

3

STACKED BAR GRAPH

PURPOSE OF GRAPHS YOU ARE ABOUT TO BUILD

- To explore for a relationship between the categories of two discrete variables

3.1 INTRODUCTION TO THE STACKED BAR GRAPH ★

As with the simple bar graph, the stacked bar graph uses rectangular boxes to represent categories of a variable. The variable located on the x -axis is known as the *stacked* variable. The stacked bar graph differs from the simple bar graph in that each rectangular box on the x -axis is made up of smaller individual boxes that SPSS calls *segments*. The *segments* are stacked on top of one another and could also be called the y -axis variable. The height of each segment represents the percentage of cases in a category of the *stacked* variable and a category of the *segment* variable. The stacked bar graph depicts the relationship between the categories of the *stacked* and *segment* variables. The stacked bar graph provides a convenient way to discover and then visualize relationships between two discrete variables. Another use of the stacked bar is when you wish to describe the relationship between two discrete and a single continuous variable. We have elected not to demonstrate this particular graph, as you can provide the same description by building a clustered bar graph as shown in the next chapter.

The reader should also be aware that the stacked bar graph can be used to mislead someone using a graph to understand data. Some readers may wish to view an example of a deceptive stacked bar graph at this time. An example may be found in Chapter 18 on deceptive uses of graphs, see Figure 18.3.

The appearance of the stacked bar can vary greatly depending on how the percentages are calculated. In the next two examples, we use the same data to demonstrate two ways to calculate percentages. The first is where you instruct SPSS to calculate percentages with each stack representing 100% of the observations within that stacked category. With this approach, certain questions can be answered, which are soon discussed. The second approach is to have SPSS calculate the percentage of the total across the segments, which results in answers to another set of questions. We next take a closer look at these two approaches.

3.1.1 Stacked Bar Graph (Percentage of Stack's Total)

To build the stacked bar, as shown in Figure 3.1, you must instruct SPSS to calculate percentages so that each *stack* represents 100% of the observations in that stack. You will soon learn how to accomplish this when building the stacked bar.

The stacked bar graph in Figure 3.1 summarizes data for students in statistics classes collected over an 11-year period. The intention of this current analysis is to compare gender and earned grade for 1,050 students in these statistics classes.¹ The *stacked* variable in Figure 3.1 is *gender*, having two categories—male and female. The *segment* variable is the *grade* earned and has five categories—A, B, C, D, and F.

In this example, the major question we are trying to answer is whether there is a difference in the performance for the males and females. For instance, is the percentage of A's the same for males and females taking the class? This question may be repeated for each of the four remaining grades. We may also examine how the grades are distributed within each category of the stacked variable, in this case *gender*. Are the distributions of grades for the males and females similar?

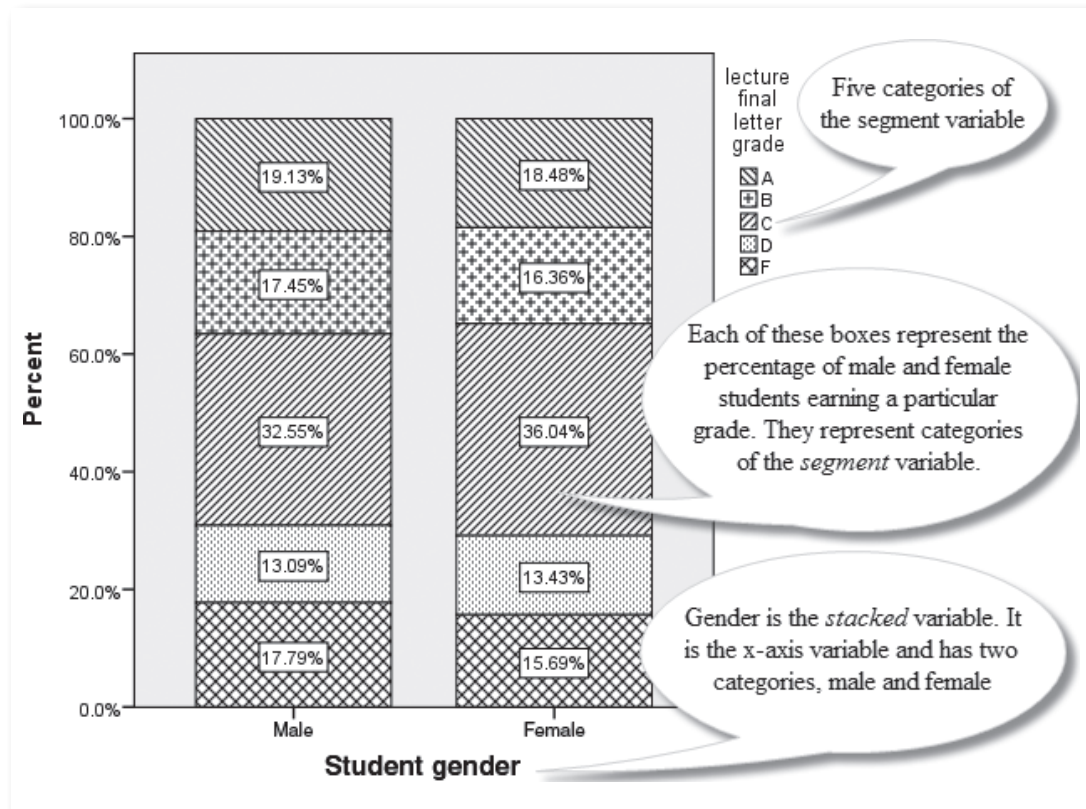
The stacked bar is especially useful when seeking to discover relationships between two variables. In our example, the variable gender was chosen for the *x*-axis, having the two categories of male and female. Each segment within the rectangular boxes for male and female represents the percentage of students earning a specific grade (second variable) *within* their gender category. For an example, look at Figure 3.1 and notice that

¹These data were obtained from records of 1,050 students taking an intermediate level social statistics class with the authors.

19.13% of males earned A, while 18.48% of females earned this same grade—very close to identical.

The stacked bar graph shown in Figure 3.1 answers the major question that sought to determine if grade earned was related to gender. Do males and females differ in the proportion of the various grades that were earned? Stated another way, we could seek to discover if the distributions of grades for males and females were approximately the same. One quick glance at the graph shows that there is very little gender difference in the proportion of grades earned, and hence, the shapes of the distributions are approximately identical.² In the following section, the database used to build a stacked bar graph and the questions we seek to answer are presented.

Figure 3.1 Stacked Bar Graph (Percentage of Stack's Total) for Gender and Grade



²Although there appears to be little difference in values, it would be appropriate to test for significance using chi-square analysis to confirm your visual interpretation.

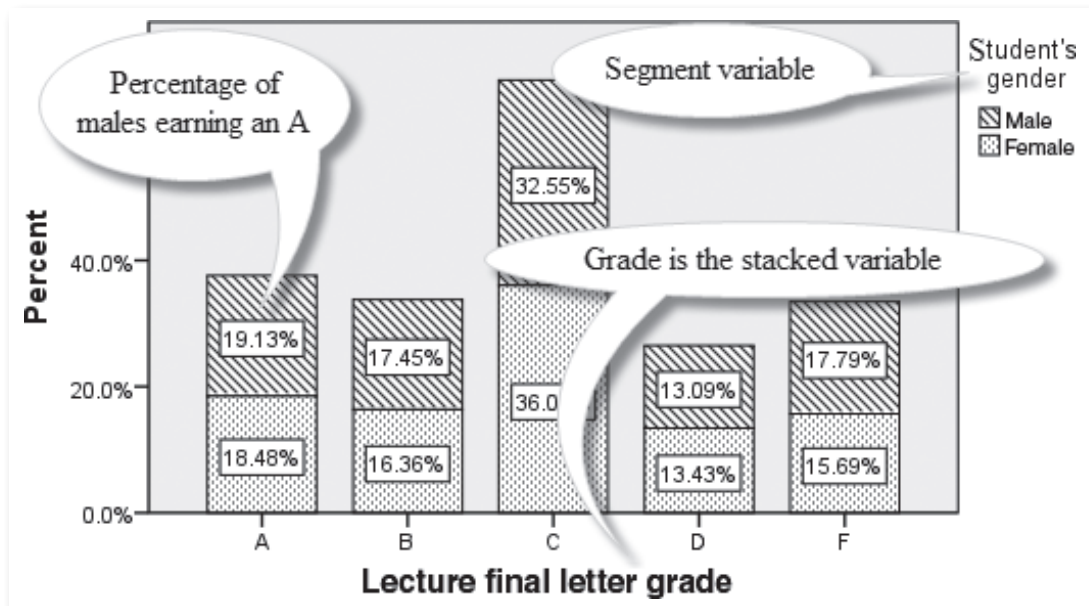
3.1.2 Stacked Bar Graph (Percentage of Segment's Total)

To build the stacked bar shown in Figure 3.2, you must instruct SPSS to calculate percentages so that each *segment* represents 100% of the observations. You will soon use a few clicks of the mouse button to accomplish this calculation.

Figure 3.2 uses the identical data used in Figure 3.1 except that the *x-axis* or *stacked* variable is now *grade* and the *segment* variable is *gender*. The percentages shown in each bar now represent the percentage of the total observations in all segments. Go ahead; add the segment percentages' together to confirm this fact.

In this example, it is acceptable to directly look at each bar (grade) and determine the correct percentage of males and females earning that grade. As in the previous graph, 19.13% represents the percentage of males who earned the grade A. The first segment of the "A" stack shows 18.48%, which is the percent of all female students earning that grade. By reading each segment horizontally, we can see the percentage of males earning the various grades. The same can be done for the females.

Figure 3.2 Stacked Bar Graph (Percentage of Segment's Total) for Gender and Grade



There are many variations in how to place your variables (*x-axis* or *legend*) as well as ways in which to calculate percentages (total, total of segment, or total of stacked variable). Each of the variations will produce unique graphs that may be tailored to answer different sets of questions about the data.

The reader will soon have the knowledge to experiment with the various possibilities of this versatile graph.

3.2 DATABASE AND QUESTIONS ★

The data used to build a stacked bar graph is found in the SPSS Sample files and is named **1991 U.S. General Social Survey.sav**.³ You should open this database now. If you need to review the opening procedure, refer to Section 2.2 of Chapter 2.

The General Social Survey (GSS) has been conducted by the National Opinion Research Center every year since 1972. There are 43 variables and 1,517 cases in the current SPSS version of a small portion of this database. We use two discrete variables to produce a stacked bar graph. One variable is named *race* and labeled *Race of the respondent*, which will be the *x-axis* variable (*stacked* variable). The *segment* variable is named *region* and labeled *Region of the United States*.

3.2.1 Questions for the Stacked Bar Graph (Percentage of Stack's Total)

1. How are the three races (White, Black, and Other) distributed between the three regions of North East, South East, and West.
2. What region do those in the White race category prefer?
3. What are the first, second, and third choices of those in the Black category?

³Some readers may feel that this database is just a little outdated. However, there are a couple of reasons we have chosen to use this database dating back to the 1991 census. First of all, it was included with SPSS and contained useful data that was measured at the levels of measurement needed for our graph-making demonstrations. On a more serious note, one should remember that to measure any change in society, one must have a starting point. Thus, the data presented in 1991 are extremely important in many situations where the amount of change is important.

3.2.2 Questions for the Stacked Bar Graph (Percentage of Segment's Total)

1. Which region had the highest percentage of those categorized as White?
2. Rank the regions, in order from the least chosen to the most chosen by those respondents who were classified as “Other.”
3. Which region is the least desirable for the White category?

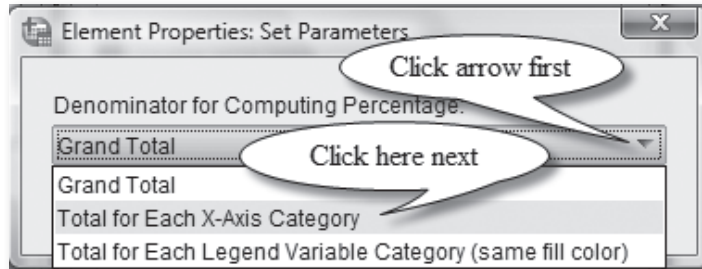
In the next section, you produce stacked bar graphs to answer these questions about the GSS database.

★ 3.3 USING SPSS TO BUILD THE STACKED BAR GRAPH

In this section, you build a stacked bar graph that requires the use of the *Chart Builder*, the *Element Properties* feature, and finally the *Chart Editor* to embellish the graph. The finished stacked bar graph will then answer the questions concerning a possible relationship between race and region of residence.

3.3.1 Building the Stacked Bar Graph (Percentage of Stack's Total)

- Open **1991 U.S. General Social Survey.sav** (found in the SPSS Sample file).
- Click **Graphs**, and then click **Chart Builder**. When the *Chart Builder* window opens, make certain that **Gallery** is selected—if not, then click **Gallery**.
- Click **Bar** (found in the *Choose from:* list in the lower left portion of the window), and drag the third **graph icon** (*Stacked Bar*) to the large panel on the top right of the *Chart Builder* window.
- Click **Race of Respondent** in the *Variables* panel, and drag it to the *X-Axis* box.
- Click **Region of the United States**, and drag it to *stack: set color* box.
- If *Elements Properties* window is not open, then click **Element Properties button** to open.
- In the *Statistics* panel, click the **down arrow** and click **percentage (?)**.
- Click **Set Parameters** (*Element Properties: Set Parameters* window opens) (see Figure 3.3).

Figure 3.3 *Element Properties: Set Parameters Window*

- Click the **black arrow** to the right of *Grand Total* and click **Total for Each X-Axis Category** (doing this calculates the percentage of each category of the *segment* variable within the *stacked* variable category).
- Click **Continue**.
- Click **Apply**, then Click **OK**.

Once **OK** is clicked, a basic draft of the stacked bar graph appears in the *SPSS Output Viewer*. The following procedures will enhance this graph and make it easier to answer our questions.

- Double click the **graph** in the *Output Viewer* (*Chart Editor* window opens).
- Click the **Data Label Mode icon** as shown in Figure 3.4, and then click on each **segment**, which inserts a percentage amount for each category within the *stacked* variable.
- Click the **Data Label Mode icon**, which turns it off.

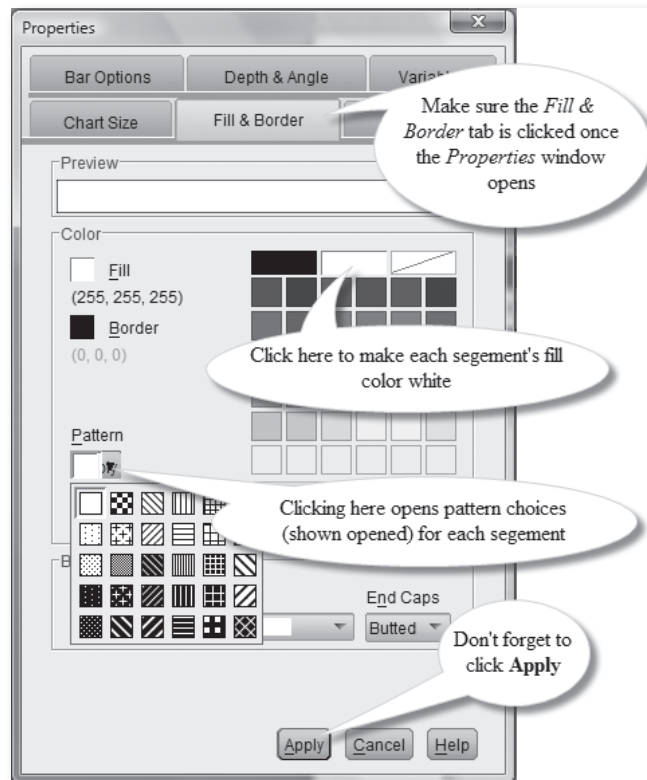
Figure 3.4 *Chart Editor Window and Show Properties Window and Data Label Mode Icons*

- Click the **rectangular box** for the **White** category (a faint yellow border surrounds all categories for the *x-axis* variable).
- Click the **upper segment** of any of the stacked categories (a faint yellow border appears on all the upper segments of your *stacked* variable categories).
- If *Element Properties* window is not open, you must click the **Show Properties Window icon** in the *Chart Editor* menu as shown in Figure 3.4 (once clicked the *Properties* window opens as shown in Figure 3.4).

We next utilize the *Properties* window (shown in Figure 3.5) to make some improvements to the stacked bar graph. These improvements will enhance the question answering ability of the graph.

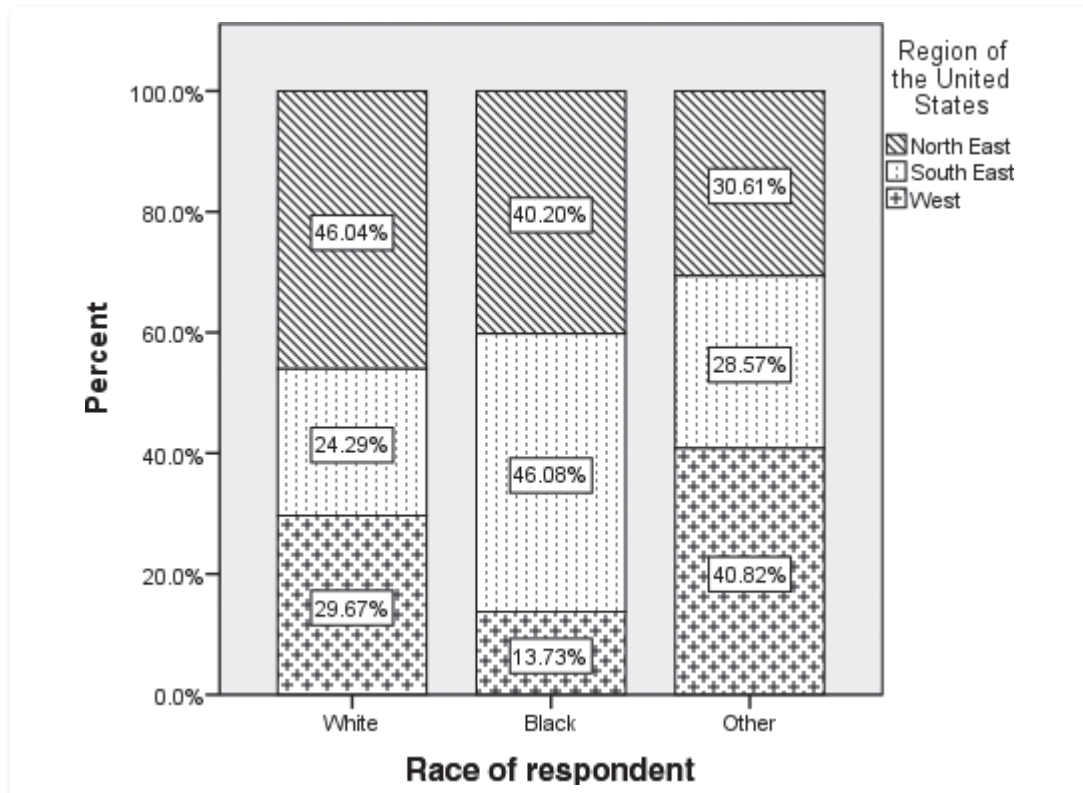
- In the *Color* panel, click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (Figure 3.5).
- Click the **black arrow** beneath *Pattern* in the *Properties* window, then click the **third pattern** in the **top** row.

Figure 3.5 *Properties* Window as Used to Change Fill and Border



- Click **Apply** (next we repeat this pattern selection process for each of the segments).
- Click the **middle segment** (all these segments are now framed with a yellow border).
- In the *Color* panel, click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (Figure 3.5).
- Click the **black arrow** beneath *Pattern*, and click the **first pattern** in the **second** row.
- Click **Apply**.
- Click the **lowest segment** (all these final segments are framed in yellow).
- In the *Color* panel click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (Figure 3.5).
- Click the **black arrow** beneath *Pattern*, and click the **second pattern** in the **second** row.
- Click **Apply**.
- Click the **X** in the box in the top-right portion of the *Chart Editor* (the finished graph now appears in the *Output Viewer* as shown in Figure 3.6).

Figure 3.6 Stacked Bar Graph (Percentage of Stack's Total) for Race and Region

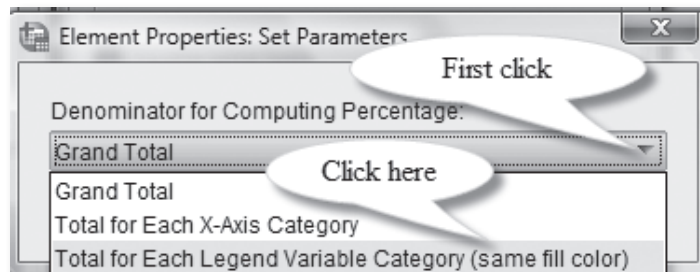


The stacked bar graph in Figure 3.6 depicts the relationship of race of respondents and the region in which they reside.

3.3.2 Building the Stacked Bar Graph (Percentage of Segment's Total)

- Open **1991 U.S. General Social Survey.sav** (found in the SPSS Sample file).
- Click **Graphs**, and then click **Chart Builder**. When the *Chart Builder* window opens, make certain that **Gallery** is selected—if not, then click **Gallery**.
- Click **Reset button** (clears previous work from *Chart preview* window).
- Click **Bar** (found in the *Choose from:* list in the lower left portion of the window), and drag the third **graph icon** (*Stacked Bar*) to the large panel on the top right of the *Chart Builder* window.
- Click **Race of respondent** in the *Variables* panel, and drag it to the *stack: set color* box.
- Click **Region of the United States**, and drag it to *XAxis* box.
- If *Elements Properties* window is not open, then click **Element Properties button** to open the window.
- In the *Statistics* panel, click the **down arrow** and click **percentage (?)**.
- Click **Set Parameters** (*Element Properties: Set Parameters* window opens) (see Figure 3.7 for this window).
- Click the **down arrow** to the right of where it says *Grand Total*, and click **Total for Each Legend Variable Category (same fill color)** (doing this calculates the percentage of each category of the *segment* variable within the *stacked* variable category)

Figure 3.7 *Element Properties: Set Parameters* Window



- Click **Continue**.
- Click **Apply**, then Click **OK**.
- Double click the **graph** in the *Output Viewer* (*Chart Editor* window opens).
- Click the **Show Data Labels icon** as shown in Figure 3.8 (data labels, as percentages, are added to all segments).
- Click the **upper segment** of any of the stacked categories (a faint yellow border appears on all bars for the *stacked* variable categories), click the **same segment** a second time, and only the horizontal segment remains framed.
- If *Element Properties* window is not open, you must click the **Show Properties Window icon** in the *Chart Editor* menu as shown in Figure 3.8.

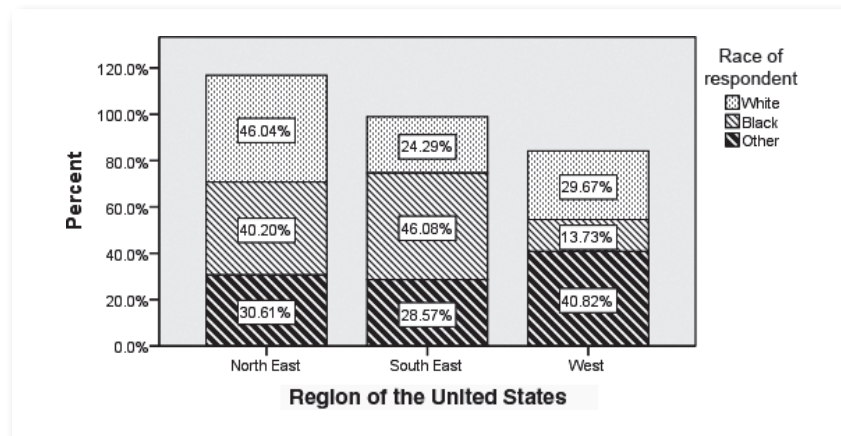
Figure 3.8 *Chart Editor* (Upper Portion) Indicating the *Show Data Labels* icon



- In the *Properties* window, click the **Fill & Border tab** (once clicked, the *Properties* window should appear as shown in Figure 3.5).
- In the *Color* panel, click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (see Figure 3.5 and be sure to avoid clicking the white rectangular box with the **red** line through it).
- Click the **black arrow** beneath *Pattern* in the *Properties* window, then click the **first pattern** in the **third** row.
- Click **Apply** (we next repeat this pattern selection process for each of the segments).
- Click the **middle segment** (all these segments are now framed with a faint border).
- In the *Color* panel, click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (Figure 3.5).
- Click the **black arrow** beneath *Pattern*, and click the **third pattern** in the **first** row.

- Click **Apply**.
- Click the **lowest segment** (all these final segments are framed in yellow).
- In the *Color* panel, click the **small box** to the left of *Fill*, then click the **white rectangular box** to the right (Figure 3.5).
- Click the **black arrow** beneath *Pattern*, then click the **second pattern** in the **fifth** row.
- Click **Apply**.
- Click the **X** in the box in the top-right portion of the *Chart Editor* (the finished graph now appears in the *Output Viewer* as shown in Figure 3.9).

Figure 3.9 Stacked Bar Graph (Percentage of Segment's Total) for Region and Race



★ 3.4 INTERPRETATION OF THE STACKED BAR GRAPH

Let's examine the graphs in Figures 3.6 and 3.9 to see if they answer our questions about the data. Remember that the graph in Figure 3.6 shows the segment variable as the region of residence, while the stacked variable is race. For the graph in Figure 3.9, the segment variable is race and the stacked variable is the region of residence.

One of the first things to note is that the appearance of both graphs differs from the examples we presented in Figures 3.1 and 3.2, when gender and grade were compared. In the "gender and grade" graphs, the segments

are approximately the same size sending the message that males and females performed equally. The graphs just produced for “race and region” show segments that are not equal. The message here is that those in the three race categories choose different regions of the United States in which to reside. Let’s examine this more closely by revisiting our original questions.

3.4.1 Questions and Answers for the Stacked Bar (Percentage of Stack’s Total)

The answers to the following questions are provided by the graph in Figure 3.6.

1. How are the three races (White, Black, and Other) distributed between the three regions of North East, South East, and West?

The graph shown in Figure 3.6 clearly indicates that the race categories are not evenly distributed between the regions of residence. It is the proportional sizes of the segments in the various stacks that tell us that the race categories are *not* evenly distributed. It is also informative to note the percentages of the race categories (Whites, Blacks, and Other) choosing one of three regions (North East, South East, and West) to reside.⁴

2. What region do those in the White race category prefer?

Forty-six percent of those in the White category choose the North East as a place to live.

3. What are the first, second, and third choices for those in the Black category?

The first, second, and third choices for those in the Black category are shown in the Black stack. Those in the Black category have selected their Number 1 region of choice as the South East (46.08%), the second choice was the North East (40.20%), and the third was the West (13.73%).

The stacked bar graph is a useful tool when attempting to describe the relationship between two variables. In the following chapter, we discuss another form of bar graph—clustered bar.

3.4.2 Questions and Answers for the Stacked Bar (Percentage of Segment’s Total)

The answers to the following questions are provided by the graph in Figure 3.9.

⁴You could at this point, under certain conditions, conduct a statistical test and determine the likelihood that the observed differences are statistically significant.

1. Which region had the highest percentage of those categorized as White?

The highest percentage recorded for whites was the North East at 46.04%.

2. Rank the regions, in order from the least chosen to the most chosen by those respondents who were classified as “Other.”

The smallest percentage of the “Other” category chose the South East (28.57%), next was the North East at 30.61%, and the most preferred choice of residency was the West at 40.82%.

3. Which region is the least desirable for the White category?

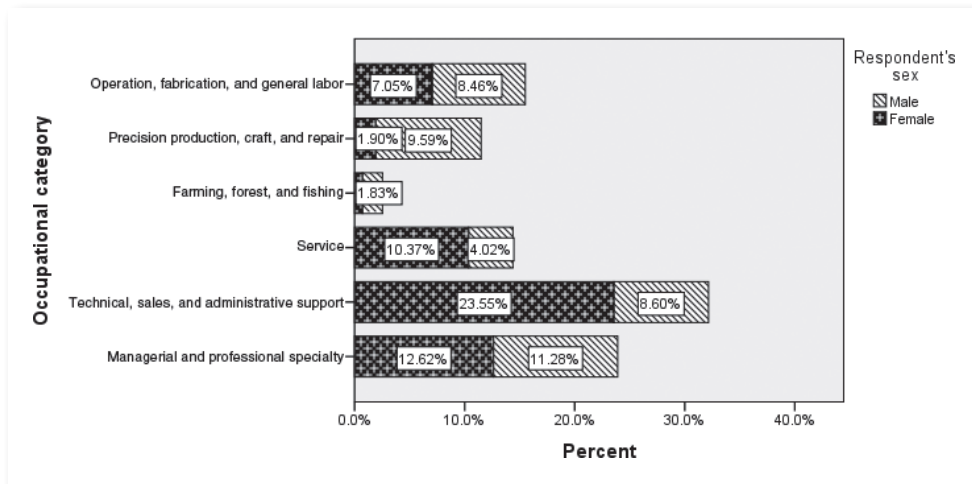
The respondents of the South East region represent 24.29%, making it the least desirable choice of residency by those categorized as White.

★ 3.5 REVIEW EXERCISES

1. Open the SPSS Sample file **1991 U.S. General Social Survey.sav**. Select the two discrete variables of *Sex* (Variable 1) and *occcat80* (Variable 14) to construct a stacked bar graph as in Figure 3.10. Place the *occcat80* variable on the *x*-axis, change the *y*-axis to percent, and make the *segment (color)* variable the *Respondent's sex*. Use a horizontal orientation as one way to accommodate the long category names.

Question: What are the proportions of males and females in each of the six occupational categories?

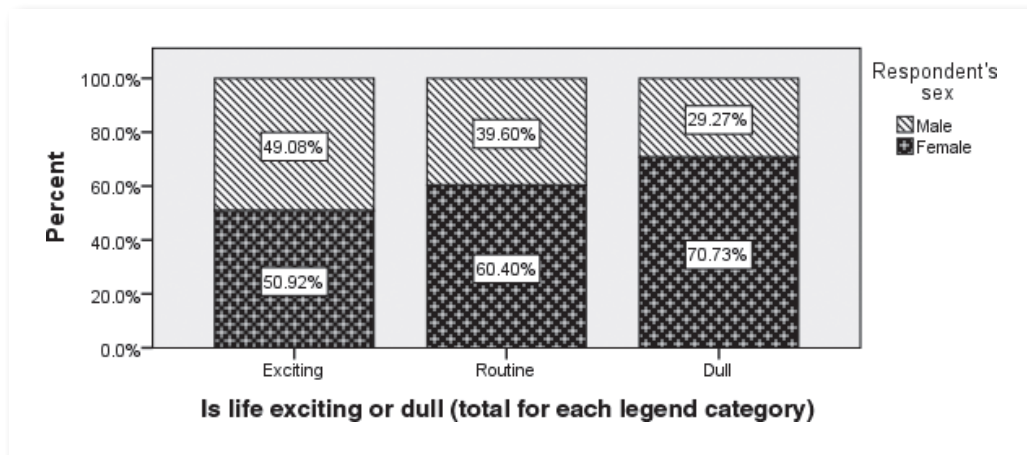
Figure 3.10 Review Exercise: Stacked Bar Graph for Occupation and Sex



2. Open the SPSS Sample file **1991 U.S. General Social Survey.sav**. Select the two discrete variables of *Sex* (Variable 1) and *life* (Variable 5) to construct a stacked bar graph as in Figure 3.11. Place the *life* variable on the *x*-axis, change the *y*-axis to percent, and make the *segment (color)* variable the *Respondent's sex*.

Questions: You want to explore the database seeking to discover any relationship between gender and outlook on life. (a) What are the gender differences for each category? (b) Which of the three life outlook categories has the greatest gender difference? (c) Which has the least difference?

Figure 3.11 Review Exercise: Stacked Bar Graph for Sex and Outlook on Life

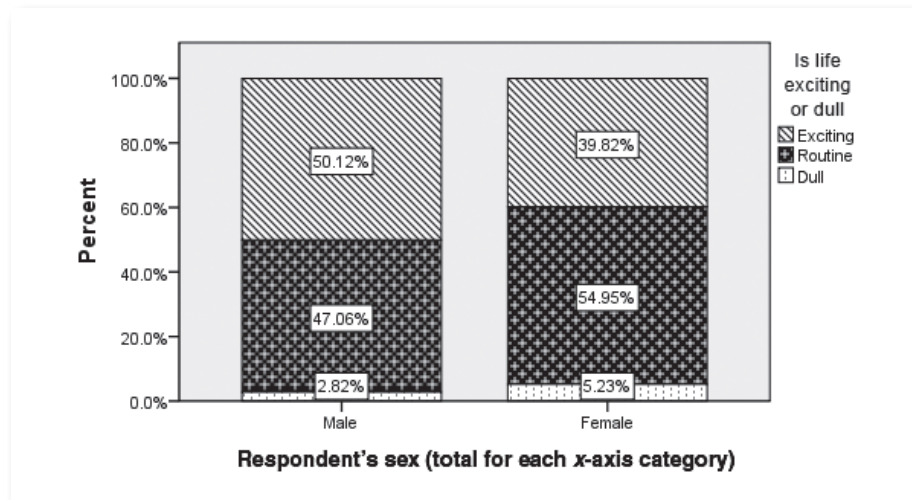


3. Open SPSS Sample file **1991 U.S. General Social Survey.sav**. Select the two discrete variables of *Sex* (labeled *Respondent's sex*) and *life* (labeled *Is life exciting or dull*) to construct a stacked bar graph as in Figure 3.12. Place the *Sex* variable on the *x*-axis, the *segment (color)* variable is *life*, and change the *y*-axis to percent.

Questions: You want to explore the database seeking to discover any relationship between gender and outlook on life. (a) What percentage of males think that life is Exciting? (b) What percentage of females think that

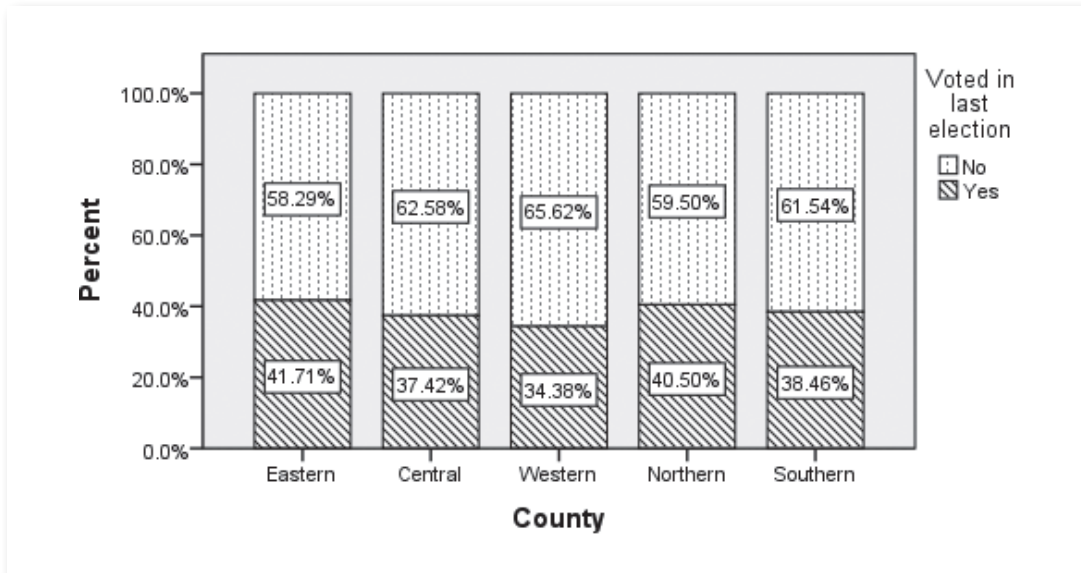
life is Exciting? (c) What percentage of males think that life is Routine? (d) What percentage of females think that life is Routine (e) What percentage of males think that life is Dull? (f) What percentage of females think that life is Dull? (g) Do males and females have the same outlook on life in this study?

Figure 3.12 Review Exercise: Stacked Bar Graph for Sex and Outlook on Life



4. Open the SPSS Sample file [poll_cs_sample.sav](#) to build a stacked bar graph as in Figure 3.13. Select the discrete variable named *county* and labeled *County* (Variable 4) and place it on the *x*-axis. Select the discrete variable named *votelast* and labeled *Voted in last election* (Variable 7) and place it on the *stack: set color* box. Change the *y*-axis to percent.

Questions: (a) What percentage of individuals who live in the *Eastern* county voted in the last election? (b) What percentage of individuals who live in the *Northern* county did not vote in the last election? (c) What percentage of individuals who live in the *Central* county voted in the last election? (d) What percentage of individuals who live in the *Southern* county did not vote in the last election? (e) What percentage of individuals who live in the *Western* county voted in the last elections? (f) Which county region has the largest percentage of individuals who voted in the last election?

Figure 3.13 Review Exercise: Stacked Bar Graph for County Region and Voting Behavior

5. Open the SPSS Sample file [poll_cs_sample.sav](#) and build a stacked bar graph as in Figure 3.14. Select the discrete variable named *opinion_gastax* and labeled *The legislature should enact a gas tax* (Variable 9) and place it on the *x-axis*. Select the discrete variable named *agecat* and labeled *Age category* (Variable 5) and place it on the *stack: set color* box. Change the *y-axis* to percent.

Questions: (a) What percentage of individuals who strongly disagree that there should be a gas tax are older than 60? (b) What percentage of individuals who agree that there should be a new gas tax are aged 18 to 30? (c) What percentage of individuals who strongly agree that there should be a new gas tax are aged 46 to 60? (d) What percentage of individuals who strongly disagree that there should be a new gas tax are aged 18 to 30 years? (e) Which of the age categories, within the strongly agree that there should be a new tax group, contains the highest percentage of individuals? (f) Which of the age categories have the lowest percentage of individuals who disagree that the legislature should enact a gas tax?

Figure 3.14 Review Exercise: Stacked Bar Graph for Gas Tax and Age Category

