



CREATING AND USING GRAPHS

INTRODUCTION TO SPSS GRAPHS

Marilyn vos Savant made it into the Guinness Book of World Records as the person with the highest IQ ever recorded. She once said about statistics that they “can be used to support or undercut almost any argument.” Another brainy person (Albert Einstein) said, “Everything should be made as simple as possible, but not simpler.” Taking these two ideas together, the purpose of displaying data in the form of a graph should be to simplify information and never to exaggerate or deceive.

Graphs are often used in conjunction with statistical procedures to visually explain or support findings from hypothesis testing or descriptive studies. To that end, this chapter explains how to use SPSS to display various types of graphs in ways that tell an accurate story of your data. To use graphs effectively (and truthfully), we present the following guidelines.

GUIDELINES FOR CREATING AND USING GRAPHS

There are (at least) two uses for graphs. First, they can be useful for identifying problems or interesting data points in your data set. Second, when reporting your results, graphs are useful for clarifying findings. Example 2.1 in the previous chapter illustrated the use of graphs to visualize the distribution of data. This chapter concentrates on the use of graphs as a way of describing statistical results.

SIDEBAR

Use of graphs as alternatives to tables is especially recommended when giving presentations (using PowerPoint, PDF reports, etc.).

In general, graphs should be used in reports or presentations as an alternative to tables when the table would contain too many numerical entries to be easily understood or when the graph more clearly illustrates your results. Any number of textbooks and journal specifications contain

guidelines relating to the use of graphs (see Tufte, 2001). Here are a few general guidelines for using and reporting graphs:

1. Use simple graphs when possible. Avoid three-dimensional graphs since they often distort your message and contain spurious and distracting information.
2. Label all plots and axes clearly.
3. When creating two or more graphs that will be compared in some way, the range of values for each axis on every graph should be the same. Axes should generally begin with zero if that is the natural minimum. Otherwise, use the minimum value of the measurement as the minimum value for the axis.
4. Axis intervals on plots should be equal.
5. Many statisticians recommend the use of bar charts instead of pie charts. (In *The Visual Display of Quantitative Data*, Edward Tufte [2001] wrote, “The only worse design than a pie chart is several of them” [p. 178].)
6. Stick with standard charts when possible. Avoid custom complex charts that attempt to display several messages at once.

As IBM SPSS evolved over the years, there have been several generations of graphs included in the program. In the Graphs menu, you will find the following graphing methods that allow you to create graphs:

- Chart Builder
- Graphboard Template Chooser
- Legacy Dialog (Plots)

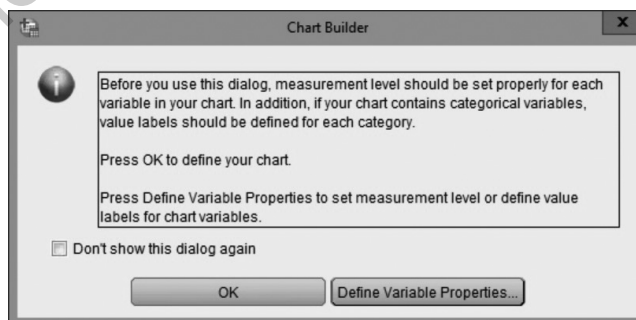
These three graphing methods often create exactly the same plots, but there are also differences. There are some plots that are not available under all graphing methods, and for similar plots across graphing methods, there may be options in one method that are not in the others. Each method has a different flavor and origin. Some of these differences will be illustrated in the upcoming examples. However, our intention is not to illustrate the differences or similarities among graphing methods for each type of graph. Our purpose is to provide you with the basics for producing standard graphs that should enable you to figure out how to create the graphs that are not covered here.

In addition, a number of statistical procedures have plots “built in” to the analyses. This chapter concentrates on the options in the Graph menu. All three types of graphing methods are illustrated in Example 3.1 (scatterplot), and following examples will (mostly) use the Chart Builder to illustrate how to create graphs.

CHART BUILDER

The SPSS Chart Builder is an interactive window that allows you to drag and click on options to build a chart by specifying variables from your data and features of the chart. When you select **Graphs/Chart Builder** the first time, Figure 3.1 is displayed. This message warns you that the SPSS data types for all of your variables must be set up properly (scale, ordinal, or nominal in SPSS terminology) and that any categorical variables should include category labels.

FIGURE 3.1 Chart Builder Warning Message



You can select the checkbox “Don’t show the dialog again” to avoid seeing this box in the future. If needed, set up the measure types (scale, ordinal, or nominal) for variables by clicking on the “Define Variable Properties” or by selecting the Variable View tab on the data screen.

When you click OK on the initial Chart Builder Box (while a dataset is open), the main Chart Builder interactive dialog box appears as shown in Figure 3.2. In upcoming examples, we will illustrate how to select variables, chart types, and options to create the graph you need.

GRAPHBOARD TEMPLATE CHOOSER

The Graphboard Template Chooser is a second method for creating SPSS charts or graphs. Using this method, you can select from ready-made templates called “Graphboard Visualizations” that contain graphs, charts, and plots and customize the plots appropriate for your variables. When you select **Graphs/Graphboard Template Chooser**, you see the dialog box shown in Figure 3.3. Upcoming examples will illustrate how to use this dialog box.

FIGURE 3.2 Chart Builder Interactive Dialog Box

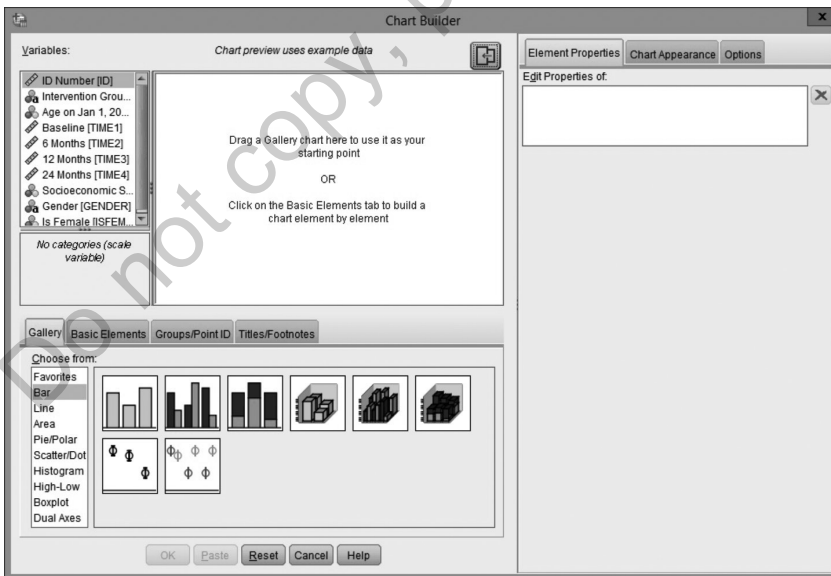
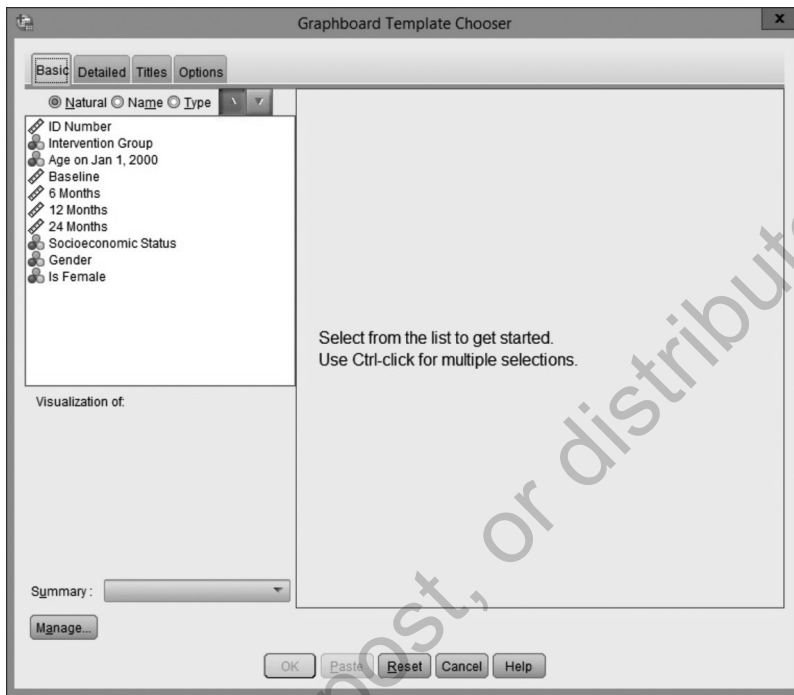


FIGURE 3.3 Graphboard Template Chooser



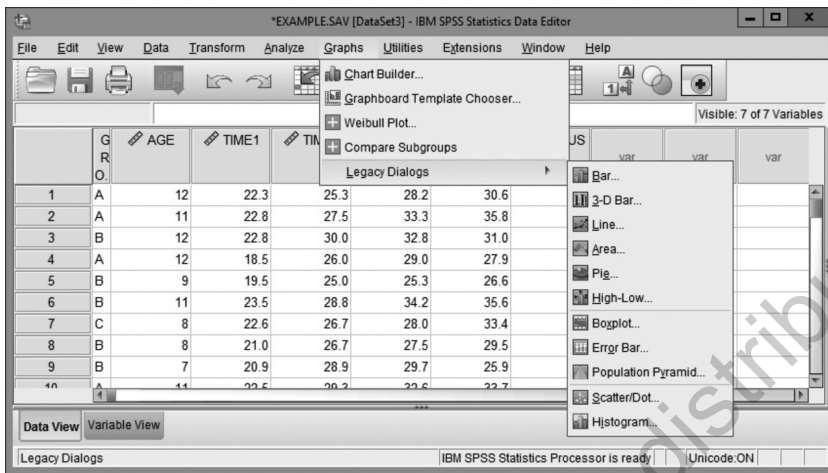
LEGACY PLOTS

Legacy Dialog plots are, as their name implies, plots that have been around a while and continue to be used in current SPSS versions, although the newer (Chart Builder and Graphboard) options are meant to replace most of these old procedures. However, the older plots are still useful, and some users prefer them to the newfangled options. When you select the **Graphs/Legacy Dialogs** method, submenus appear as shown in Figure 3.4.

SCATTERPLOTS

A scatterplot displays the relationship between two scale (quantitative) variables on an x - y coordinate system. It is often used in conjunction with a correlation or regression analysis.

FIGURE 3.4 Legacy Dialog Plots



Appropriate Applications for a Scatterplot

The following are examples of situations in which a scatterplot might be appropriate to visually display a relationship between variables:

- *Examine the relationship between grades on the midterm Chemistry exam and the final.* You are interested in seeing if people who received high scores on the midterm also get high scores on the final.
- *Does running help SBP?* A number of subjects' systolic blood pressures (SBP) are measured at baseline. Over a few weeks, you observe average time spent on a treadmill and observe differences in SBP. Is the amount of time spent on the treadmill related to differences in SBP?
- *Are opinions about welfare related to age?* You want to visualize this relationship to see if older people are more favorable to a proposed change than younger people.

Design Considerations for a Scatterplot

The two variables of interest in a scatterplot should be scale variables or at least ordinal. If you want to test hypotheses about the existence of a linear relationship between two variables, they should both be scale (quantitative) and normally distributed.

EXAMPLE 3.1

Chart Builder Scatterplot Example

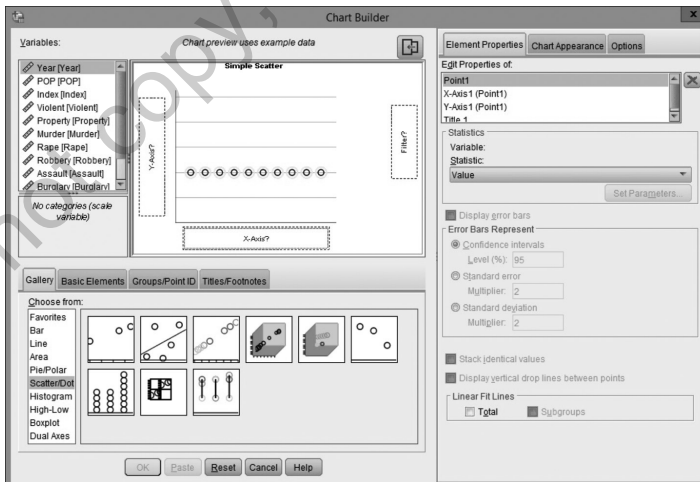
SPSS Step-by-Step. EXAMPLE 3.1a: Creating a Scatterplot Using the Chart Builder

In this example, we will examine crime data from Washington, DC for the years from 1978 to 2012.

1. Open the data DC_CRIME.SAV, and select **Graphs/Chart Builder**. If a warning message (as illustrated in Figure 3.1) appears, click OK.
2. On the Chart Builder Dialog, select **Scatter/Dot** from the list of graph types in

the lower left of the dialog. Locate the Simple Scatter option in the list of graph types (the upper left graph icon), and drag that option into the Chart Preview window at the top of the dialog. When you place the Simple Scatter icon in the Chart Preview box, notice that a new section of the dialog appears to the right of the original dialog, labeled “Element Properties.” These combined dialog boxes are shown in Figure 3.5.

FIGURE 3.5 Initial Chart Builder Selection for a Simple Scatterplot



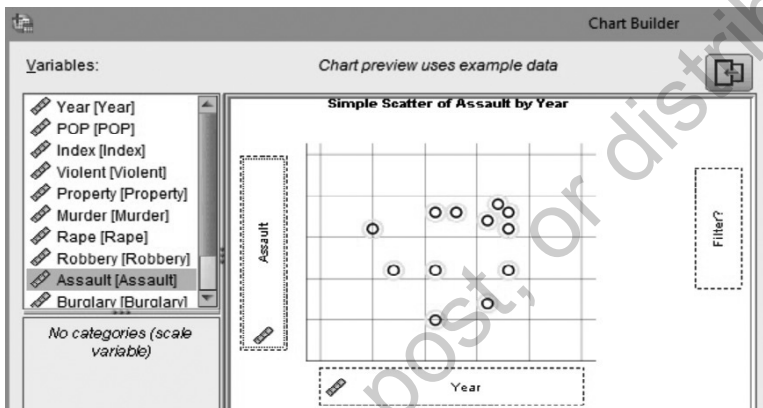
(Continued)

(Continued)

- From the Variables list on the left of the dialog, drag the variable *Year* to the *X*-axis position at the bottom of the preview graph. Similarly, drag

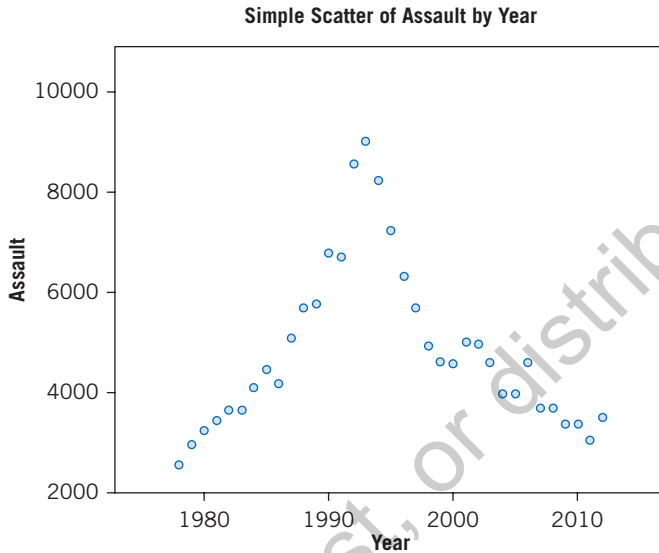
Assault to the *Y*-axis position. The results are shown in Figure 3.6. (The sample plot shown on the screen is not of the actual data at this point.)

FIGURE 3.6 Preview of Graph With Year and Assault



- Click OK. The desired scatterplot is shown in Figure 3.7 where you may observe that Assaults peaked around 1994 and have declined since.
- To modify this graph, select **Graphs/Chart Builder** again, and the dialog box reappears with all of the information you have previously selected still there. Select the **Title/Footnotes** tab just below the chart viewer. The Title 1 checkbox has by default

been selected, and the graph has the default title, "Simple Scatter of Assault by Year." To change this title, click on the Title 1 checkbox to deselect it and remove the title. Now click the Title 1 checkbox and a "Custom" text box appears to the right where you can enter a title for the chart. Enter "DC Crime Statistics 1978 to 2012," and click OK. A new version of the chart appears with the new title at the top. (In some older

FIGURE 3.7 Scatterplot of Year by Assault

versions of the program, you may not need to delete an existing title before adding your new title.)

- Other options (tabs) in the Chart Builder dialog allow you to make additional changes. For example, in the Elements Properties dialog (on the right), select Y-axis, and change the axis label to "Criminal Assaults." You can select the Basic Elements tab and then Transpose to transpose the x and y axes. There are many more

options you can select that are not illustrated here. Once you make any changes and select OK, the new graph appears with those changes made.

Basic chart types supported by Chart Builder include bar, line, area, pie/polar, scatter/dot, histogram, high-low, boxplots, and dual axes plots. Each of these plots can be edited in a Chart Editor to select other options to modify basic plots. (The Chart Editor is introduced in upcoming examples.)

(Continued)

(Continued)

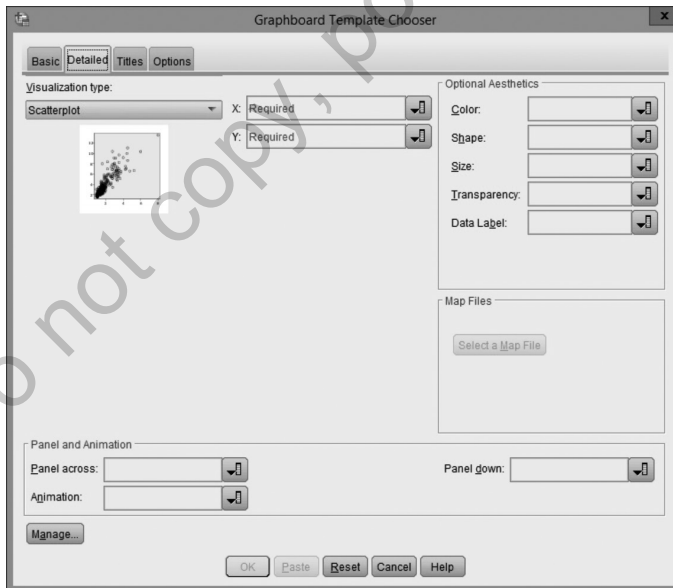
Graphboard Template Chooser Scatterplot Example

SPSS Step-by-Step. EXAMPLE 3.1b: Creating a Scatterplot Using the Graphboard Template Chooser

1. Using the same data, DC_CRIME.SAV, as in the previous example, select **Graphs/Graphboard Template Chooser**. Click on the Detailed tab at the top left of the dialog, and select Scatterplot from the pull-down "Visualization
2. In the X (required) option box, select *Year*, and in the Y (required) option box, select *Assault*.
3. If you click OK, you get the same scatterplot you previously created (Figure 3.7). To illustrate how you can select other options, change the Visualization type to Scatterplot matrix. In the variable selection box, add *Violent*, *Property*,

type" options as shown in Figure 3.8.

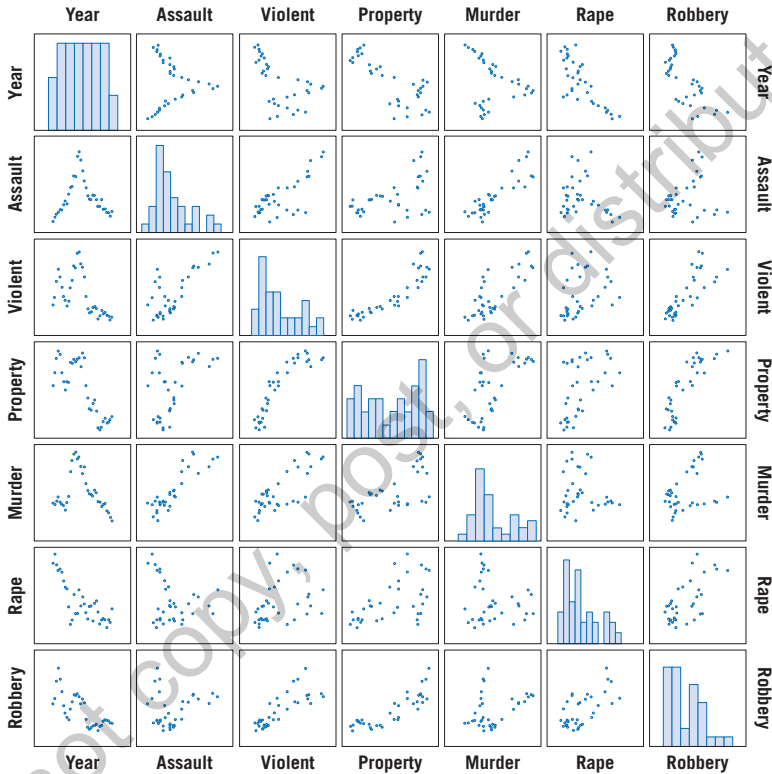
FIGURE 3.8 Select Type of Plot for Graphboard Template Chooser



Murder, Rape, and Robbery.
Click OK, and a matrix of

scatterplots appears, as shown
in Figure 3.9.

FIGURE 3.9 Scatterplot Matrix From Graphboard Template Chooser



The histograms on the diagonal for each variable allow you to visualize the distribution of that variable. There are a number of other Graphboard types you can choose in the Template Chooser, including the following: Bar, Bar of Counts, Pie, Pie of Counts, 3-D Bar, 3-D

Pie, Line, Area, 3-D Area, Path, Ribbon, Surface, Scatterplot, Bubbleplots, 3-D Scatterplot, Scatterplot Matrix, Histogram, Histogram with Normal Distribution, 3-D Histogram, 3-D Density, Dot Plot, 2-D Dot Plot, Boxplot, Heat Map, Parallel, and Maps (Chloropleths).

(Continued)

(Continued)

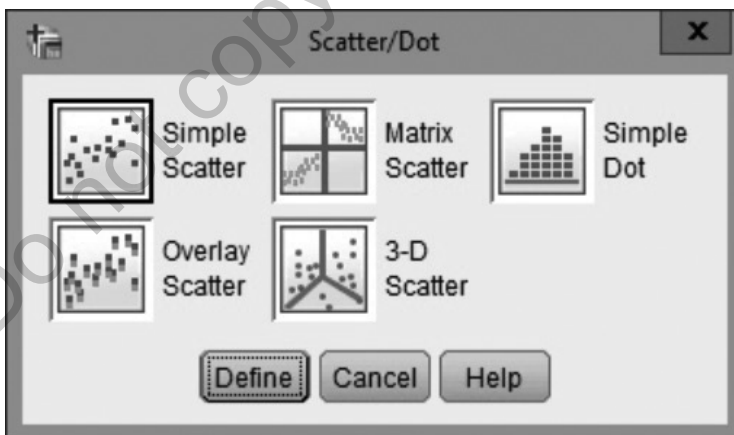
Instead of showing the same plot over and over again, in the examples here, we use different graphical methods to illustrate different plots. However, not every method can create every plot. For example, a scatterplot matrix similar to the one shown in Figure 3.9 can be designed using the Chart Builder and Legacy Dialog methods, but the option to include the histograms along the dialog is not available in either of these two methods. On the other hand, the grouped scatterplots illustrated in the next example are available in all three methods. There are too many differences in the capabilities of the three methods to mention in this chapter. If you are unable to create the plot you desire with one method, try the other methods to see if the options you want are available there.

Legacy Dialog Scatterplot Example

SPSS Step-by-Step. EXAMPLE 3.1c: Creating a Scatterplot Using Legacy Dialogs

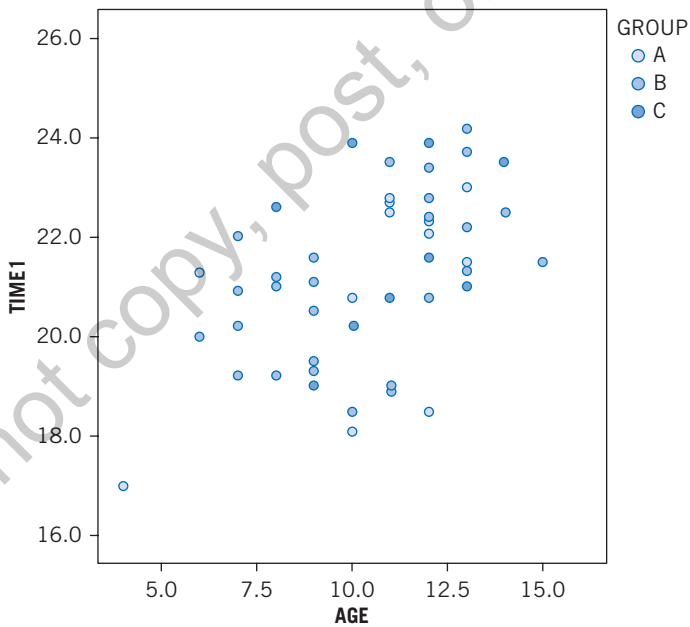
1. Open the data set named EXAMPLE.SAV, and select **Graphs/Legacy Dialogs . . .** and **Scatter/Dot**, and you will see the dialog box in Figure 3.10.

FIGURE 3.10 Select Type of Scatterplot



2. Select the Simple Scatter icon, and click Define. In the resulting dialog box, select *TIME1* as the Y-axis, *AGE* as the X-axis, and *GROUP* as the Set Markers by variable.
3. Click OK to display the graph in Figure 3.11. Notice that the dots appear in three colors keyed to the three categories for the *GROUP* variable.
4. Because it may be difficult to see the group differences by color (if your presentation/printout is in black and white as it is in this book), edit the chart to change the dots to other shapes. To do this, double-click on the scatterplot to display the Chart Editor. In the Chart Editor, double-click on the scatterplot to display the Chart Properties dialog box

FIGURE 3.11 Scatterplot With Graphed Data Using Legacy Dialogs



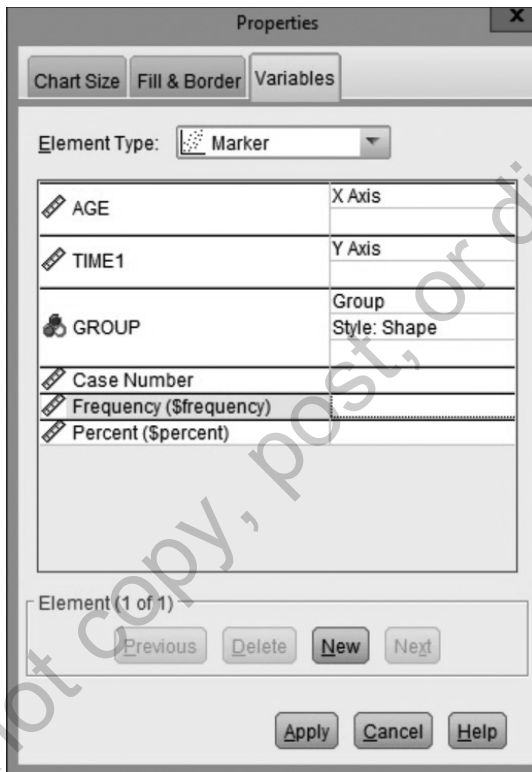
(Continued)

(Continued)

shown in Figure 3.12. In this new dialog, click on the Variables tab at the top.

- In the Variables Tab, click on Style for the Group variable, and select the Style: Shape

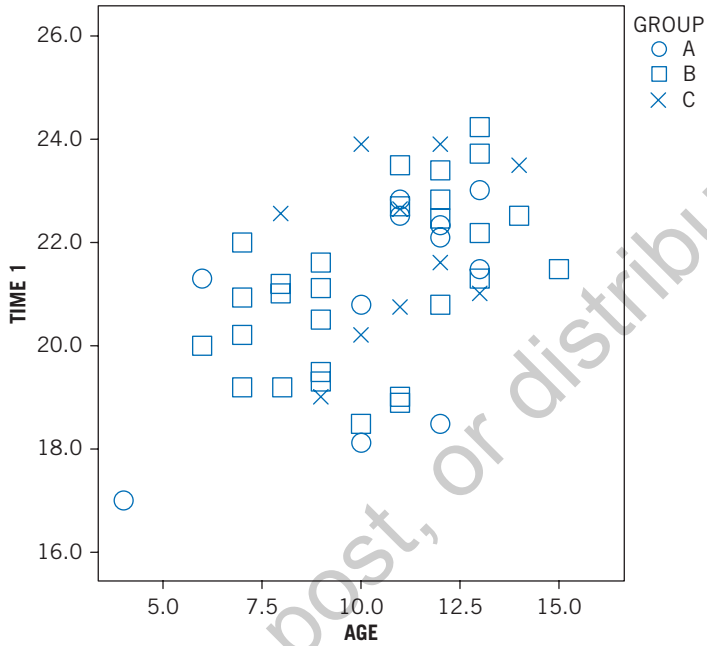
FIGURE 3.12 Variable Tab in Chart Editor



option. Click Apply. A new version of the plot with groups indicated by shaped dots appears. To make the difference more dramatic,

click on the Marker tab, and change the marker size to 15. Click on Apply, and the plot appears as shown in Figure 3.13.

FIGURE 3.13 Scatterplot With Shaped Dots



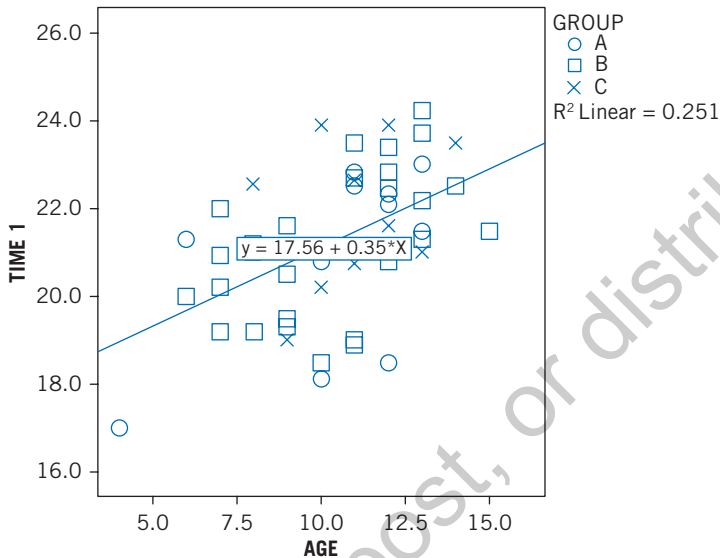
6. While still in the Chart Editor, to fit a linear regression line through the scatter of points, select **Elements/Fit Line at**

Total to display the regression line. Close the Chart Editor. The resulting graph is shown in in Figure 3.14.

In the upcoming chapter on Analysis of Variance and Covariance, an option to display separate regression lines by groups will be illustrated.

(Continued)

(Continued)

FIGURE 3.14 Scatterplot With Fitted Regression Line

HISTOGRAMS

A histogram is composed of bars that illustrate the frequency distribution of a scale (quantitative) variable across (usually) equal divisions (sometimes called bins) of the variable. The width of each bar is based on the size of the divisions, and the height of the bars represents the relative frequency of the observations within those divisions (if the bin widths are equal). A histogram is often used to visualize the distribution of a variable, and in particular, to assess the statistical normality of a variable's distribution. A histogram is also a tool for summarizing large amounts of quantitative data into a single graph.

Appropriate Applications for a Histogram

- *Examine the distribution of test scores for a standardized biology test given to all sophomores at a university.* This might be used to determine if there are natural cut-points for assignment of grades and to locate any students with particularly low grades who might need additional tutoring.
- *Visually inspect the distribution of the variable AGE in a data set to determine if the data have an approximate normal distribution.* Knowing the distribution of a variable may impact which statistical test is appropriate for analysis.
- *Display the distribution of salaries for 20,000 employees.* The histogram is a way to visualize large amounts of data in a single graph.

Design Considerations for a Histogram

The display of data as a histogram may be affected by the selection of bin size (width of divisions). Differing bin sizes can change the shape of the histogram. There are several design issues that can lead to misleading interpretations for histograms. These include the following:

- The shape of the histogram is dependent on the bin size, so it is important to choose a bin size that illustrates the distribution of the data. A good starting point is 8 to 12 bins.
- Axes should include a zero base, and there should be a scale on the vertical and horizontal axes.
- 3-D histogram charts may look more interesting, but the volume of taller bars may lead to an incorrect interpretation that they represent more information than is correct.
- All bins (class intervals) should represent the same width. (There is a programming way to create unequal bin sizes in SPSS but be aware that if bin sizes are unequal, the frequency is represented by area and not height.)

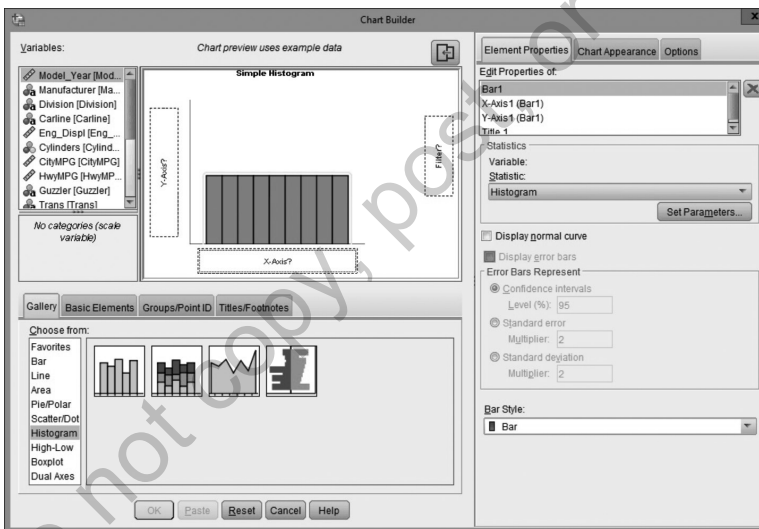
When you create a histogram, SPSS chooses bin sizes for you unless you specify a custom bin size. Information on how to select bin size is shown in the following example.

EXAMPLE 3.2

SPSS Step-by-Step. EXAMPLE 3.2: Creating a Histogram Using the Chart Builder

1. Open the data CARS2014.SAV and select **Graphs/Chart Builder**. If the warning dialog box (Figure 3.1) appears, click OK.
2. Make sure the Gallery tab is selected, and choose Histogram. Drag the Simple Histogram (leftmost icon) into the Chart Preview box as shown in Figure 3.15.
3. Drag the variable *CityMPG* from the Variables list on the

FIGURE 3.15 Drag the Simple Histogram Icon Into the Chart Preview Box

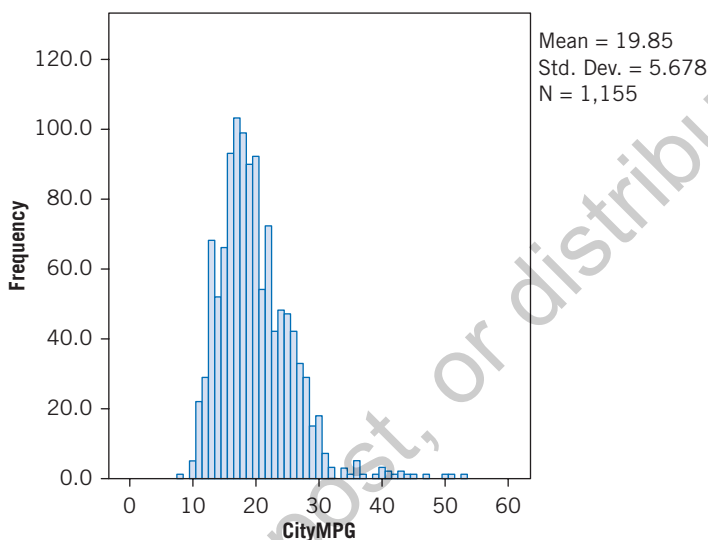


left to the X-axis box in the Chart Preview. Click OK to display the initial histogram shown in Figure 3.16.

4. To modify the graph, double-click on the displayed

histogram to enter the Chart Editor. To modify the appearance of the bars double-click on the bars and the Properties dialog box (on the right) is displayed

FIGURE 3.16 Histogram of City MPG



with several tab options. Be sure the Fill&Borders tab is selected at the top of the Properties box. As a default the Fill color is set to blue and the Border is set to black. However, you will notice that there are no black borders around the bars in Figure 3.16. This is because the Weight of the Border is set to zero by default in the Border Style section at the bottom of the Properties box. Select 0.5 in the Weight pull-down menu. You can

change the color of the bars by choosing a color from the options available. In the Pattern pull-down menu, select a crosshatch pattern. Click Apply to implement your changes.

- To display a distribution curve (while still in the Chart Editor), select **Elements/Show Distribution Curve**. In the displayed Properties dialog box, select the Normal radio button and click Close. A normal curve (based on

(Continued)

(Continued)

the mean and standard deviation for the current data) is superimposed on the histogram, as shown in Figure 3.17. Note that the histogram bars are crosshatched with black borders as previously specified in Step 4.

6. To change bin widths in a histogram, double-click on the bars in the Chart Editor to highlight them. In the displayed

Properties dialog box, choose the Binning tab. In the X-axis box, select custom, Interval width, and enter 2. Click Apply and Close. Close the Chart Editor and the bins in the histogram will reflect that change as shown in Figure 3.18.

There are many other options in the Chart Editor. We suggest that you use the Chart Editor to put a main title on the histogram and change the axis titles.

FIGURE 3.17 Histogram With Crosshatched Bars and a Superimposed Normal Curve

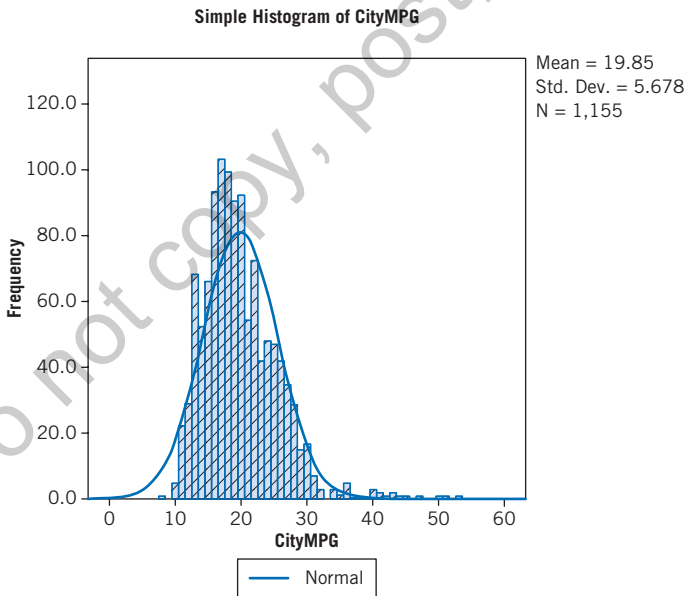
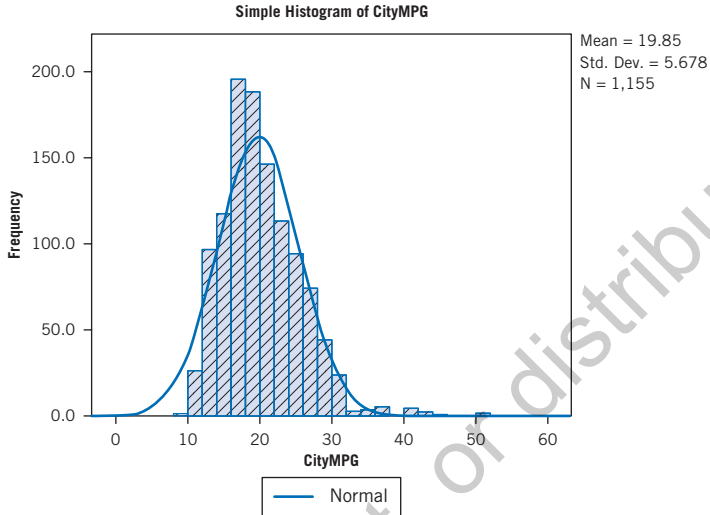


FIGURE 3.18 Histogram With Custom Bin Size

BAR CHARTS

A bar chart is not a histogram, although they are often confused. In a bar chart, each bar represents a discrete value (category), and the height of each bar represents the count for that category. The order of the bars is usually not critical. (In a histogram, bars represent the count of items within a range of ordered values.) In the bar chart, bars may also represent means by group, and in this case, may also include error bars. Examples of charts for categorical data are provided in Chapter 6: Analysis of Categorical Data. This section shows how to build a bar chart based on frequencies and another bar chart reporting means with error bars.

Appropriate Applications for a Bar Chart

- Visualize the number of M&M's in a typical bag of candy by color. Each bar represents the frequency of a particular color.

- *Visualize the proportion of five different hair colors in a population of 20,000 students.* Each bar represents the proportion ($100 * [\text{count}/\text{total}]$) for each hair color.
- *Visualize the relationship of means by group.* Each bar represents a mean value for a group.

Design Considerations for Bar Charts

There are several general design issues that should be considered when creating a bar chart (whether you use SPSS or some other method). These include the following:

- Bar charts should always have a zero base.
- Always include a scale on the vertical axis.
- 3-D bar charts may look more interesting, but the volume of taller bars may lead to an incorrect interpretation that they represent more information than is correct.
- All bars should be of the same width.
- The order of the bars is generally not important when categories represented have no particular order (such as in eye color or race). A Pareto chart arranges bars from tallest to shortest.
- Bars are typically separated with a blank area between to indicate that the data are categorical.

EXAMPLE 3.3

SPSS Step-by-Step.
EXAMPLE 3.3a: Creating a Bar Chart on Frequencies Using the Chart Builder

1. Open the data SOMEDATA.SAV, and select **Graphs/Chart Builder**. If the warning

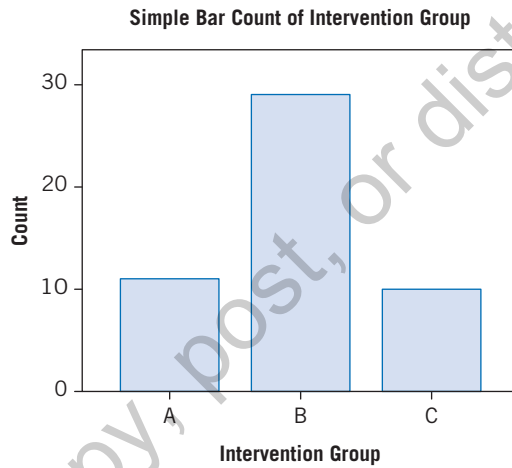
dialog box (Figure 3.1) appears, click OK.

2. Choose Bar as the chart type, and drag the Simple Bar (leftmost) icon into the Chart Preview window. This selects a single bar type.

- From the variable list, drag *Intervention Group* to the bottom axis in the Chart Preview window. Click OK, and the simple bar chart showing the counts for the three intervention groups is

shown in Figure 3.19. As in the histogram example, you could double-click on this chart, enter the Chart Editor, and change the appearance of the chart, such as bar colors, titles, and so on.

FIGURE 3.19 Initial Bar Chart

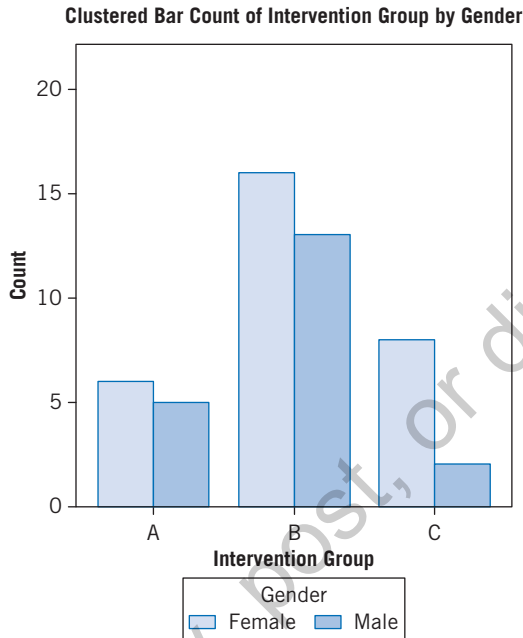


- To enhance this chart by displaying intervention group by Gender, return to the Chart Builder. Select the Cluster Bar (second to the left) icon, and drag it into the Chart Preview Window. Drag *Gender* into the “Cluster on X” specification (in the upper right of the Chart Preview). Click OK to display the graph in Figure 3.20. This

chart provides a comparison of gender for each of the three Intervention Groups showing that there are fewer male than females in each group. You could enhance this chart (for example) by double-clicking on it to open the Chart Editor and choosing **Elements/Show Data Labels** to display counts for each bar.

(Continued)

(Continued)

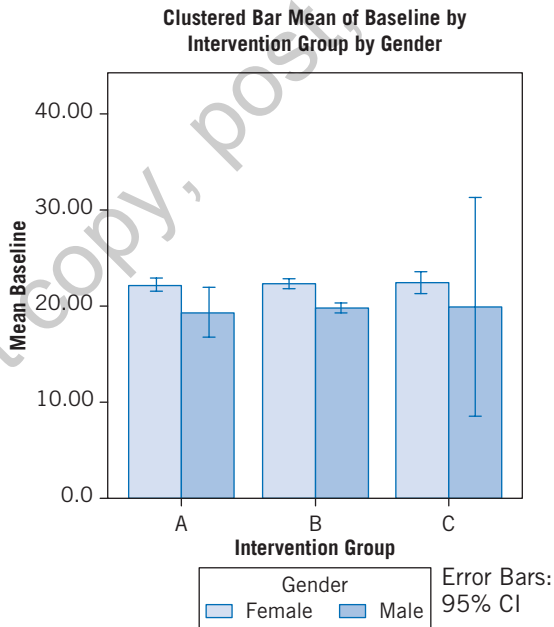
FIGURE 3.20 Clustered Bar Chart**SPSS Step-by-Step****EXAMPLE 3.3b: Creating a Bar Chart Reporting Means Using the Chart Builder**

1. Open the data SOMEDATA.SAV, and select **Graphs/Chart Builder**. If the warning dialog box (Figure 3.1) appears, click OK.
2. Choose Bar as the chart type, and drag the Clustered Bar (second from left) icon into the Chart Preview window.
3. We will use the three variables *Intervention*, *Baseline*, and *Gender*. (Note that the X-axis and Cluster on X variables must be set as nominal or ordinal SPSS Measurement types, and the Y-axis variable must be of scale type.)
 - a) Drag the *Intervention group* variable to the lower (X-axis) position.
 - b) Drag *Baseline* into the Y-axis (Count) position on

the chart. It appears as “Mean Baseline,” which indicates that the chart will display mean values for Baseline in the chart.

- c) Drag *Gender* into the “Cluster on X: Set Color” position at the top right of the Chart Preview.
4. To include error bars on the graph, click the “Display error bars” (Confidence Intervals Level % 95) in the Element Properties dialog (at the right of the viewer).
5. Click OK in the Chart Builder to create the resulting chart shown in Figure 3.21. This chart displays means (by heights of the bars) by Gender for each intervention along with error bars that indicate the amount of variability in each cluster with a 95% Confidence Interval on the mean.

FIGURE 3.21 Clusted Means Chart With Error Bars



PIE CHARTS

People have a love–hate relationship with pie charts. Some people like them for their simplicity. Others call them the worst way to display data. Some objections are that pie charts are difficult to compare to other pie charts. Most people can't tell the difference in sizes between similarly sized slices, and it is difficult to determine the number of items in each category. Technically, there are issues in the way people visually interpret areas and degrees in the chart that make it easy to manipulate its meaning (Cleveland, 1985). Nevertheless, pie charts are often used. We'll illustrate how you can create a pie chart using Chart Builder. Slices in pie charts are used to illustrate a proportion of a whole. Many times, the data represented in a bar chart can also be displayed as a pie chart.

Appropriate Applications for a Pie Chart

- *Visualize the proportion of five different hair colors in a population of 20,000 students.* Each pie slice represents the proportion ($100 * [\text{count} / \text{total}]$) of each hair color.
- *Visualize the parts of a budget.* Each pie slice represents a proportion (a slice) of an entire budget.

Design Considerations for Pie Charts

There are several design issues that can lead to misleading interpretations for pie charts. These include the following:

- Any time two pie charts are compared, the circle should be the same diameter; otherwise the larger circle appears to carry more weight than a smaller one.
- When pie charts have many slices, it becomes increasingly difficult to compare the sizes of each slice. Pie charts work best when there is one large slice among several and the chart illustrates which category is dominant.
- 3-D pie charts distort wedges making some appear more important (or less important) than they are.
- By selecting specific colors for slices and designing where a slice is located in the chart, it is possible to manipulate the visual perception of the chart's message (Cleveland, 1985).

EXAMPLE 3.4

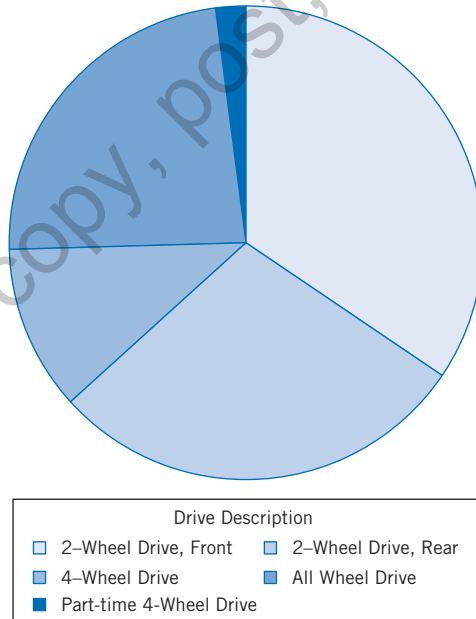
SPSS Step-by-Step. EXAMPLE 3.4: Creating a Pie Chart Using the Chart Builder

1. Open the data CARS2014.SAV and select **Graphs/Chart Builder**. If the warning dialog box (Figure 3.1) appears, click OK.
2. Select the **Pie/Polar** type graph from the options, and drag the icon into the Chart

Preview Box. Select the *Drive Description* variable from the list, and drag it to the “Slice by” box at the bottom of the sample pie chart. Click OK, and the pie chart shown in Figure 3.22 appears.

This chart indicates the proportions of drive types for various vehicles in the CARS2014 data set showing that 2-Wheel Drive, Front is the most

FIGURE 3.22 Pie Chart of Drive Description



(Continued)

[Continued]

common type. You can double-click on the chart to enhance the chart using the Chart Editor. For example, experiment with putting the number of observations for each type in the pie slices by choosing **Elements/Show Data Labels**, and (in the Properties

dialog) put the Count option in the Displayed list and remove Percent. Also, double-click on the green pie slice to select it, and choose **Elements/Explode Slice** to separate that slice from the others for the purpose of highlighting that group.

BOXPLOTS

Boxplots (also called box-and-whiskers plots) are a representation of the quartiles of a scale variable with 50% of the data values inside the middle box and 25% of the data (represented by lines, or whiskers) on each side of the box. In general, a boxplot shows the 0th, 25th, 50th, 75th, and 100th percentile of the data (sometimes called the Tukey five-number summary, named after its originator John Tukey). A horizontal line inside the box represents the median.

Appropriate Applications for Boxplots

- *Visualize the distribution of the data.* Assess the approximate normality of data by looking for a symmetric box with a median centered within the box, a box whose length (called the interquartile range, IRQ) is about 25% of the total length (range), and with equal-length whiskers.
- *Locate outliers in a set of data.* Use boxplots to identify data points that are possible outliers. In SPSS, an outlier is defined as a point beyond $1.5 \times \text{IRQ}$ from the top or bottom of the box, and an extreme outlier is $3 \times \text{IRQ}$ beyond the box.

Design Considerations for Boxplots

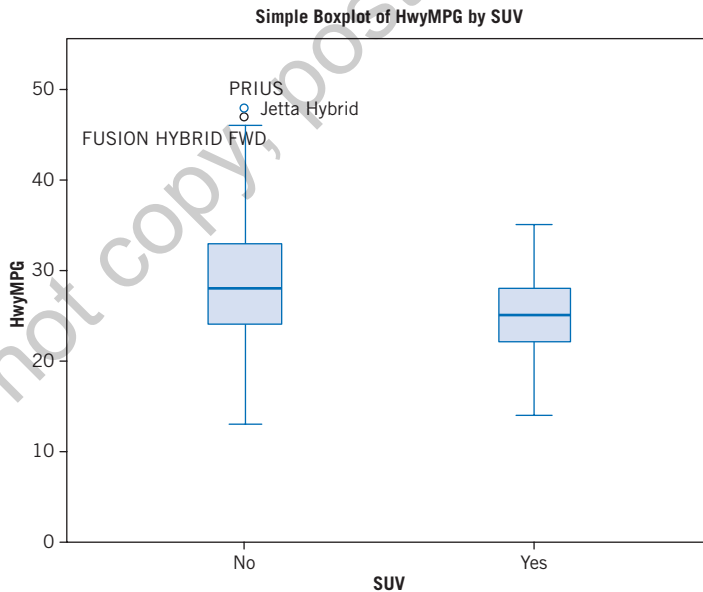
Boxplots are a simple way to view a variable's distribution, central tendency, and, when displayed by group, allow you to make a visual comparison across groups.

EXAMPLE 3.5

SPSS Step-by-Step. EXAMPLE 3.5: Creating Comparative Boxplots Using the Chart Builder

1. Open the data CARS2014.SAV, and select **Graphs/Chart Builder**. If the warning dialog box (Figure 3.1) appears, click OK.
2. Select the boxplot icon from the graph types. Choose the leftmost (Simple Boxplot) icon, and drag it into the Chart Preview box.
3. Select (drag) *SUV* to the X-axis and *HwyMPG* into the Y-axis.
4. Select the **Groups/Point ID** tab just below the Chart Preview box, and check the Point ID label checkbox. A new variable box labeled Point Label Variable appears in the Chart Preview. Drag the variable *Carline* into that box.
5. Click OK to display the comparative boxplots shown in Figure 3.23.

FIGURE 3.23 Comparative Boxplots Using Cars2014 Data Set



(Continued)

(Continued)

These comparative boxplots provide visual evidence that (although there is a lot of overlap) SUVs (SUV = Yes on the plot) tend to get lower highway miles per gallon than non-SUVs (but not always). The

boxplots also indicate that several non-SUVs have unusually high miles per gallon. When you examine the labels for these outliers, you see that they are all hybrid vehicles.

Summary

A number of additional charts and graphs are created throughout the text using the basic three graphing methods illustrated in this chapter:

- Chart Builder
- Graphboard Template Chooser
- Legacy Dialogs

Most charts created by any of these techniques can be edited using the SPSS Chart Editor. All three techniques were illustrated in this chapter, but most examples used in this chapter and throughout the book use the Chart Builder, which (we believe) has the most straightforward design mechanism of the three.

References

Cleveland, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Tufte, E. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.