each specific series. Line charts will probably provide a better visual presentation of the data than column charts, especially in **time series data**, if this is our goal.

Figure 2.6 shows the five-product data in a simple and direct format. Note that the graph provides information in the three areas we have identified as important:

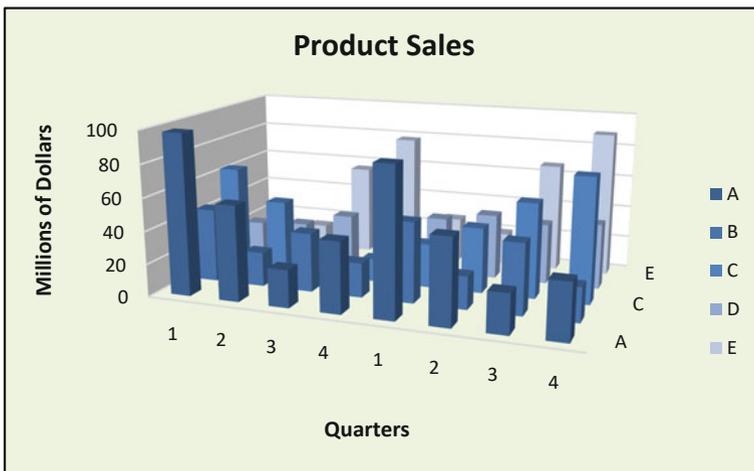


Fig. 2.5 3-D column chart for products A–E

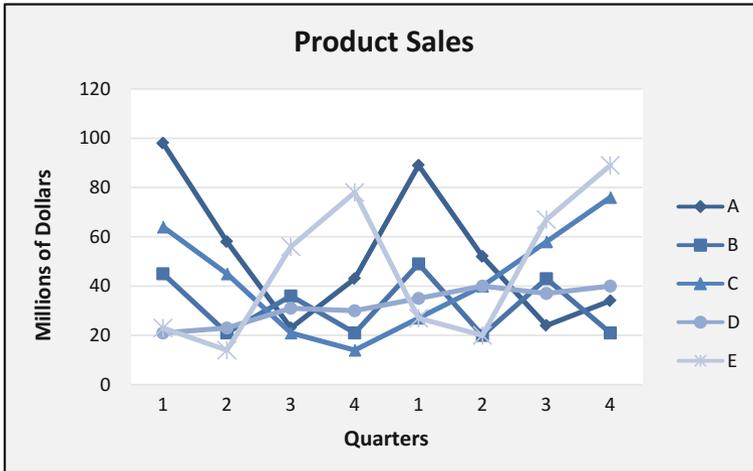


Fig. 2.6 Line chart for products A–E

(1) the relative value of a product’s sales to other products *within* each quarter, (2) the relative value of a product’s sales to other products *across* quarters, and (3) the behavior of the individual product’s sales over quarters. The line graph provides some interesting insights related to the data. For example:

1. Products A and E both appear to figure seasonal behavior that achieves highs and lows approximately every four quarters (e.g. highs in Quarter 1 for A and Quarter 4 for E).
2. The high for product E is offset approximately one quarter from that of the high for A. Thus, the peak in sales for Product A **lags** (occurs later) the peak for E by one quarter.
3. Product D seems to show little seasonality, but does appear to have a slight **linear trend** (increases at a *constant* rate over time). The trend is positive; that is, sales increase over time.
4. Product B has a stable pattern of quarterly, alternating increases and decreases, and it may have very slight positive trend from year 1 to year 2.

Needless to say, line graphs can be quite revealing, even if the behavior is based on scant data. Yet, we must also be careful not to convince ourselves of **systematic behavior** (regular or predictable) based on little data; more data will be needed to convince ourselves of true systematic behavior.

Finally, Fig. 2.7 is also a line graph, but in 3-D. It suffers from the same visual obstructions that we experienced in the 3-D column graph—possibly appealing from a visual perspective, but providing less information content than the simple line graph in Fig. 2.6, due to the obstructed view. It is difficult to see values of product E (the rear-most line) in early quarters. As I stated earlier, *simple* graphs are often *better* from a presentation perspective.

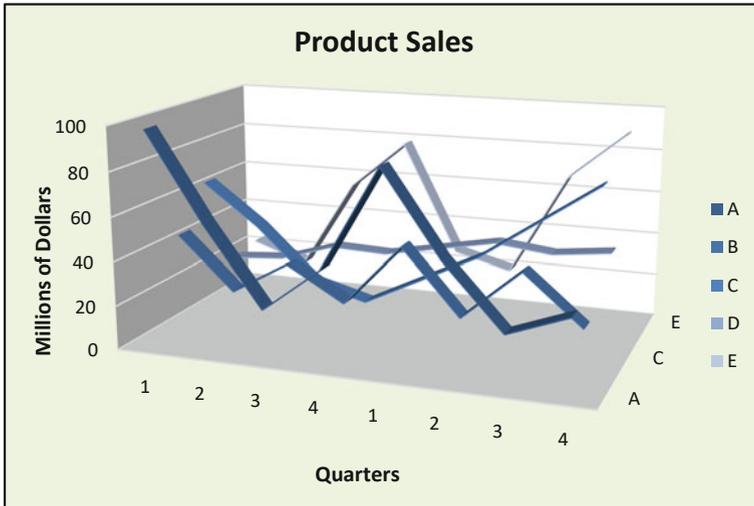


Fig. 2.7 3-D line chart for products A–E

2.4.3 Specific Steps for Creating a Chart

We have seen the results of creating a chart in Figs. 2.3, 2.4, 2.5, 2.6, and 2.7. Now, let us create a chart from beginning to end for Fig. 2.6—the Line chart for all products. The process we will use includes the following steps: (1) select a chart type, (2) identify the data to be charted including the x-axis, and (3) provide titles for the axes, series, and chart. For step 1, Fig. 2.8 shows the selection of the Line chart format (see arrow) within the Charts group of the Insert tab. Note that there are also custom charts available for specialized circumstances. In pre-2007 Excel, these were a separate category of charts, but after 2007, they were incorporated into the Format options.

The next step—selection of the data—has a number of possible options. The option shown in Fig. 2.9 has a blank chart selected, and the chart is engaged (arrows indicate engaged blank chart). A right click of the mouse produces a set of options, including *Select Data*, where the data range can be copied into the blank chart. By capturing the data range, including the titles (B1:F9), a series title (A, B, etc.) is automatically provided. Alternatively, if the data range had been selected prior to selecting the chart type, the data also would have been automatically captured in the line chart.

Note that the X-axis (horizontal) for Fig. 2.9 is represented by the quarters of each of the 2 years, 1–4 for each year. In order to reflect this in the chart, you must specify the range where the axis labels are located. You can see that our axis labels are in range A2:A9. This could also be done in the prior step of *Select Data*.

In Fig. 2.10 we can see the partially completed chart. A right click on the chart area permits you to once again use the *Select Data* function. The dialogue box that

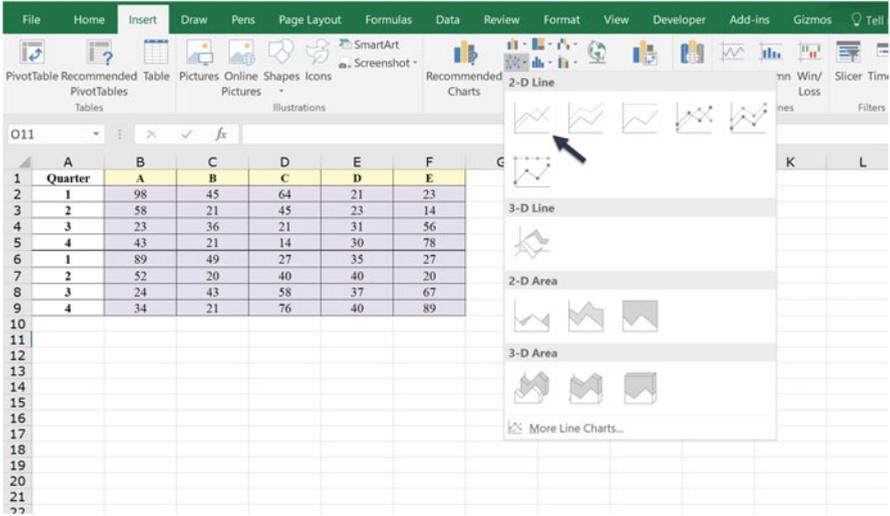


Fig. 2.8 Step 1-selection of line chart from charts group

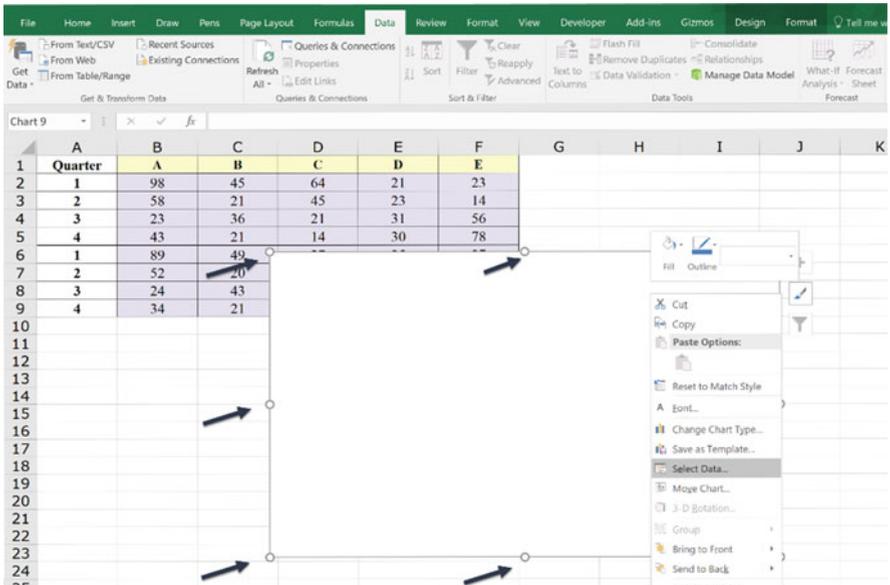


Fig. 2.9 Step 1-selection of data range for engaged chart

appears permits you to select the new **Horizontal (Category) Axis Labels**. By pressing the Edit button in the Horizontal Axis Labels window, you can capture the appropriate range (A2:A9) to change the x-axis. This is shown in Fig. 2.11.

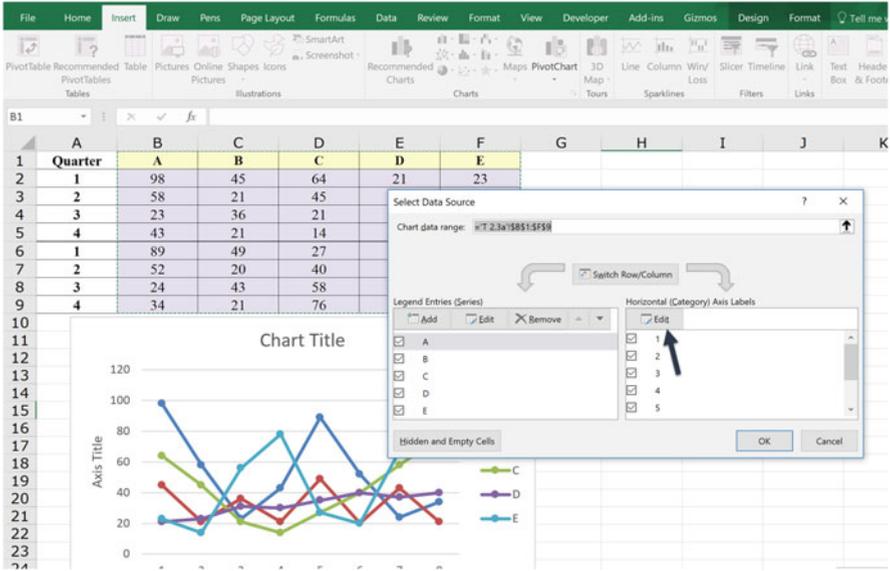


Fig. 2.10 Step 2-select data source dialogue box

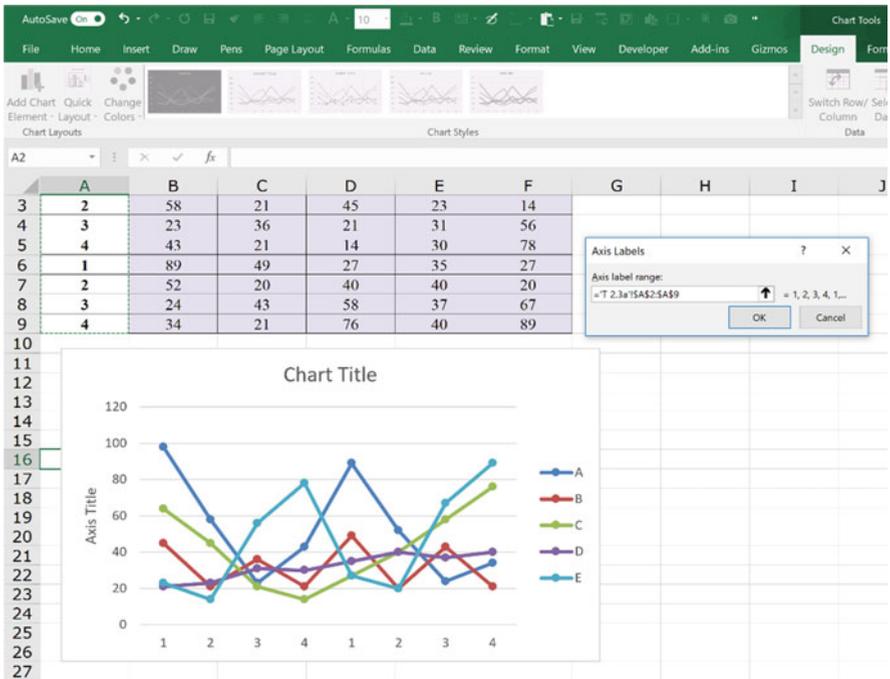


Fig. 2.11 Step 3-selection of X-axis data labels

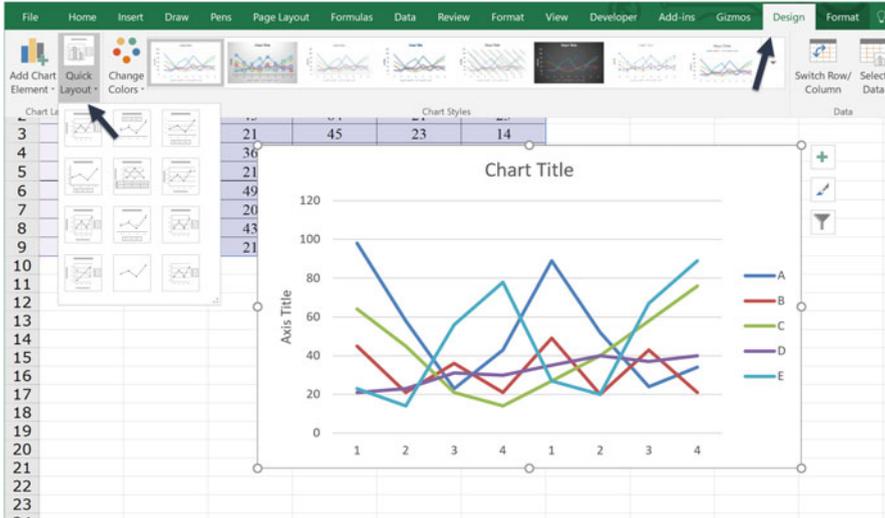


Fig. 2.12 Step 3-chart design, layout, and format

Step 3 of the process permits titles for the chart and axes. Figure 2.12 shows the selection of Quick Layout in the Design tab group. The Design and Format tabs provide many more options for customizing the look of the chart to your needs.

As we mentioned earlier, many details for a chart can be handled by pointing and right-clicking; for example, **Select Data**, **Format Chart Area**, and **Chart Type** changes. Selecting a part of the graph or chart with a left click, for example an axis or a Chart Title, then right clicking, also permits changes to the look of the chart or changes in the axes scale, pattern, font, etc. A more direct way to make changes is to use the **Add Chart Element** subgroup in the Design tab. I would suggest that you take a simple data set, similar to the one I have provided and experiment with all the options available. Also, try the various chart types to see how the data can be displayed.

2.5 An Example of Graphical Data Analysis and Presentation

Before we begin a full-scale example of graphical data analysis and presentation/ visualization, let's consider the task we have before us. We are engaging in an exercise, **Data Analysis**, which can be organized into four basic activities: *collecting*, *summarizing*, *analyzing*, and *presenting* data. Our example will be organized into each of these steps, all of which are essential to successful graphical data analysis.

Collecting data not only involves the act of gathering, but also includes careful planning for the types of data to be collected (interval, ordinal, etc.). Data collection

can be quite costly; thus, if we gather the wrong data or omit necessary data, we may have to make a costly future investment to repeat this activity. Some important questions we should ask before collecting are:

1. What data are necessary to achieve our analysis goals?
2. Where and when will we collect the data?
3. How many and what types of data elements related to an observation (e.g. customer name, date, etc.) are needed to describe the context or orientation? For example, each record of the 39 total records in Table 2.2 represents an invoice or receipt observation with seven data fields of nominal, interval, and ratio data elements.

Summarizing data can be as simple as placing primary data elements in a worksheet, but also it can include numerous modifications that make the data more useable. For example, if we collect data related to a date (1/23/2013), should the date also be represented as a day of the week (Wednesday)? This may sound redundant, since a date implies a day of the week, but the data collector must often make these conversions of the data. Summarizing prepares the data for the analysis that is to follow. It is also possible that during analysis, the data will need further summarization or modification to suit our goals.

There are many techniques for **analyzing** data. Not surprisingly, valuable analysis can be performed by simply **eyeballing** (careful visual examination) the data. We can place the data in a table, make charts of the data, and look for patterns of behavior or movement in the data. Of course, there are also formal mathematical techniques of analyzing data with descriptive or inferential statistics. Also, we can use powerful modeling techniques like Monte Carlo simulation and constrained optimization for analysis. We will see more of these topics later.

Once we have collected, summarized, and analyzed our data, we are ready for **presenting** data results. Much of what we have discussed in this chapter is related to graphical visualization of data and represents a distillation of our understanding of the data. The goal of visualization is to inform and influence our audience. If our preliminary steps are performed well, the visualization of results should be relatively straightforward.

With this simple model in mind—collect, summarize, analyze, and present—let's apply what we have learned to an example problem. We will begin with the collection of data, proceed to a data summarization phase, perform some simple analysis, and then select various graphical visualization formats that will highlight the insights we have gained.

2.5.1 Example—Tere's Budget for the 2nd Semester of College

This example is motivated by a concerned parent, Dad, monitoring the second semester college expenditures for his daughter, Tere (short for Teresa). Tere is in her first year of university, and her first semester expenditures have far exceeded

Dad's planned budget. Therefore, Dad has decided to monitor how much she spends during the second semester. The second semester will constitute a data collection period to study expenditures. Dad is skilled in data analytics, and whatever he learns from this semester will become the basis of his advice to Tere regarding her future spending. Table 2.4 provides a detailed breakdown of the expenditures that result from the 2nd semester: 60 expenditures that Tere incurred. Dad has set as his goal for the analysis the determination of how and why expenditures occur over time. The following sections take us step by step through the data analysis process, with special emphasis on the visualization of results.

2.5.2 *Collecting Data*

Dad meets with Tere to discuss the data collection effort. Dad convinces Tere that she should keep a detailed log of data regarding second semester expenditures, either paid for with a credit card or cash. Although Tere is reluctant (and upset), Dad convinces her that he will be fair in his analysis. Finally, they agree on a list of the most important issues and concerns he wants to address regarding expenditures:

1. What types of purchases are being made?
2. Are there spending patterns occurring during the week, month, and semester?
3. How are the payments of expenditures divided among the credit card and cash?
4. Finally, can some of the expenditures be identified as unnecessary?

To answer these questions, Dad assumes that each time an expenditure occurs, with either cash or credit card, an observation is generated. This is known as **transactional data**. Next, he selects 6 data fields to describe each observation: (1) the number identifier of the week (1–15) for the 15-week semester in which the expenditure occurs, (2) the date, (3) the weekday (Sunday = Sn, Monday = M, etc.) corresponding to the date, (4) the amount of the expenditure in dollars, (5) whether cash (C) or credit card (R) was used for payment, and finally, (6) one of three categories of expenditure types: food (F), personal (P), and school (S). Note that these data elements represent a wide variety of data types, from ratio data related to Amount, to categorical data representing food/personal/school, to interval data for the date. In Table 2.4 we see that the first observation in the first week was made by credit card on Sunday, January 6th for food in the amount \$111.46. Thus, we have collected our data and now we can begin to consider summarization.

2.5.3 *Summarizing Data*

Let's begin the process of data analysis with some basic exploration; this is often referred to as a *fishing expedition*. It is called a fishing expedition, because we simply want to perform a cursory examination of the expenditures with no particular

Table 2.4 2nd semester university student expenses

Obs.	Week	Date	Weekday	Amount	Cash/Credit Card	Food/Personal/School
1	Week 01	6-Jan	Sn	111.46	R	F
2		7-Jan	M	43.23	C	S
3		8-Jan	T	17.11	C	S
4		10-Jan	Th	17.67	C	P
5	Week 02	13-Jan	Sn	107.00	R	F
6		14-Jan	M	36.65	C	P
7		14-Jan	M	33.91	C	P
8		17-Jan	Th	17.67	C	P
9		18-Jan	F	41.17	R	F
10	Week 03	20-Jan	Sn	91.53	R	F
11		21-Jan	M	49.76	C	P
12		21-Jan	M	32.97	C	S
13		22-Jan	T	14.03	C	P
14		24-Jan	Th	17.67	C	P
15		24-Jan	Th	17.67	C	P
16	Week 04	27-Jan	Sn	76.19	R	F
17		31-Jan	Th	17.67	C	P
18		31-Jan	Th	17.67	C	P
19		1-Feb	F	33.03	R	F
20	Week 05	3-Feb	Sn	66.63	R	F
21		5-Feb	T	15.23	C	P
22		7-Feb	Th	17.67	C	P
23	Week 06	10-Feb	Sn	96.19	R	F
24		12-Feb	T	14.91	C	P
25		14-Feb	Th	17.67	C	P
26		15-Feb	F	40.30	R	F
27	Week 07	17-Feb	Sn	96.26	R	F
28		18-Feb	M	36.37	C	S
29		18-Feb	M	46.19	C	P
30		19-Feb	T	18.03	C	P
31		21-Feb	Th	17.67	C	P
32		22-Feb	F	28.49	R	F
33	Week 08	24-Feb	Sn	75.21	R	F
34		24-Feb	Sn	58.22	R	F
35		28-Feb	Th	17.67	C	P
36	Week 09	3-Mar	Sn	90.09	R	F
37		4-Mar	M	38.91	C	P
38		8-Mar	F	39.63	R	F
39	Week 10	10-Mar	Sn	106.49	R	F
40		11-Mar	M	27.64	C	S
41		11-Mar	M	34.36	C	P
42		16-Mar	S	53.32	R	S

(continued)

Table 2.4 (continued)

Obs.	Week	Date	Weekday	Amount	Cash/Credit Card	Food/Personal/School
43	<i>Week 11</i>	17-Mar	Sn	111.78	R	F
44		19-Mar	T	17.91	C	P
45		23-Mar	S	53.52	R	P
46	<i>Week 12</i>	24-Mar	Sn	69.00	R	F
47		28-Mar	Th	17.67	C	P
48	<i>Week 13</i>	31-Mar	Sn	56.12	R	F
49		1-Apr	M	48.24	C	S
50		4-Apr	Th	17.67	C	P
51		6-Apr	S	55.79	R	S
52	<i>Week 14</i>	7-Apr	Sn	107.88	R	F
53		8-Apr	M	47.37	C	P
54		13-Apr	S	39.05	R	P
55	<i>Week 15</i>	14-Apr	Sn	85.95	R	F
56		16-Apr	T	22.37	C	S
57		16-Apr	T	23.86	C	P
58		18-Apr	Th	17.67	C	P
59		19-Apr	F	28.60	R	F
60		20-Apr	S	48.82	R	S

analytical direction in mind other than becoming acquainted with the data. This initial process should then lead to more explicit directions for the analysis; that is, we will go where the fishing expedition leads us. Summarization of the data will be important to us at this stage. Figure 2.13 displays the data in a loose chronological order, and it does not provide a great deal of information, for a number of reasons. First, each successive observation does not correspond to a strict chronological order. For example, the first seven observations in Fig. 2.13 represent Sunday, Monday, Tuesday, Thursday, Sunday, Monday, and Monday expenditures, respectively. Thus, there are situations where no expenditures occur on a day, and there are days where multiple transactions occur. If Dad's second question about patterns of expenditures is to be answered, we will have to modify the data to include all days of the week and impose strict chronological order; thus, our chart should include days where there are no expenditures. When multiple daily transactions occur, we will have to decide on whether to aggregate data into a single day.

Table 2.5 displays a small portion of our expenditure data in this more rigid format, which has inserted days for which there are no expenditures. Note, for example, that a new observation has been added for Wednesday (now categorized as day 4), 9-Jan for zero dollars. Every day of the week will have an entry, although it may be zero dollars in expenditures, and there may be multiple expenditures on a day. Finally, although we are interested in individual expenditure observations, weekly, and even daily, totals could also be quite valuable. In summary, the original data collected needed substantial adjustment and summarization to organize it into more meaningful and informative data that will achieve our stated goals.

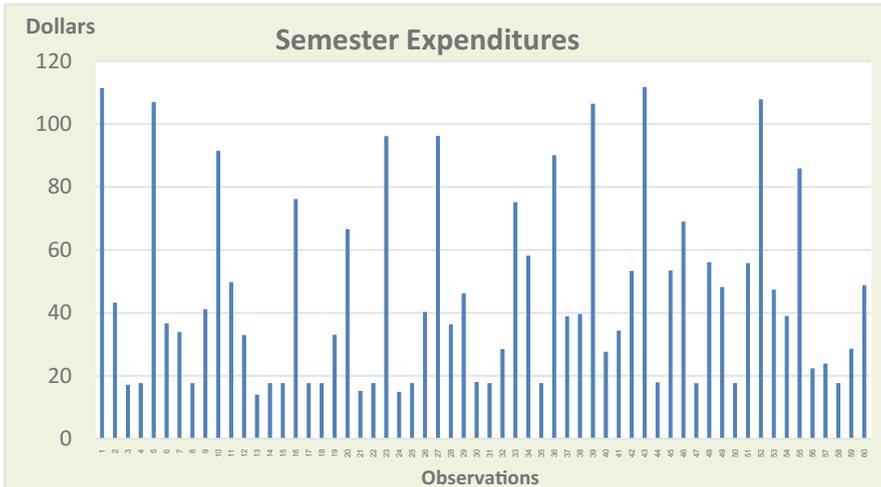


Fig. 2.13 Chronological (by transaction number) display of expenditure data

Let us assume that we have reorganized our data into the format shown in Table 2.5. As before, these data are arranged in columnar format and each observation has six fields plus an observation number. We have made one more change to the data in anticipation of the analysis we will perform. The Weekday field has been converted to a numerical value, with Sunday being replaced with 1, Monday with 2, etc. We will discuss the reason for this change later, but for now understand that it will permit easier analysis of the data.

2.5.4 Analyzing Data

Now, we are ready to look for insights in the data we have collected and summarized (that is, perform analysis). First, focusing on the dollar value of the observations, we see considerable variation in amounts of expenditures. This is not unexpected, given the relatively small number of observations in the semester. If we want to graphically analyze the data by type of payment (credit card or cash payments) *and* the category of expenditure (F, P, S), then we will have to further reorganize the data to provide this information. We will see that this can be managed with the **Sort tool** in the Data tab. The Sort tool permits us to *rearrange* our overall spreadsheet table of data observations into the observations of specific interest for our analysis.

Dad suspects that the expenditures for particular days of the week are higher than others from the data in Table 2.5. He begins by organizing the data according to day of the week—all Sundays (1), all Mondays (2), etc. To Sort the data by day, we first capture the entire data range we are interested in sorting. We can either include the header row that contains column titles (the fields names Weekday, Amount, etc.) or

Table 2.5 Portion of modified expenditure data including no expenditure days

Obs.	Week	Date	Weekday	Amount	Cash/Credit Card	Food/Personal/School
1	Week 01	6-Jan	1	111.46	R	F
2		7-Jan	2	43.23	C	S
3		8-Jan	3	17.11	C	S
		9-Jan	4	0.00		
4		10-Jan	5	17.67	C	P
		11-Jan	6	0.00		
		12-Jan	7	0.00		
5	Week 02	13-Jan	1	107.00	R	F
6		14-Jan	2	36.65	C	P
7		14-Jan	2	33.91	C	P
		15-Jan	3	0.00		
		16-Jan	4	0.00		
8		17-Jan	5	17.67	C	P
9		18-Jan	6	41.17	R	F
	19-Jan	7	0.00			
10	Week 03	20-Jan	1	91.53	R	F
11		21-Jan	2	49.76	C	P
12		21-Jan	2	32.97	C	S
13		22-Jan	3	14.03	C	P
		23-Jan	4	0.00		

not include them. If the header row is not captured, then it is assumed that the row immediately above the data is the header field names. Then, we select the Sort tool in the Sort and Filter group of the Data tab. Sort permits us to set sort keys (the titles in the header row) that can then be selected, as well as an option for executing ascending or descending sorts. An ascending sort of text arranges data in ascending alphabetical order (a to z) and an ascending sort of numerical data is analogous. Now, we can see that converting the Weekday field to a numerical value ensures a Sort that places in ascending order. If the field values had remained Sn, M, etc., the sort would lead to an alphabetic sort and loss of the consecutive order of days—Friday as day 1 and Wednesday as day 7.

Figure 2.14 shows the data sort procedure for our original data. We begin by capturing the spreadsheet range of interest containing the observations. In the Sort and Filter group, we select the Sort tool. Figure 2.14 shows the dialog boxes that ask the user for the key for sorting the data. The key used is Day #. The complete sorted data are shown in Table 2.6. As you can see in Table 2.6, the first 16 sorted observations are for Sunday (Day 1).

At this point, our data have come a long way from the original 60 basic observations and are ready to reveal some expenditure behavior. First, notice in Table 2.6 that all expenditures on Sunday are for food (F), made with a credit card, and are generally the highest \$ values. This pattern occurs every Sunday of every week in the data. Immediately, Dad is alerted to this curious behavior—is it possible that Tere

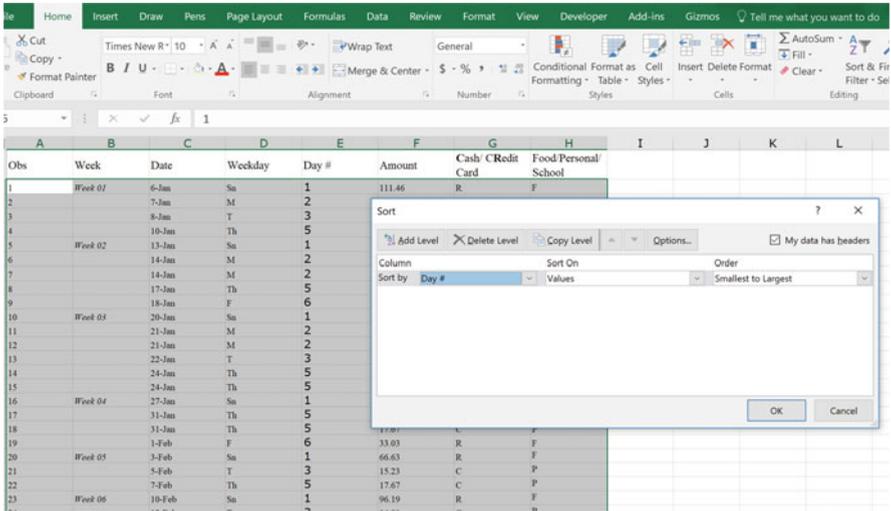


Fig. 2.14 Data sort procedure

reserves grocery shopping for Sundays? Also, note that Monday’s cash expenditures are of lesser value and never for food. Additionally, there are several multiple Monday expenditures, and they occur irregularly over the weeks of the semester. There does not appear to be much of a regular pattern of behavior for Mondays.

Figure 2.15 provides a graph of this Sunday and Monday data comparison, and Fig. 2.16 compares Sunday and Thursday. In each case, Dad has organized the data series by the specific day of each week. Also, he has aggregated multiple expenditures for a single day, such as Monday, Jan-14 expenditures of \$33.91 and \$36.65 (total \$70.56). The Jan-14 quantity can be seen in Fig. 2.15 in week 2 for Monday, and this has required manual summarization of the data in Table 2.6. Obviously, there are many other possible daily comparisons that can be performed, and they will also require manual summarization.

Now, let’s summarize some of Dad’s early findings. Below are some of the most obvious results:

- 1) All Sunday expenditures (16 observations) are high dollar value, Credit Card, Food, and occur consistently every Sunday.
- 2) Monday expenditures (12) are Cash, School, and Personal, and occur frequently, though less frequently than Sunday expenditures.
- 3) Tuesday expenditures (8) are Cash and predominantly Personal.
- 4) There are no Wednesday (0) expenditures.
- 5) Thursday expenditures (13) are all Personal, Cash, and exactly the same value (\$17.67), although there are multiple expenditures on some Thursdays.
- 6) Friday expenditures (6) are all for Food and paid with Credit Card.
- 7) Saturday expenditures (5) are Credit Card and a mix of School and Personal.

Table 2.6 Modified expenditure data sorted by weekday and date

Date	Weekday	Amount	Cash/C <u>R</u> edit Card	Food/ <u>P</u> ersonal/ <u>S</u> chool
6-Jan	1	111.46	R	F
13-Jan	1	107	R	F
20-Jan	1	91.53	R	F
27-Jan	1	76.19	R	F
3-Feb	1	66.63	R	F
10-Feb	1	96.19	R	F
17-Feb	1	96.26	R	F
24-Feb	1	58.22	R	F
24-Feb	1	75.21	R	F
3-Mar	1	90.09	R	F
10-Mar	1	106.49	R	F
17-Mar	1	111.78	R	F
24-Mar	1	69	R	F
31-Mar	1	56.12	R	F
7-Apr	1	107.88	R	F
14-Apr	1	85.95	R	F
7-Jan	2	43.23	C	S
14-Jan	2	33.91	C	P
14-Jan	2	36.65	C	P
21-Jan	2	32.97	C	S
21-Jan	2	49.76	C	P
28-Jan	2	0		
4-Feb	2	0		
11-Feb	2	0		
18-Feb	2	36.37	C	S
18-Feb	2	46.19	C	P
25-Feb	2	0		
4-Mar	2	38.91	C	P
11-Mar	2	27.64	C	S
11-Mar	2	34.36	C	P
18-Mar	2	0		
25-Mar	2	0		
1-Apr	2	48.24	C	S
8-Apr	2	47.37	C	P
15-Apr	2	0		
8-Jan	3	17.11	C	S
15-Jan	3	0		
22-Jan	3	14.03	C	P
29-Jan	3	0		
5-Feb	3	15.23	C	P
12-Feb	3	14.91	C	P
19-Feb	3	18.03	C	P

(continued)

Table 2.6 (continued)

Date	Weekday	Amount	Cash/Credit Card	Food/Personal/School
26-Feb	3	0		
5-Mar	3	0		
12-Mar	3	0		
19-Mar	3	17.91	C	P
26-Mar	3	0		
2-Apr	3	0		
9-Apr	3	0		
16-Apr	3	22.37	C	S
16-Apr	3	23.86	C	P
9-Jan	4	0		
16-Jan	4	0		
23-Jan	4	0		
30-Jan	4	0		
6-Feb	4	0		
13-Feb	4	0		
20-Feb	4	0		
27-Feb	4	0		
6-Mar	4	0		
13-Mar	4	0		
20-Mar	4	0		
27-Mar	4	0		
3-Apr	4	0		
10-Apr	4	0		
17-Apr	4	0		
10-Jan	5	17.67	C	P
17-Jan	5	17.67	C	P
24-Jan	5	17.67	C	P
24-Jan	5	17.67	C	P
31-Jan	5	17.67	C	P
31-Jan	5	17.67	C	P
7-Feb	5	17.67	C	P
14-Feb	5	17.67	C	P
21-Feb	5	17.67	C	P
28-Feb	5	17.67	C	P
7-Mar	5	0		
14-Mar	5	0		
21-Mar	5	0		
28-Mar	5	17.67	C	P
4-Apr	5	17.67	C	P
11-Apr	5	0		
18-Apr	5	17.67	C	P
11-Jan	6	0		

(continued)

Table 2.6 (continued)

Date	Weekday	Amount	Cash/Credit Card	Food/Personal/School
18-Jan	6	41.17	R	F
25-Jan	6	0		
1-Feb	6	33.03	R	F
8-Feb	6	0		
15-Feb	6	40.3	R	F
22-Feb	6	28.49	R	F
1-Mar	6	0		
8-Mar	6	39.63	R	F
15-Mar	6	0		
22-Mar	6	0		
29-Mar	6	0		
5-Apr	6	0		
12-Apr	6	0		
19-Apr	6	28.6	R	F
12-Jan	7	0		
19-Jan	7	0		
26-Jan	7	0		
2-Feb	7	0		
9-Feb	7	0		
16-Feb	7	0		
23-Feb	7	0		
2-Mar	7	0		
9-Mar	7	0		
16-Mar	7	53.32	R	S
23-Mar	7	53.52	R	P
30-Mar	7	0		
6-Apr	7	55.79	R	S
13-Apr	7	39.05	R	P
20-Apr	7	48.82	R	S

- 8) The distribution of the *number* of expenditure types (Food, Personal, and School) is not proportionate to the *dollars spent* on each type. (See Figs. 2.17 and 2.18). Food accounts for fewer numbers of expenditures (36% of total) than personal, but for a greater percentage (60%) of the total dollar of expenditures.

2.5.5 Presenting Data

Figures 2.13, 2.14, 2.15, 2.16, 2.17, and 2.18 and Tables 2.4, 2.5, and 2.6 are a few examples of the many possible graphs and data tables that can be presented to explore the questions originally asked by Dad. Each graph requires data preparation

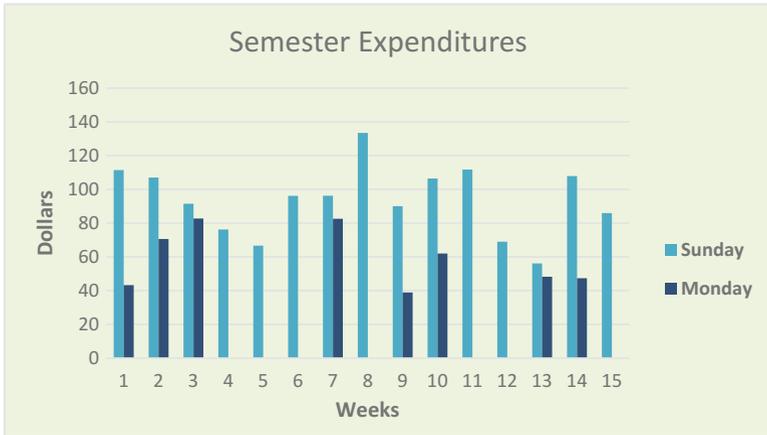


Fig. 2.15 Modified expenditure data sorted by Sunday and Monday

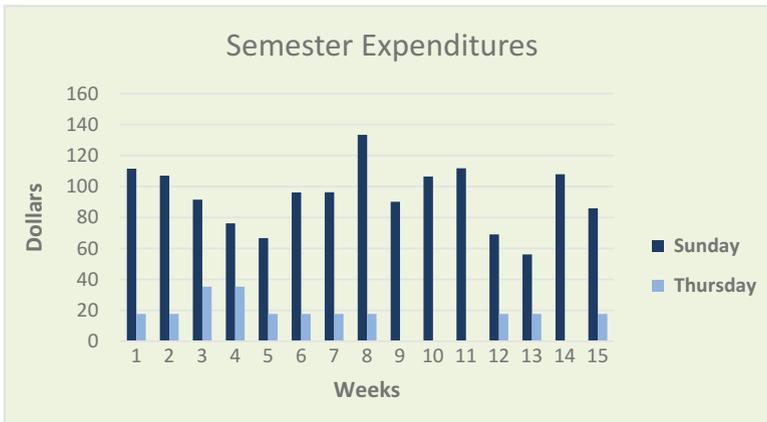


Fig. 2.16 Modified expenditure data sorted by Sunday and Thursday

to fit the analytical goal. For example, the construction of the pie charts in Figs. 2.17 and 2.18 required that we count the number of expenditures of each type (Food, School, and Personal) and that we sum the dollar expenditures for each type, respectively.

Dad is now able to examine Tere’s buying patterns more closely, and through discussion with Tere, he finds some interesting behavior related to the data he has assembled:

- 1) The \$17.67 Thursday expenditures are related to Tere’s favorite personal activity—having a manicure and pedicure. The duplicate charge on a single Thursday represents a return to have her nails redone once she determines she is not happy

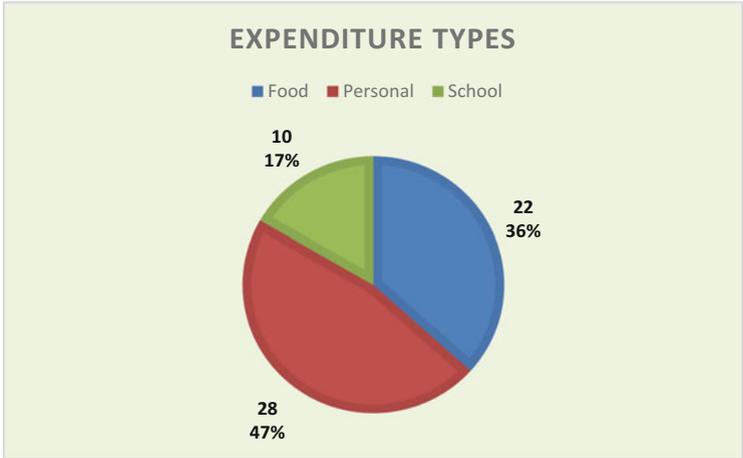


Fig. 2.17 Number of expenditure types

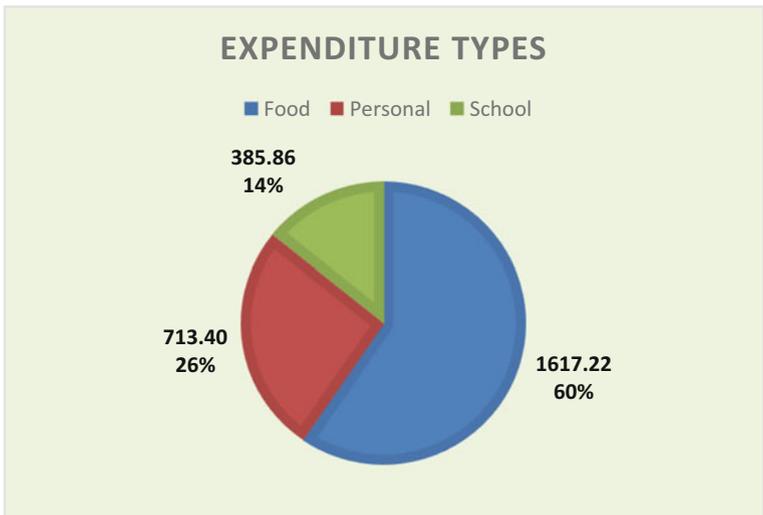


Fig. 2.18 Dollar expenditures by type

with the first outcome. Dad believes she should ask for a refund or free pedicure and manicure.

- 2) Sunday expenditures are dinners (not grocery shopping as Dad initially believed) at her favorite sushi restaurant. The dollar amount of each expenditure is always high because she treats her friends, Dave and Suzanne, to dinner. Dad determines that this is a magnanimous, but fiscally irresponsible, gesture. She agrees with Dad that she should stop paying for her friends.

- 3) There are no expenditures on Wednesday, because she has class all day long and can do little but study and attend class. Dad is happy with Wednesdays.
- 4) To avoid carrying a lot of cash, Tere generally prefers to use a credit card for larger expenditures. She is adhering to a bit of advice received from Dad for her own personal security. Dad is pleased she is taking his advice and carrying little cash.
- 5) She makes fewer expenditures near the end of the week because she is generally exhausted by her school work. Sunday dinner is a form of self-reward that she has established at the start to a new week. Dad has no problem with the reward, but he is concerned about extending the reward to her friends.
- 6) Friday food expenditures, she explains to Dad, are due to grocery shopping. Dad sees food shopping as an alternative to eating out.

Once Dad has obtained this information, he negotiates several money-saving concessions. First, she agrees to not treat Dave and Suzanne to dinner every Sunday; every other Sunday is sufficiently generous. She agrees to reduce her manicure visits to every other week, and she also agrees that cooking for her friends is equally entertaining, and possibly more socially meaningful, than eating out.

We have not gone into great detail on the preparation of data to produce Figs. 2.15, 2.16, 2.17, and 2.18, other than the sorting exercise we performed. Later in Chap. 4, we will learn to use the Filter and Advanced Filter capabilities of Excel. This will provide a simple method for preparing our data for graphical presentation/visualization.

2.6 Some Final Practical Graphical Presentation Advice

This chapter has presented a number of topics related to graphical presentation and visualization of quantitative data. Many of the topics are an introduction to data analysis, which we will visit in far greater depth in later chapters. Before we move on, let me leave you with a set of suggestions that might guide you in your presentation choices. Over time you will develop a sense of your own presentation style and preferences for presenting data in effective formats. Don't be afraid to experiment as you explore your own style and taste.

1. *Some charts and graphs deserve their own worksheet*—Often a graph fits nicely on a worksheet that contains the data series that generate the graph. But also, it is quite acceptable to dedicate a separate worksheet to the graph if the data series make viewing the graph difficult or distracting. This is particularly true when the graph represents the important *results* of a worksheet. (Later we will discuss static versus dynamic graphs, which make the choice relatively straightforward.)
2. *Axis labels are essential*—Some creators of graphs are lax in the identification of graph axes, both with the units associated with the axis scale and the verbal description of the axis dimension. Because they are intimately acquainted with

the data generating the graph, they forget that the viewer may not be similarly acquainted. Always provide clear and concise identification of axes, and remember that you are not the only one who will view the graph.

3. *Scale differences in values can be confusing*—Often graphs are used as tools for visual comparison. Sometimes this is done by plotting multiple series of interest on a single graph or by comparing individual series on separate graphs. In doing so, we may not be able to note series behavior due to scale differences for the graphs. This suggests that we may want to use multiple scales on a graph to compare several series. This is shown in Fig. 2.19. Select series 1 by right-clicking on the line, and then select the Secondary Axis (see arrow) in the Format Data series window. Additionally, if we display series on separate graphs, we can impose similar scales on the graphs to facilitate equitable comparison. Being alert to these differences can seriously affect our assessment of results.
4. *Fit the Chart Type by considering the graph's purpose*—The choice of the chart type should invariably be guided by one principle—*keep it simple*. There are often many ways to visualize data, whether the data are cross-sectional or time series. Consider the following ideas and questions relating to chart type selection.

Time Series Data (data related to a time axis)

- (a) Will the data be displayed over a chronological time horizon? If so, it is considered time series data.
- (b) In business or economics, time series data are invariably displayed with time on the horizontal axis.
- (c) With time series, we can either display data discretely (bars) or continuously (lines and area). If the flow or continuity of data is important, then Line and Area graphs are preferred. Be careful that viewers not assume that they can locate values between time increments, if these intermediate values are not meaningful.

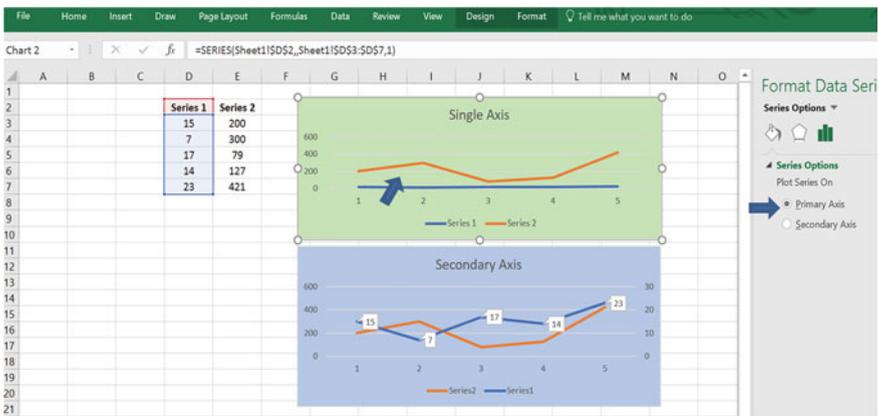


Fig. 2.19 Using secondary axis with data of varying magnitudes

Cross-sectional Data Time Snap-shot or (time dimension is not of primary importance)

- (a) For data that is a single snap-shot of time or where time is not our focus, column or bar graphs are used most frequently. If you use column or bar graphs, it is important to have category titles on axes (horizontal or vertical). If you do not use a column or bar graph, then a Pie, Doughnut, Cone, or Pyramid graph may be appropriate. Line graphs are usually not advisable for cross-sectional data.
- (b) Flat Pie graphs are far easier to read and interpret than 3-D Pie graphs. Also, when data result in several very small pie segments, relative to others, then precise comparisons can be difficult.
- (c) Is the categorical order of the data important? There may be a natural order in categories that should be preserved in the data presentation—e.g. the application of chronologically successive marketing promotions in a sales campaign.
- (d) Displaying multiple series in a Doughnut graph can be confusing. The creation of Doughnuts within Doughnuts can lead to implied proportional relationships which do not exist.

Co-related Data (Correlation Analysis)

- (a) Scatter diagrams are excellent tools for viewing the co-relationship (correlation in statistical jargon) of one variable to another. They represent two associated data items on a two-dimensional surface—e.g. the number of housing starts in a time period and the corresponding purchase of plumbing supplies.
- (b) Bubble diagrams assume that the two values discussed in scatter diagrams also have a third value (relative size of the bubble) that relates to the frequency or strength of the point located on two dimensions—e.g. a study that tracks combinations of data pairs that occur frequently. In this case, the size of the bubble is the frequency of the occurrence of specific combinations. This chart is found in the Scatter Diagram group. See Fig. 2.20.

General Issues

- (a) Is the magnitude of a data value important relative to other data values occurring in the same category or at the same time? (This was the case in Fig. 2.4.) If so, then consider the Stacked and 100% Stacked graph. The Stacked graph preserves the option to compare *across* various time periods or categories—e.g. the revenue contribution of five categories of products for eight quarters provides not only the relative importance of products within a quarter, but also shows how the various quarters compare. Note that this last feature (comparison across quarters) will be lost in a 100% Stacked graph. See Figs. 2.4 and 2.21
- (b) In general, I find that 3-D graphs can be potentially distracting. The one exception is the display of multiple series of data (usually less than five or six) where the overall pattern of behavior is important to the viewer. Here a 3-D Line graph (ribbon graph) or an Area graph is appropriate, as long as the series do not

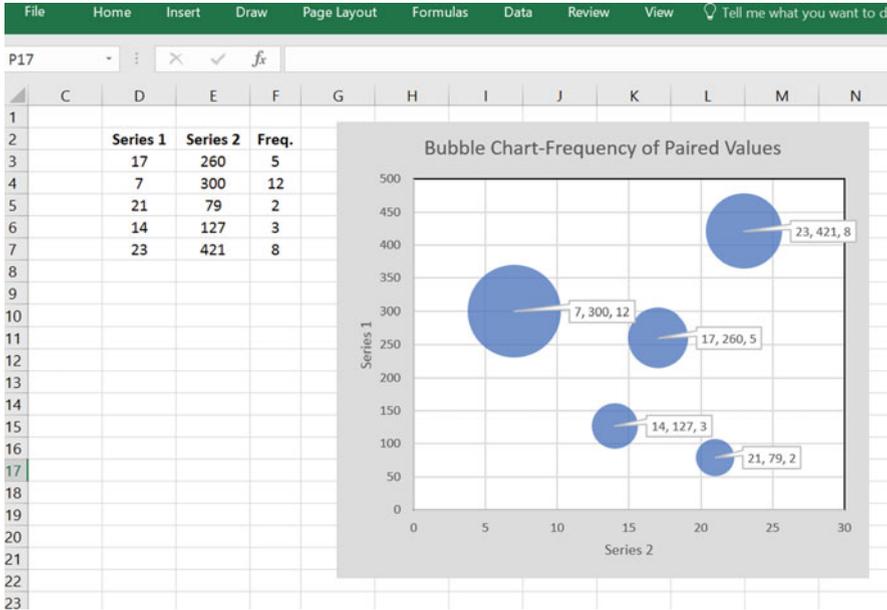


Fig. 2.20 Bubble chart—frequency of data pairs

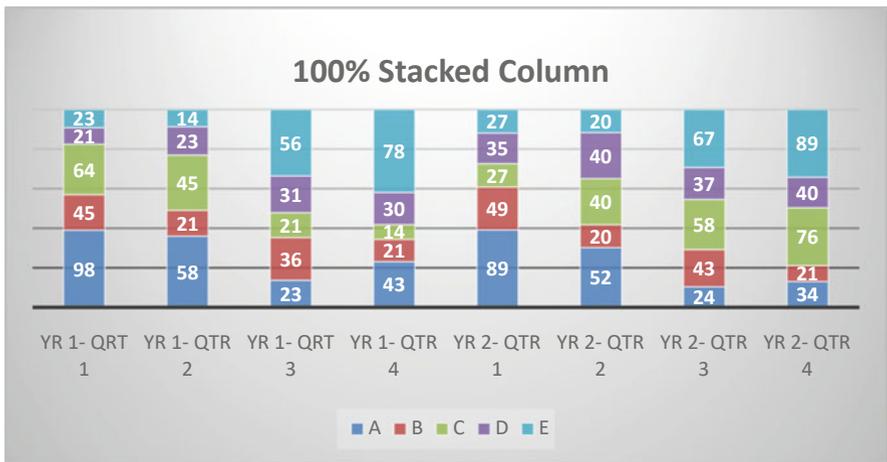


Fig. 2.21 100% stacked column graph of 5 product data

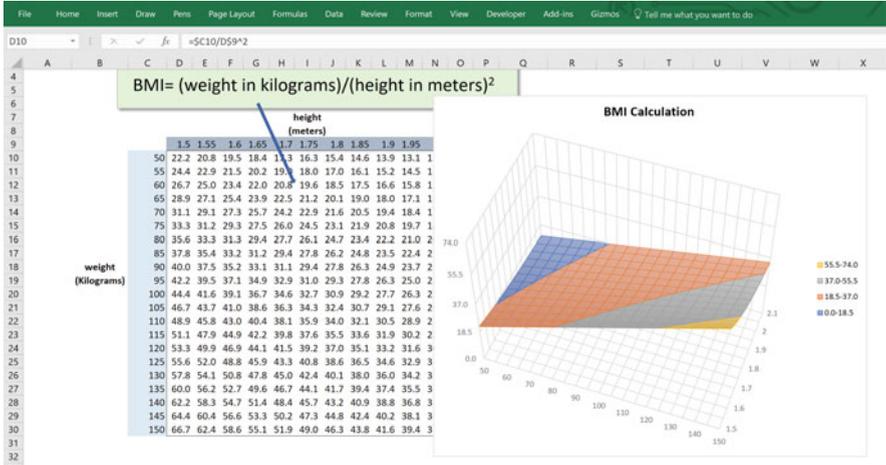


Fig. 2.22 3-D surface graph of BMI calculation

obscure the view of series with lesser values. If a 3-D graph is still your choice, exercise the 3-D View options that reorient the view of the graph or point and grab a corner of the graph to rotate the axes. This may clarify the visual issues that make a 3-D graph distracting. See Figs. 2.5 and 2.7. The 3-D Surface graph is a sophisticated visualization of data on three axes. See Fig. 2.22 for an example based on the calculation of Body Mass Index (BMI).

- (c) It may be necessary to use several chart types to fully convey the desired information. Don't be reluctant to organize data into several graphical formats; this is more desirable than creating a single, overly complex graph.
- (d) Once again, it is wise to invoke a philosophy of simplicity and parsimony—less is often more.

2.7 Summary

In the next chapter, we will concentrate on numerical *analysis* of quantitative data. Chap. 3 and the two chapters that follow contain techniques and tools that are applicable to the material in this chapter. You may want to return and review what you have learned in this chapter in light of what is to come; this is good advice for all chapters. It is practically impossible to present all the relevant tools for analysis in a single chapter, so I have chosen to “spread the wealth” among the seven chapters that remain.

Key Terms

Ratio data	Stacked column
Interval data	3-D column
Categorical/nominal data	Line chart
Ordinal data	Time series data
Data visualization	Lags
Data context	Linear trend
Records	Systematic behavior
Fields	Horizontal (category) Axis labels
Comment field	Select data, format chart area, chart type
SQL	Add chart elements
Relational database	Chart title
Charts and graphs	Data analysis
Chart wizard	Collecting
Ribbons	Summarizing
Tabs	Analyzing
Groups	Eyeballing
Chart tools	Presenting
Data range	Transactional data
Column or Bar chart	Fishing expedition
Negative correlation	Sort tool
Positive correlation	Sort keys

Problems and Exercises

- Consider the data in Table 2.3 of this chapter.
 - Replicate the charts that appeared in the chapter and attempt as many other chart types and variations as you like. Use new chart types—pie, doughnut, pyramid, etc.—to see the difference in appearance and appeal of the graphs.
 - Add another series to the data for a new product, F. What changes in graph characteristics are necessary to display this new series with A–E? (Hint: scale will be an issue in the display).

F
425
560
893
1025
1206
837
451
283

2. Can you find any interesting relationships in Tere’s expenditures that Dad has not noticed?
3. Create a graph similar to Figs. 2.15 and 2.16 that compares *Friday* and *Saturday*.
4. Perform a single sort of the data in Table 2.6 to reflect the following three conditions: 1st—credit card expenditures, 2nd—in chronological order, 3rd—if there are multiple entries for a day, sort by quantity in ascending fashion.
5. Create a pie chart reflecting the proportion of all expenditures related to *Food*, *Personal*, and *School* for Dad and Tere’s example.
6. Create a scatter diagram of *Day #* and *Amount* for Dad and Tere’s example.
7. The data below represent information on bank customers at 4 branch locations, their deposits at the branch, and the percent of the customers over 60 years of age at the branch. Create graphs that show: (1) line graph for the series No. Customers and \$ Deposits for the various branches and (2) pie graphs for each quantitative series. Finally, consider how to create a graph that incorporates all the quantitative series (hint: bubble graph).

Branch	No. customers	\$ Deposits	Percent of customers over 60 years of age
A	1268	23,452,872	0.34
B	3421	123,876,985	0.57
C	1009	12,452,198	0.23
D	3187	97,923,652	0.41

8. For the following data, provide the summarization and manipulation that will permit you to sort the data by day-of-the-week. Thus, you can sort all Mondays, Tuesdays, etc. (hint: you will need a good day calculator).

Last name, First name	Date of birth	Contribution
Laffercar, Carole	1/24/76	10,000
Lopez, Hector	9/13/64	12,000
Rose, Kaitlin	2/16/84	34,500
LaMumba, Patty	11/15/46	126,000
Roach, Tere	5/7/70	43,000
Guerrero, Lili	10/12/72	23,000
Bradley, James	1/23/48	100,500
Mooradian, Addison	12/25/97	1000
Brown, Mac	4/17/99	2000
Gomez, Pepper	8/30/34	250,000
Kikapoo, Rob	7/13/25	340,000

9. *Advanced Problem*—Isla Mujeres is an island paradise located very near Cancun, Mexico. The island government has been run by a prominent family, the Murillos, for most of four generations. During that time, the island has become a major tourist destination for many foreigners and Mexicans. One evening, while

vacationing there, you are dining in a local restaurant. A young man seated at a table next to yours overhears you boasting about your prowess as a quantitative data analyst. He is local politician that is running for the position of Island President, the highest office on Isla Mujeres. He explains how difficult it is to unseat the Murillos, but he believes that he has some evidence that will persuade voters that it is time for a change. He produces a report that documents quantitative data related to the island's administration over 46 years. The data represent 11 4-year presidential terms and the initial 2 years of the current term. Presidents are designated as A–D, for which all are members of the Murillo clan, except for B. President B is the only non-Murillo to be elected and was the uncle of the young politician. Additionally, all quantities have been converted to 2008 USD (US Dollars)

- The raw data represent important economic development relationships for the Island. How will you use the raw data to provide the young politician information on the various presidents that have served the Island? Hint—Think as an economist might, and consider how the president's investment in the island might lead to improved economic results.
- Use your ideas in a. to prepare a graphical analysis for the young politician. This will require you to use the raw data in different and clever ways.
- Compare the various presidents, through the use of graphical analysis, for their effectiveness in running the island. How will you describe the young politician's Uncle?
- How do you explain the changes in Per Capita Income given that it is stated in 2008 dollars? Hint—There appears to be a sizable increase over time. What might be responsible for this improvement?

Years	President	Municipal tax collected	Salary of island president	Island infrastructure investment	Per capita income
1963–1966	A	120,000	15,000	60,000	1900
1967–1970	A	186,000	15,000	100,000	2100
1971–1974	A	250,000	18,000	140,000	2500
1975–1978	B	150,000	31,000	60,000	1300
1979–1982	B	130,000	39,000	54,000	1000
1983–1986	C	230,000	24,000	180,000	1800
1987–1990	C	310,000	26,000	230,000	2300
1991–1994	C	350,000	34,000	225,000	3400
1995–1998	C	450,000	43,000	320,000	4100
1999–2002	D	830,000	68,000	500,000	4900
2003–2006	D	1,200,000	70,000	790,000	5300
2007–2008 ^a	D	890,000	72,000	530,000	6100

^aNote: This is 2 years of data versus 4 years