
SPSS Step-by-Step

Tutorial: Part 1

© DataStep Development 2004

Table of Contents

1 **SPSS Step-by-Step 5**

Introduction 5

Installing the Data 6

Installing files from the Internet 6

Installing files from the diskette 6

Introducing the interface 6

The data view 7

The variable view 7

The output view 7

The draft view 10

The syntax view 10

What the heck is a crosstab? 12

2 **Entering and modifying data 13**

Creating the data definitions: the variable view 13

Variable types 13

<i>Variable names and labels</i>	15
<i>Missing values</i>	15
<i>Non-numeric numbers, or when is a number not a number?</i>	15
<i>Binary variables</i>	15
Creating a new data set	16
Getting help in creating data sets and defining variables	22
Creating primary reference lists	24
<i>Frequencies</i>	24
<i>Descriptive statistics: descriptives (univariate)</i>	25
Recodes and Transformations	26
<i>Backup the original file</i>	26
<i>Recoding existing variables</i>	27
Recode income data	27
Recoding variables revisited	37
<i>The one exception in recoding variables</i>	37
<i>The other exception</i>	37

3

Charting your data 39

Using the automated chart function	40
<i>Using the Interactive Chart function</i>	42
<i>Creating a chart from scratch</i>	45

1

SPSS Step-by-Step

Introduction

SPSS (Statistical Package for the Social Sciences) has now been in development for more than thirty years. Originally developed as a programming language for conducting statistical analysis, it has grown into a complex and powerful application with now uses both a graphical and a syntactical interface and provides dozens of functions for managing, analyzing, and presenting data. Its statistical capabilities alone range from simple percentages to complex analyses of variance, multiple regressions, and general linear models. You can use data ranging from simple integers or binary variables to multiple response or logarithmic variables. SPSS also provides extensive data management functions, along with a complex and powerful programming language. In fact, a search at Amazon.com for SPSS books returns 2,034 listings as of March 15, 2004.

In these two sessions, you won't become an SPSS or data analysis guru, but you will learn your way around the program, exploring the various functions for managing your data, conducting statistical analyses, creating tables and charts, and preparing your output for incorporation into external files such as spreadsheets and word processors. Most importantly, you'll learn how to learn more about SPSS.

Installing the Data

The data for this tutorial is available on floppy disk (if you received this tutorial as part of a class) and on the Internet. Use one of the following procedures to install the data on your computer.

Installing files from the Internet

Before you begin to download the files, create a new folder on your computer's hard disk named **SPSSTutorialData**. When you download the files, you'll save them in this directory.

1. Go to ftp.datastep.com. Do not type "www" or "http://".
2. When prompted for a user name, enter: **SPSS**
3. For the password, enter: **tutorial**.
4. To download each file, click it once, press Ctrl-C or select Edit > Copy from the menu.
5. Switch to a window for your computer and save the file in the directory named **SPSSTutorialData**. To do so, open the folder and press Ctrl-V or select Edit > Paste from the menu.
6. Once you have downloaded all the files, close the Internet browser.

Installing files from the diskette

1. With the diskette in the floppy drive, open a window for the C drive on your computer (or any other drive where you want to save the data files).
2. Create a new folder named **SPSSTutorialData**.
3. Copy the files from the floppy disk to the **SPSSTutorialData** folder by copying and pasting or by dragging.

Introducing the interface

When you use SPSS, you work in one of several windows: the data view, the variable view, the output view, the draft output view, and the script view. Eventually you'll also use the syntax editor (think: code) to save or refine your queries.

The data view

The data view displays your actual data and any new variables you have created (we'll discuss creating new variables later on in this session).

1. From the menu, select File > Open > Data.
2. In the Open File window, navigate to C:\SPSSTutorialData\Employee data.sav and open it by double-clicking. SPSS opens a window that looks like a standard spreadsheet. In SPSS, columns are used for variables, while rows are used for cases (also called records).
3. Press **Ctrl-Home** to move to the first cell of the data view.
4. Press **Ctrl-End** to move to the last cell of the data view.
5. Press **Ctrl-Home** again to move back to the first cell.

The variable view

'At the bottom of the data window, you'll notice a tab labeled **Variable View**. The variable view window contains the definitions of each variable in your data set, including its name, type, label, size, alignment, and other information.

1. Click the Variable View tab.
2. Review the information in the rows for each variable.

Note: While the variables are listed as columns in the Data View, they are listed as rows in the Variable View. In the Variable View, each column is a kind of variable itself, containing a specific type of information.

3. Click the Data View tab to return to the data.
4. Double-click the label **id** at the top of the id column. Notice that double-clicking the name of a variable in the data view opens the variable view window to the definition of that variable.
5. Click the data view tab again.

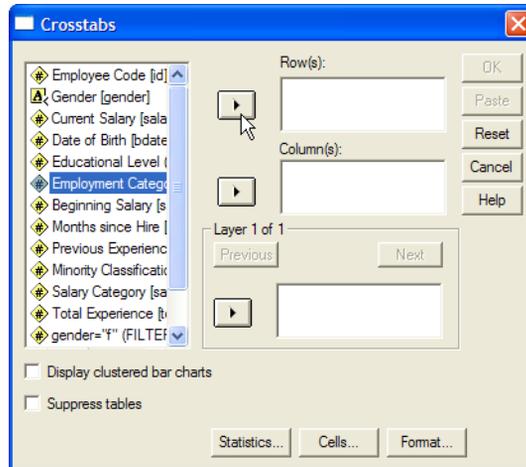
The output view

The output window is where you see the results of your various queries such as frequency distributions, cross-tabs, statistical tests, and charts. If you've worked with Excel, you're probably used to seeing all your work on one page, charts, data, and calculations. In SPSS, each window handles a separate task. The output window is where you see your results.

Try it:

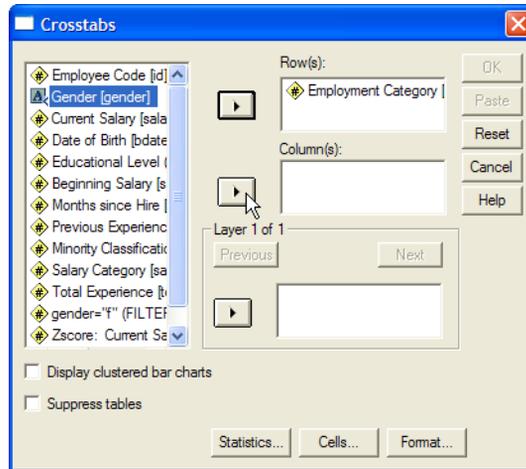
1. From the menu, select Analyze > Descriptive Statistics> Crosstabs.
2. Click once on Employment, then click the small right arrow next to Rows to move the variable to the Rows pane (Figure 1).

FIGURE 1. Moving a variable to the Rows pane



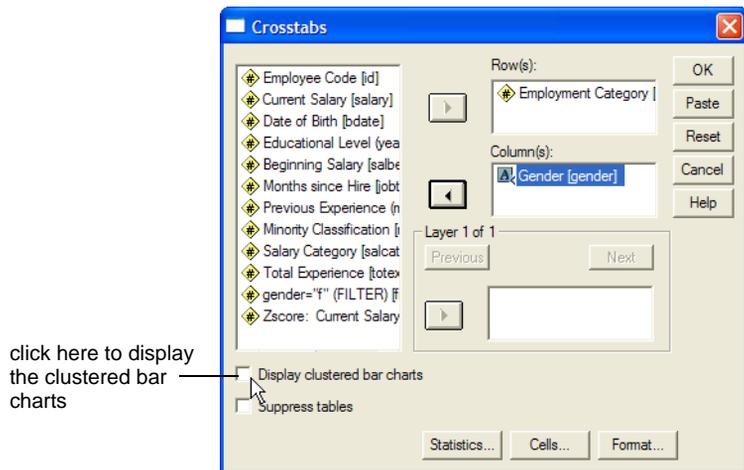
3. Click **Gender**, then click the small right arrow next to Columns to move the Gender variable to the Columns pane. (Figure 2).

FIGURE 2. Selecting a column variable



4. Click **Display clustered bar charts** (Figure 3).

FIGURE 3. Selecting clustered bar charts



5. Click **OK**. SPSS brings the output window to the front displaying two tables and the clustered bar chart you requested. Take a moment to review the contents of the tables and the chart. Notice that the red arrow next to the title Crosstabs corresponds to the Crosstabs icon in the left pane of the window. The left pane displays the contents of the right pane and is a convenient method of moving around among the various output you'll be generating.

Note: In some cases, you may see asterisks instead of numbers in a table cell. Asterisks indicate that the current column is too narrow to display the complete number. Widen the column by dragging its margin to the right.

The draft view

The draft view is where you can look at output as it is generated for printing. The draft view does not contain the contents pane or some of the notations present in the output pane.

Try it:

1. From the menu, select File > New > Draft Output. SPSS opens a Draft Output window that contains its own menu.
2. From the menu, select Analyze > Descriptive Statistics > Crosstabs. Notice that the dialog box opens with your previous selections.
3. Click **OK**. From here you can select charts or tables, copy them, and paste them into other applications like spreadsheets or word processors.

Note: If you want to maintain the correct spacing of the tables, use a non-proportional font like Courier New.

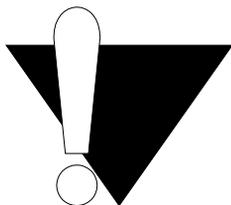
The syntax view

SPSS has never lost its roots as a programming language. Although most of your daily work will be done using the graphical interface, from time to time you'll want to make sure that you can exactly reproduce the steps involved in arriving at certain conclusions. In other words, you'll want to replicate your analysis. The best method of preserving the exact steps of a particular analysis is the syntax view. In the syntax view, you'll preserve the code used to generate any set of tables or charts.

Syntax is basically the actual computer code that produces a specific output. It looks like this:

```
CROSSTABS
  /TABLES=jobcat  BY gender
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT
  /BARChart .
```

In the code shown above, SPSS is instructed to create crosstabs, using the variable *jobcat*, sorting the crosstabs by gender using a specific format, to put a count into each cell, and then to create a corresponding barchart.



Note: Preserving the steps you take in arriving at a conclusion is especially important if you are writing for publication, peer review, or any other situation in which others might want to test your conclusions.

In the next steps, you'll create a simple chart and frequency distribution, save the syntax and then recreate the chart and frequency distribution by running the saved syntax.

Try it:

1. From the menu, select Analyze > Descriptive Statistics > Crosstabs. Notice that your previous selections are still present.
2. This time, instead of clicking OK, click **Paste**.
3. SPSS opens the Syntax Editor with the code you just pasted.

When you select **Paste** instead of **OK** in dialog boxes like the Crosstabs box, the code generated by the function is pasted into the Syntax Editor. If you wanted, you could generate all your output from the syntax window alone. Generally, however, you will probably work with your data and output until they are just the way you want them, then repeat the steps you took and paste the code into the syntax editor. You can then run the syntax at any time to recreate the output.

4. If the cursor is not already located somewhere in the syntax, click anywhere in the word CROSSTABS, then click the small right arrow on the toolbar or select from the menu Run > Current. SPSS opens the draft output window with the same chart you created using the menu commands. This time, however, you created the output by running the syntax (code) you created with the Paste function. Notice that you can scroll up to the previous output you created using the menu commands.

What the heck is a crosstab?

A crosstab (short for *cross tabulation*) is a summary table, with the emphasis on *summary*. Here's an example:

Employment Category * Gender Crosstabulation

Count		Gender		Total
		Female	Male	
Employment Category	Clerical	206	157	363
	Custodial	0	27	27
	Manager	10	74	84
Total		216	258	474

Notice that the rows contain one set of categories (employment category) while the columns contain another (gender). In this crosstab, the cells contain counts, but in others you can use percentages, means, standard deviations, and the like.

Here's the important part: crosstabs are used for only categorical (discrete) data, that is, groups like employment categories or gender. You can't use a crosstab for continuous data like temperature or dosage or income. BUT, you *can* change data like temperature or dosage or income into categories by creating groups, like income less than \$25,000, income between 25000 and 49999, income 50000 or higher. We'll discuss these data conversions known as **transformations** or **recodes** later. For now, you just need to understand that crosstabs deal with groups or categories.

Now that you've seen the various windows you'll be using, we'll move on to the techniques you'll use in SPSS for managing your data files.

2

Entering and modifying data

In this section, you'll learn how to define variables and create a data set from scratch.

Creating the data definitions: the variable view

It's impossible to talk about SPSS (or any analysis program) without talking about data and types of data. So here goes. Each particular type of information (such as income or gender or temperature or dosage) is called a variable. You can have various types of variables such as numeric variables (any number that you can use in a calculation), string variables (text **or** numbers that you can't use in calculations), currency (numbers with two and only two decimal places) and variables with specific formats.

Variable types

SPSS uses (and insists upon) what are called *strongly typed* variables. *Strongly typed* means that you must define your variables according to the type of data they will contain. You can use any of the following types, as defined by the SPSS Help file.

- **Numeric.** A variable whose values are numbers. Values are displayed in standard numeric format. The Data Editor accepts numeric values in standard format or in scientific notation.
- **Comma.** A numeric variable whose values are displayed with commas delimiting every three places, and with the period as a decimal delimiter. The Data Editor accepts numeric values for comma variables with or without commas; or in scientific notation.
- **Dot.** A numeric variable whose values are displayed with periods delimiting every three places, and with the comma as a decimal delimiter. The Data Editor accepts numeric values for dot variables with or without dots; or in scientific notation. (*Sometimes known as European notation.*)
- **Scientific notation.** A numeric variable whose values are displayed with an embedded E and a signed power-of-ten exponent. The Data Editor accepts numeric values for such variables with or without an exponent. The exponent can be preceded either by E or D with an optional sign, or by the sign alone--for example, 123, 1.23E2, 1.23D2, 1.23E+2, and even 1.23+2.
- **Date.** A numeric variable whose values are displayed in one of several calendar-date or clock-time formats. Select a format from the list. You can enter dates with slashes, hyphens, periods, commas, or blank spaces as delimiters. The century range for 2-digit year values is determined by your Options settings (from the Edit menu, choose Options and click the Data tab).
- **Custom currency.** A numeric variable whose values are displayed in one of the custom currency formats that you have defined in the Currency tab of the Options dialog box. Defined custom currency characters cannot be used in data entry but are displayed in the Data Editor.
- **String.** Values of a string variable are not numeric, and hence not used in calculations. They can contain any characters up to the defined length. Uppercase and lowercase letters are considered distinct. Also known as an alphanumeric variable.¹

Because SPSS uses strongly typed variables, you have to make sure that all the data in any field (variable) is consistent.

1. *SPSS Base System Version 11.5 Help.*

Variable names and labels

Forget all the nice conveniences you find in Windows and the Mac for handling long, interesting file names, or even the leniency that some database applications provide in naming fields. In SPSS, variable names are eight characters, period. And no funny characters like spaces or hyphens. Here are the rules:

- Names must begin with a letter.
- Names must not end with a period.
- Names must be no longer than eight characters.
- Names cannot contain blanks or special characters.
- Names must be unique.
- Names are not case sensitive. It doesn't matter if you call your variable CLI-ENT, client, or CliEnt. It's all *client* to SPSS.

Missing values

If you do not enter any data in a field, it will be considered as missing and SPSS will enter a period for you.

Non-numeric numbers, or when is a number not a number?

Some numbers aren't really numbers. That is, they're numbers but you can't use them in mathematical calculations. Take a phone number, for example, or an account number or a zip code. You can sort them, but you can't add or subtract or multiply them. Well, you could, but the result would be meaningless. In essence, these numbers are actually just text which happens to be numeric. A good example is an address, which contains both numbers and text. We refer to these variables as *string* or *text* variables.

Binary variables

Binary variables are a special subgroup of numeric variables. Sometimes you treat them as strings and sometimes you treat them as numeric. For example, yes/no, male/female, and 0/1 are all binary variables. That is, they have two and only two possible values. Obviously, you can't do a calculation on yes/no or male/female, **BUT** and this is a very big and very important **BUT**, you can recode these variables into numeric values, like assigning a value of 0 to female and 1 to male.

Recoding binary variables is a critically important part of data analysis. Suppose, for example, that all you want to know is whether a specific event occurred. For example, suppose you want to know whether a client ever applied for welfare. As it happens, the data set you're using contains the date when each person applied for welfare for the first time, if they ever did; if they didn't the field is blank. Using this date, you can create a new variable called, say, *welfapp* that contains a 1 if there is any value in the first welfare application field and a 0 if there isn't. Now you have a simple value (0 or 1) that you can use for further calculations or subsetting.

Let's get back to the male/female issue for a moment. Say you have recoded the variable into a 0 for female and a 1 for male. If you calculate an average (mean) for this variable, what you now have is a proportion. Say the average of your new variable is .45. You now know that there are somewhat more men than women in your population. In other words, you have calculated a proportion.

Creating a new data set

If you're doing original research or, in our case, creating databases for clients, there comes a time when you have to create your own data file from scratch. In this task, you'll create a new data set, define a set of variables, and then enter some data in the variables. You'll also create some automatic data entry constraints to improve the accuracy of your data entry.

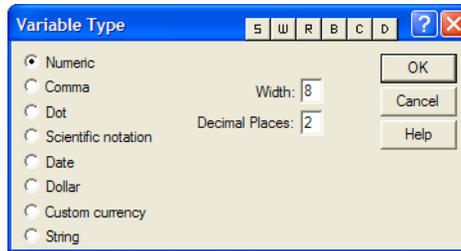
In this task, you will create four types of variables: numeric, date, string, and binary.

1. From the menu, select File > New > Data. If you're asked to save the contents of the current file, click **No**.
2. When the new file opens in the Data View, click the **Variable View** tab at the bottom of the window.
3. With the cursor in the Name column on the first row (referring to the name of the variable) type:

clientid

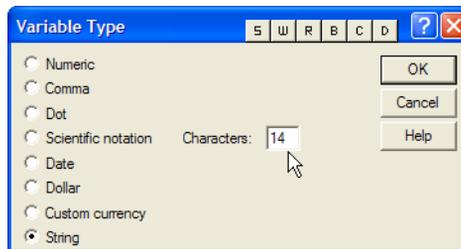
4. In the Type column, click the build button ("build button" is actually a Microsoft term, but since SPSS's documentation doesn't give the button a name, we'll use "build") to open the Variable Type dialog box. (Figure 4)

FIGURE 4. Variable type dialog box



5. Select (click) **String**. Notice that you can now define the length of the variable (Figure 5).

FIGURE 5. Defining the length of a string variable



6. Select all the text in the Characters field and type:

14

7. Click **OK**. The dialog box closes and the variable is now set to a length of 14 with no decimal places.
8. Press **tab** or **Enter** three times to move to the label column.
9. Type:

Client ID

This is the label that will appear on all output and in dialog boxes like those you used in crosstabs and charts.

10. Press **tab** or **Enter** three times to move to the “columns” column. “Columns” defines the width of the *display* of the variable, not its actual contents. The display width affects how the column will be displayed in output like crosstabs and pivot tables.

11. Select all the text in the “columns” column and type:

14

12. Leave the remaining columns as they are, with left alignment and “nominal” as the measure.

13. On the next row, click in the name column and type:

gender

14. Press **tab** or **Enter** to move to the next column.

15. Click the build button to open the Variable Type dialog box.

16. Select **String** and click **OK** to accept the width.

17. Click in the Label column for gender and type:

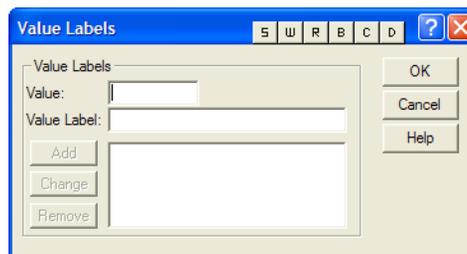
Gender

Notice that in the variable labels, you can use upper and lower case as well as spaces and punctuation.

18. Press **tab** or **Enter** to move to the Values column.

19. Click the build button in the Values column to open the Value Labels window (Figure 6).

FIGURE 6. Value labels window



Note: Variable labels determine how the name of the variable is displayed in output. Value labels determine how each value is displayed. Thus, setting a

label of “Female” for “f” in the gender variable instructs SPSS to display “Female” as a column heading for all cases with a value of *f* in gender.

20. In the Value field, type:

f

21. In the Value Label field, type:

Female

22. Click **Add**.

23. In the Value field, type:

m

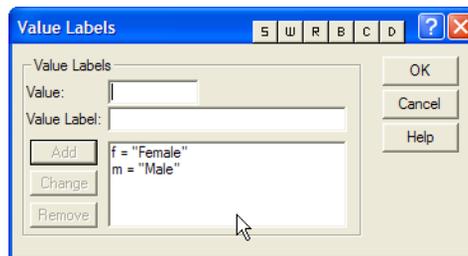
24. In the Value Label field, type:

Male

25. Click **Add**.

26. The Value Labels window should now look like Figure 7.

FIGURE 7. Completed Value Labels window



27. Click **OK**.

28. On the next row, click in the Name column and type:

employed

29. Press **tab** or **Enter** or click in the Type field.

30. Click the build button to open the Variable Type window.

31. Employed is going to be a numeric, binary variable, so leave numeric selected, but change **Width** to **1** and **Decimal Places** to **0**.

32. Click **OK**.

33. Tab to or click in the Label field and type:

Employed year-end

34. Press **tab** or **Enter** to move to the Values field and click the Build Button.

35. In the Value field, type:

1

36. In the Value Label field, type:

Yes

37. Click **Add**.

38. In the Value field, type:

0

39. In the Value Label field, type:

No

40. Click **Add**.

41. Click **OK**.

42. On the next row, click in the Name field and type:

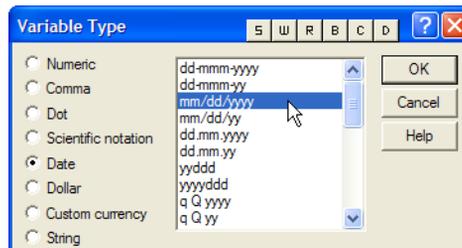
nextelig

43. Press **tab** or **Enter** or click in the Type field and click the build button to open the Variable Type window.

44. Select Date by clicking it.

45. In the pane to the right, select the date format mm/dd/yyyy as in Figure 8.

FIGURE 8. Selecting a date format



46. Click **OK**.

47. Tab to or click in the Label field and type:

Next eligibility date

48. Tab to or click in the columns field and set the column with to **12**.

49. From the menu, select File > Save As.

50. Navigate to the A:\ drive and name the file TestData.sav.

51. Click **OK**.

52. Click the Data View tab. Notice that you now have four columns in which you'll enter data for each record.

53. On the first row, click in the clientid column and type:

4839209

54. In the gender column, type:

f

55. Press **tab**. Notice that when you leave the field, SPSS updates the field to the value label you assigned for *f*.

56. In the Employed field, click the drop-down arrow and select **Yes**.

57. In the nextelig field, type 4/1/2004.

58. Use the following table to complete the data entry for this file.

clientid	gender	employed	nextelig
5533990	m	No	6/1/2004
5938209	f	Yes	4/15/2004
9583902	m	Yes	6/1/2004

You have now seen how you can define variables, assign labels to both variables and values, and define constraints that will control the type of data that can be entered. In future research, you'll probably receive a file that has already been entered in another application such as Access, Excel, or some other application. If you're doing your own data entry in SPSS, however, you should be aware of its data entry capabilities.

Getting help in creating data sets and defining variables

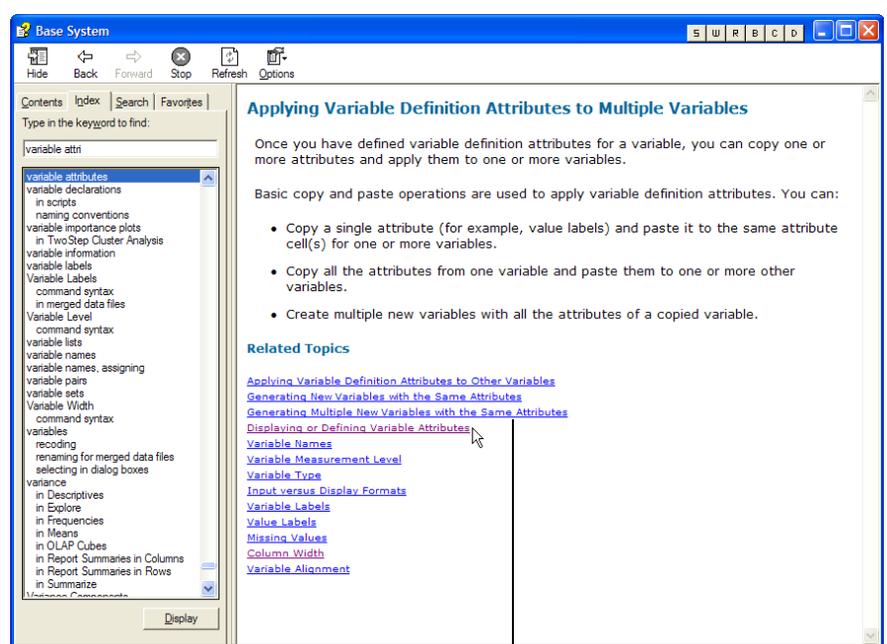
Now that you have had some practice in creating a data set, let's review the kind of help SPSS provides to answer your questions.

1. From the menu, select Help > Topics.
2. On the Index tab, click in the field named: **Type in the keyword to find:**
3. Type:

variable attri

4. Double-click the highlighted text in the list (Variable attributes) to display the topics found.
5. SPSS opens the Topics Found window with **Applying Variable Definition Attributes** highlighted.
6. Click **Display**.
7. The Help window is updated to this topic (Figure 9).

FIGURE 9. Help window for Applying Variable Definition Attributes



click one of the highlighted topics for more information

8. Click the highlighted item **Displaying or Defining Variable Attributes**.
9. When the window is updated, click **Show Me**. SPSS now opens the tutorial window that is included as part of the application. Any time you see **Show Me** in a help window, you can click it to see that specific section of the tutorial.
10. Click the **Next** arrow a few times to see how the steps are illustrated.
11. Close the Show Me window (the Internet Explorer window) to return to the standard help window.
12. Under Related Topics, click Value Labels. As you can see, you can follow the links in the help file, start a new search, or select any of the topics listed on the left to move around in the SPSS Help system.
13. Close the help window.
14. From the menu, select File > Open > Data.
15. Navigate to the Employee data.sav file on the floppy disk and open it.

Creating primary reference lists

There is one set of outputs you'll create that is more important than anything else, and that is the set of primary references. Primary references describe your overall data set. In other words, how many in all? How many in each category? What are the maximums and minimums? Means? Standard deviations? Here's our rule: list out the summaries and put them somewhere where you refer to them quickly.

You may not always want to print out all the details of your data set. For example, printing out every single income for a data set of one million people, would not be useful, economical, or nice to either your printer or the trees. So here are the basic rules: print frequencies for categorical variables and descriptive (also called univariate) statistics for continuous variables.

In this exercise, we'll use the sample Employee dat.sav file.

Frequencies

1. If it's not already open, open the Employee dat.sav file by selecting File > Open and navigating to C:\SPSSTutorialData\Employee data.sav.
2. From the menu, select File > New > Draft Output.
3. From the menu, select Analyze > Descriptive Statistics > Frequencies.
4. Double-click Gender, Employment Category, and Minority Classification to move them to the Variables list.
5. Click the check box labeled **Display frequency tables**.
6. Click **Statistics**.
7. Make sure all the check boxes are cleared (not checked).
8. Click **Continue**.
9. Click **Charts**.
10. If it is not already selected, select **None** by clicking it.
11. Click **Continue**.
12. Click **OK**.
13. From the menu, select File > Save As.
14. Navigate to C:\SPSSTutorialData\ and save the file as AllFreqs.

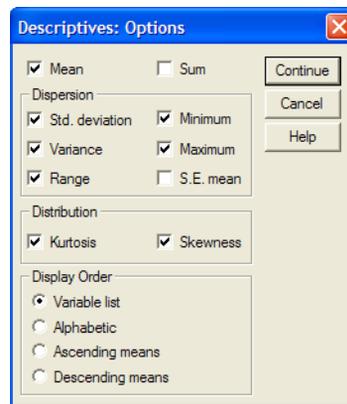
When you're working with your own file, follow these steps, then print the output so you'll have it handy.

Descriptive statistics: descriptives (univariate)

The next step is to print the descriptive or univariate statistics for the continuous variables.

1. From the menu, select File > New > Draft Output.
2. From the menu, select Analyze > Descriptive Statistics > Descriptives.
3. Click **Reset** to clear any previous selections.
4. Double-click Current Salary, Beginning Salary, Months since Hire, and Previous Experience to move them to the Variables list.
5. Click **Options**.
6. In the Descriptives: Options window, click Mean, Std. deviation, Variance, Range, Minimum, Maximum, Kurtosis, and Skewness (Figure 10).

FIGURE 10. Selected measures for Options



7. Click **Continue**.
8. Click **OK**.
9. When the resulting table is displayed, notice that the variables you selected are listed as rows, while the statistics are listed in columns.
10. From the menu, select File > Save As.
11. Navigate to C:\SPSSTutorialData\ and save the file as AllDescriptives.
12. Notice that the statistic **Range** displays the distance between the minimum and maximum.

When you are working with your own file, be sure to print this output and save it someplace handy (we use binders for, well, just about everything). These two printouts, frequencies and descriptives give you a picture of the overall shape of your data.

Recodes and Transformations

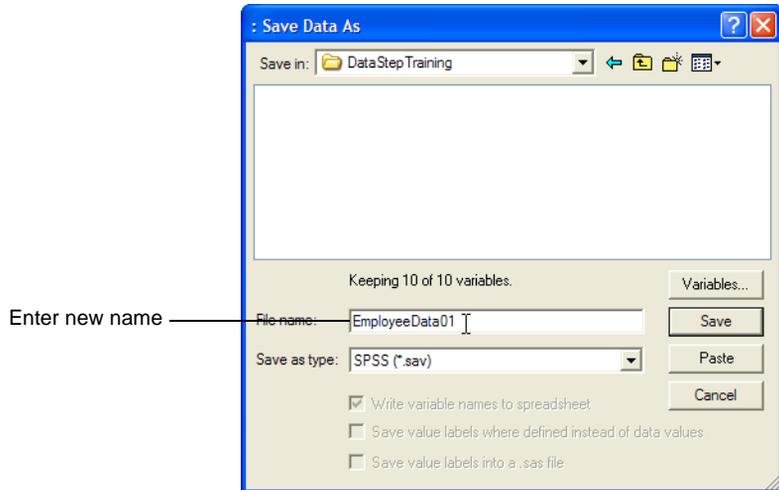
Sooner or later, no matter how carefully you planned your data design, you'll probably want to work with some variables in different forms. If you collected income or age data, for example, you might want to group the continuous variables into categories. Or you might want to create a variable that combines various conditions, say, all minority managers by gender. This type of data manipulation is called transforming or recoding. In this exercise, you'll create several new variables, some that indicate multiple conditions and some that recode continuous variables into categorical variables.

Backup the original file

The first step before making *any* changes to your data file is: **BACK UP YOUR DATA**. And the easiest way to back up your data is to save it under another name.

1. If you don't have the data view open, select it from the menu by selecting Window > Employee data.sav - SPSS Data Editor.
2. From the menu, select File > Save As, then navigate to C:\SPSSTutorialData\.
3. Name the new file EmployeeData01 (Figure 11).

FIGURE 11. Naming the new file



4. Click **Save**. Notice that the title bar of your window now identifies the file as EmployeeData01.sav. Now you can begin your transformations.

Recoding existing variables

Recoding refers to assigning codes (or different codes) to an existing variable.

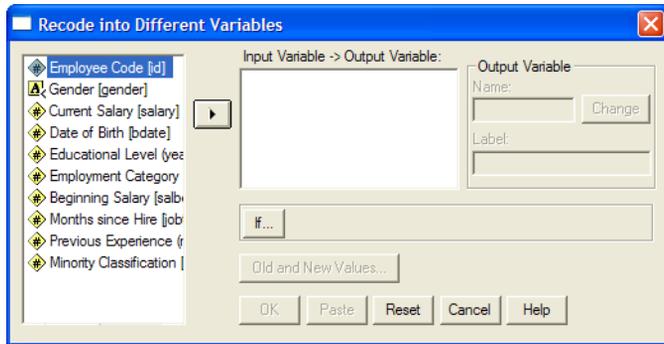


NEVER ever, ever, EVER recode your variables into the same variable name (with one exception). That way lies madness. And chaos. For one thing it deletes your existing data. And for another it destroys the history of the data. Always create a new variable to contain the new codes.

Recode income data

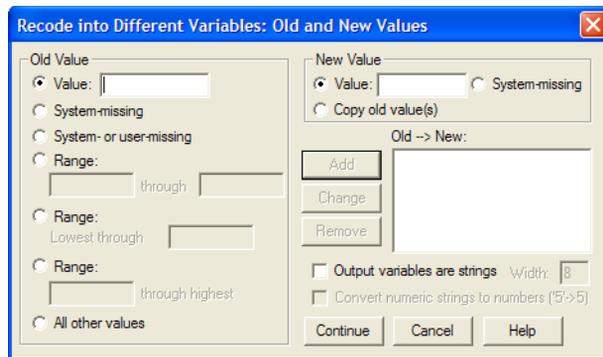
1. From the menu, select Transform > Recode > Into Different Variables to open the Recode window (Figure 12).

FIGURE 12. The SPSS Recode window



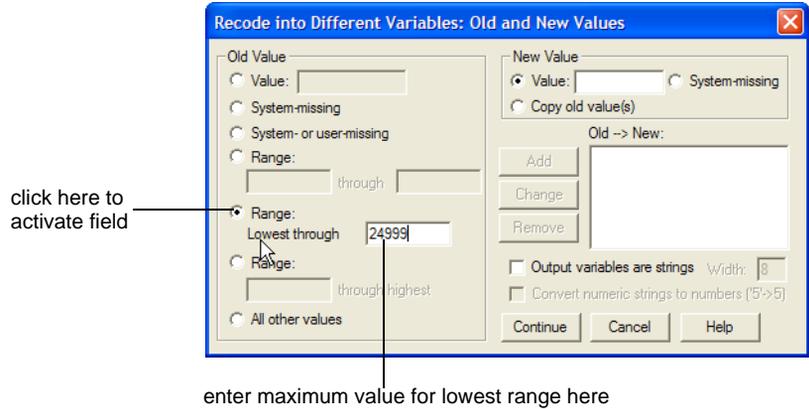
2. Double-click Current Salary to move it to the Input Variable --> Output Variables pane.
3. Click **Old and New Values** to open window where you'll create the new codes (Figure 13).

FIGURE 13. Old and New Values window



4. Click the second **Range** radio button (lowest through) to activate its field (Figure 14).

FIGURE 14. Entering lowest income range



5. Click in the Lowest through range and type:

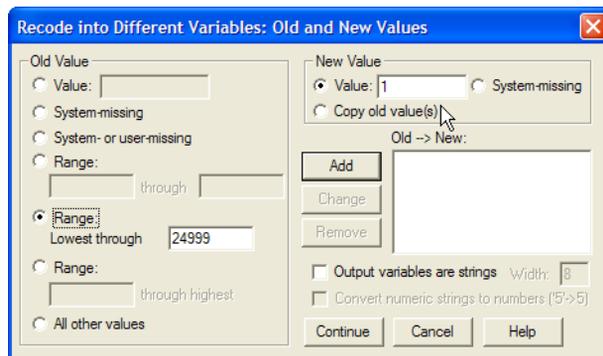
24999

6. Click in the Value field under New Value and type:

1

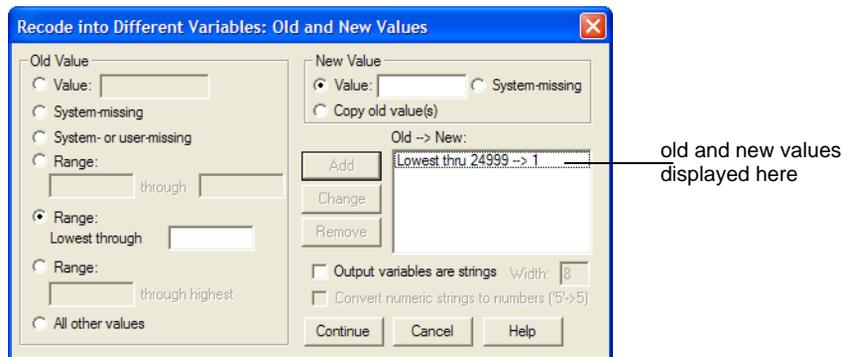
In the new field, any incomes less than 25000 will have a value of 1. (Figure 15)

FIGURE 15. Defining the new value for the lowest income range



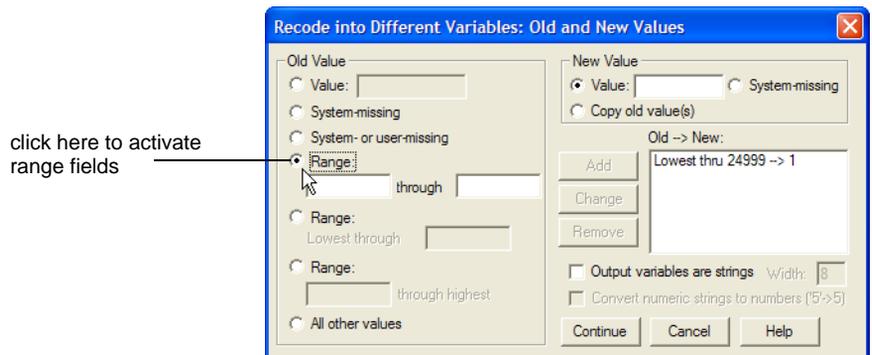
- Click **Add**. Notice that the complete definition of old and new values appears in the Old --> New pane (Figure 16). In the next steps you'll define three more income ranges.

FIGURE 16. Recode window with Old --> New values displayed



- Under Old Value, click the first Range radio button to activate the minimum and maximum values (Figure 17).

FIGURE 17. Activating minimum and maximum range values



- In the first range field, type:

25000

10. In the second range field, type:

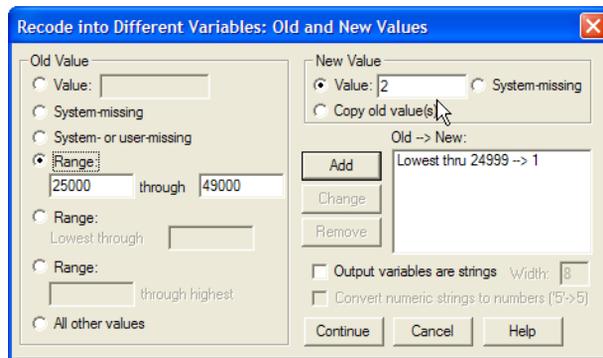
49999

11. Under New Value, type:

2

Your window should now look like Figure 18.

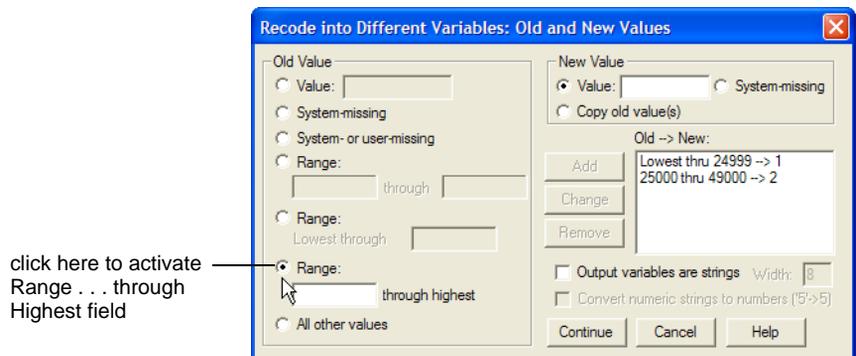
FIGURE 18. Recode window with minimum and maximum range values



12. Click **Add**. Notice that the new definition is added to the Old --> New pane.

13. Under Old Value, click the **third** range radio button to activate the Range . . . through Highest field (Figure 19).

FIGURE 19. Activating the range through highest field



14. In the field to the left of *through Highest* type:

50000

15. Under New Value, type:

3

16. Click **Add**. Again, the new definition is added to the Old --> New pane. In the next steps, you'll create a value to accommodate any odd values that might have been entered into the file. Even if you're sure there aren't any, check anyhow.

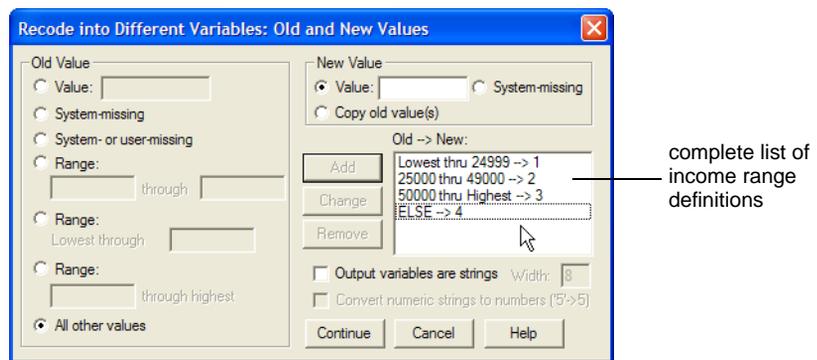
17. Under Old Value, click the radio button for **All other values**. Notice that there is no range to enter.

18. Under New Value, type:

4

19. Click **Add**. Now all possible definitions are entered under the Old --> New pane (Figure 20).

FIGURE 20. Completed definitions for income ranges



20. Click **Continue**. The Old and New Values window closes and the original Recode into Different Variables window is displayed.

21. In the field for Output Variable Name type:

incrange

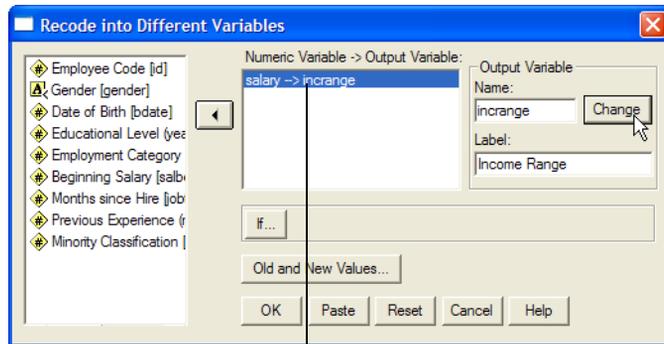
22. In the field for Output Variable Label type:

Income Range

The label is the name that will be displayed on all output, so you'll want to make sure it's informative and correctly formatted.

- Click **Change**. The new name is now listed in the Numeric Variable --> Output Variable pane (Figure 21).

FIGURE 21. Recode window with output variable defined



original name and output name displayed here

- Click **OK**. The Recode window closes and the data view is displayed. Notice that there is now a new column on the right containing the new range codes. (Figure 22)

FIGURE 22. Data view showing new variable

new variable

EmployeeData01.sav - SPSS Data Editor											
File Edit View Data Transform Analyze Graphs Utilities Window Help											
1 : id 371											
	id	gender	salary	bdate	educ	jobcat	salbegin	jobtime	prevexp	minority	incrange
1	371	Female	\$58,125	04/05/1966	16	Manager	\$18,000	71	11	No	3.00
2	348	Female	\$56,750	06/11/1962	16	Manager	\$30,000	73	15	No	3.00
3	468	Female	\$55,750	11/28/1965	16	Manager	\$19,980	64	36	No	3.00
4	240	Female	\$54,375	06/15/1961	16	Manager	\$18,750	81	81	No	3.00
5	72	Female	\$54,000	01/07/1964	16	Clerical	\$18,000	93	11	No	3.00

- Noticed that the codes are displayed with two decimal places. These should be simple integer codes, so in the next step you'll change the format of the variable.

26. Double-click the name **incrange** to open the Variable View with that variable selected (Figure 23).

FIGURE 23. Variable view with incrange selected

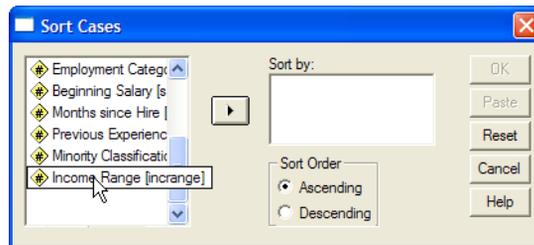


27. Double-click the **2** in the Decimals field and type:

0

28. Press **tab** or **Enter** to leave the field.
29. Click the Data View tab to see the corrected data. In the next step, you'll sort the cases in ascending and descending order of **incrange** to see how the values were applied and to see if there are any values that were not included.
30. Click anywhere in the incrange column.
31. From the menu select Date > Sort Cases to open the sort window (Figure 24).

FIGURE 24. Sort Cases window



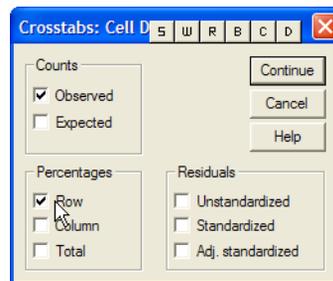
32. Scroll down to Income Range and double-click it to move it to the Sort by pane.
33. Click **OK**. The cases are now sorted according to the lowest income range value.
34. From the menu select Data > Sort cases. Notice that your previous choices are still selected.
35. Click once on Income Range in the Sort By pane.
36. Click **Descending** under Sort Order.

37. Click **OK**. The cases are now sorted with the highest income range value listed first. If there had been any entries that did not fall into the expected categories, they would be listed first, having a value of 4.

Now let's put the new variable to work and display the distribution of cases by income range.

1. From the menu, select New > Draft Output.
2. From the menu, select Analyze > Descriptive Statistics > Crosstabs.
3. In the Crosstabs window, click once on Gender, then click the right arrow to move it into the Rows pane.
4. Click once on Income Range, then click the right arrow to move it into the Columns pane.
5. Click **Cells** to open the Crosstabs: Cell Display window (Figure 25).

FIGURE 25. Crosstabs: Cell Display window



6. In the Percentages pane, click the check box for **Row**. In this case, row percentages will display the percent within gender in each income range. For counts, make sure Observe is checked.
7. Click **Continue**.
8. Click **OK**. The Crosstabs window will close and the new crosstabs will be displayed in the draft output window.
Notice that the income ranges are listed as 1, 2, and 3. Not very informative. So let's go back and assign value labels for the new variable.
9. Close the draft output window without saving.
10. In the Data View, double-click the column heading for **incrange** to open the Variable View for that variable.

11. In the Values column, click the build button to open the Value Labels dialog box.

12. In the Value field, type:

1

13. In the Value Label field, type:

< \$25,000

14. Click **Add**.

15. In the Value field, type:

2

16. In the Value Label field, type:

\$25,000 - \$49,999

17. Click **Add**.

18. In the Value field, type:

3

19. In the Value Label field, type:

\$50,000 or more

20. Click **Add**.

21. In the Value field, type:

4

22. In the Value Label field, type:

Other

23. Click **Add**.

24. Click **OK**.

25. From the menu, select File > New > Draft Output.

26. From the menu, select Analyze > Descriptive Statistics > Crosstabs.

27. Your previous selections are still available, so click **OK**.

In this exercise, you have worked through the entire process of recoding values into a new target variable and practiced sorting the cases to find the minimum and maximum values of the new variable. Finally, you used the newly defined income

ranges to create a crosstab report showing the number and percent of employees within each income category by gender.

Recoding variables revisited

Earlier, we mentioned that there is an exception regarding recoding into the same variables.

The one exception in recoding variables

In fact, there is a case where you might want to recode a value into the same variable, and that is the case of missing or unknown values.

BACK UP YOUR DATA FIRST!

In some cases, you might receive a data set which has missing or just plain odd values in one or more variables. (This will, of course, be a data set you received from someone else, as you would never do anything so silly.) What you can then do is, after identifying the odd values, recode them to a standard value that you will use to represent missing data. You can use any value you want that does not already occur as a valid value in the existing data. For example, suppose you have a field that was supposed to be numeric but someone entered hyphens to indicate that they saw the field but didn't have any data to enter. (Don't laugh. We're working on a project with just that problem.) You might then review all the data, find a value that does not already exist as a valid value and change the hyphens to that value. For example, if you work with legacy data sets, you'll probably find a value of 9 or 99 used to represent missing data.

The other exception

There are cases (we hope they're rare) where you'll want to recode unacceptable values to a flag value that you can use in subsetting your data. For example, suppose you're surveying income and you find that values range from 0 to 55000, except for ten values that are all greater than ten million. As it happens, you can't go back to confirm the odd values, but you don't want to use them in your calculations. In that case, you might want to recode anything over, say, 60000 to a flag value like 99999. In your calculations, you then have a standard value you can use to exclude any outliers or suspicious data from your analyses. As usual, however, be sure to back up your original file before recoding any variables.

On the whole, however, we prefer to recode into different variables.

BE SURE TO RECORD ALL CHANGES TO THE DATA SET.

3

Charting your data

Okay, you've done all the hard work, created the data collection instrument, conducted the interviews or surveys, entered the data, cleaned the data and made any necessary recodes or transformations, and now it's time to find out what it all means. One of the best ways to get an idea of what your data looks like (literally) is to generate charts. Charts provide visual displays of comparisons and relationships. They can also be extremely misleading in the hands of a bad analyst, so be careful. If you're going to be working with charts, two excellent sources of information and guidelines are Edward Tufte's work on visual representations and Gerald Jones's book *How to Lie with Charts*.¹

A word on charts: keep them simple. The purpose of charts is to illuminate relationships and comparisons, not to obscure them. Eschew what Tufte calls *chart junk* and go for simple, clean, and clear designs. But that's a whole other course.

SPSS excels at charts. Many of its functions exceed even those of Microsoft Excel and Microsoft Access, which are pretty good on their own. SPSS provides three methods of creating charts: you can use the automated chart function, you can use the interactive chart function, or you can start with the blank chart and build it from

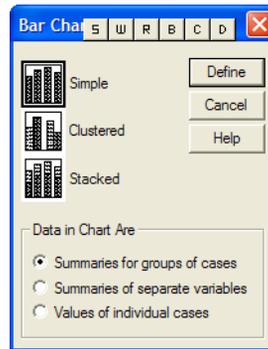
1. Jones, Gerald E., *How to Lie with Charts*, Sybex 1995. Tufte, Edward R., *The Visual Display of Quantitative Information*, Graphics Press, 1983, *Envisioning Information*, Graphics Press, 1990, *Visual Explanations*, Graphics Press, 1997.

scratch. For the charts you'll be creating here, you'll use all three methods to create charts that illustrate gender distribution within job categories.

Using the automated chart function

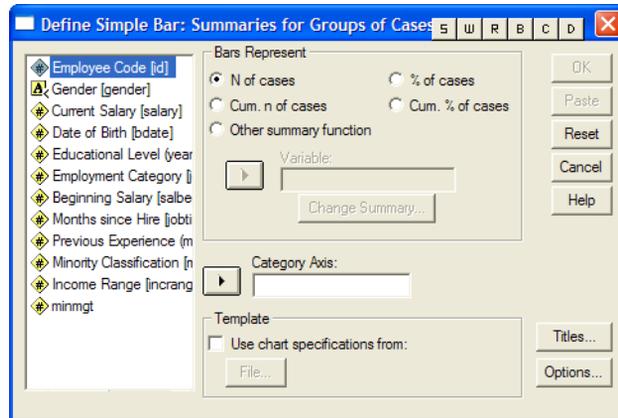
1. If it not already open, open the data file named Employee Data.sav in the folder named SPSSTutorialData.
2. From the menu, select File > New > Draft Output.
3. From the menu, select Graphs > Bar. (Figure 26)

FIGURE 26. Creating a simple bar chart



4. Make sure Simple is selected.
5. Click **Define**. SPSS then displays the Define Simple Bar window (Figure 27) where you will begin to select the variables to be displayed.

FIGURE 27. Define simple bar window



6. Click once on Employment Category, then click the right arrow to move it to the Category Axis field.
7. Click **OK**. SPSS displays the completed chart. Not too interesting, right? Let's make it a little more interesting.
8. From the menu, select Graphs > Bar.
9. This time, select (click) **Clustered**, then click **Define**.
10. Click once on Employment Category, then click the right arrow to move it to the **category axis**.
11. Click once on Gender, then click the right arrow to move it to **Define Clusters By**.
12. Click **OK**. Much more interesting. In the next step, you'll improve your chart with explanatory titles.
13. From the menu, select Graphs > Bar.
14. This time, select (click) **Clustered**, then click **Define**.
15. Click once on Employment Category, then click the right arrow to move it to the category axis.
16. Click once on Gender, then click the right arrow to move it to Define Clusters By.
17. Click **Titles**.

18. On the first line, type:

Job Category Distribution by Gender

19. On the first line under Footnote, type:

Print date: 3/16/2004

20. On the second line under Footnote, type:

Data source: SPSS sample data file, Employee dat.sav

21. Click **Continue**.

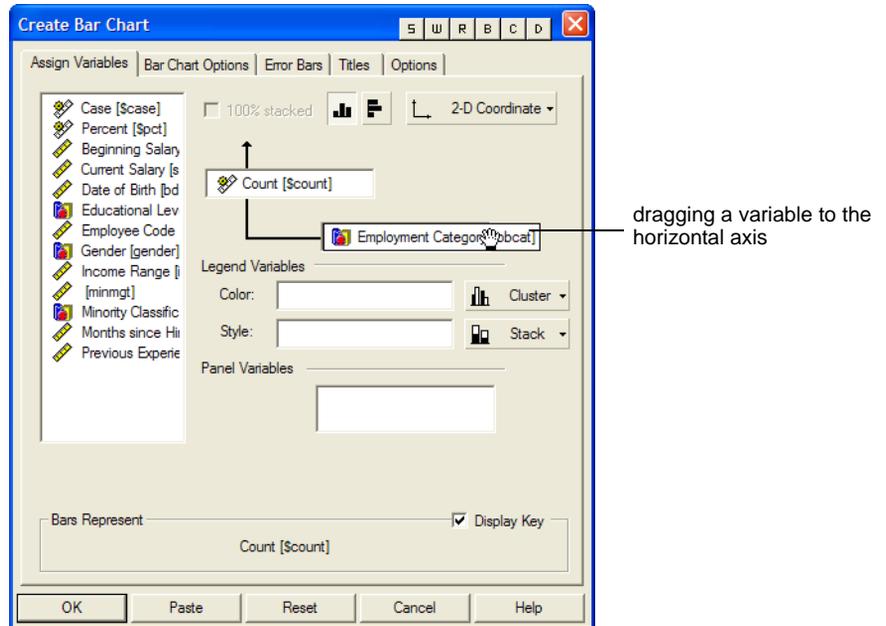
22. Click **OK**.

Using the Interactive Chart function

While the automated chart function is handy, it limits the degree to which you can customize your chart. The interactive function provides more flexibility and power. In the interactive chart window, you'll drag variables to the axis where you want them displayed.

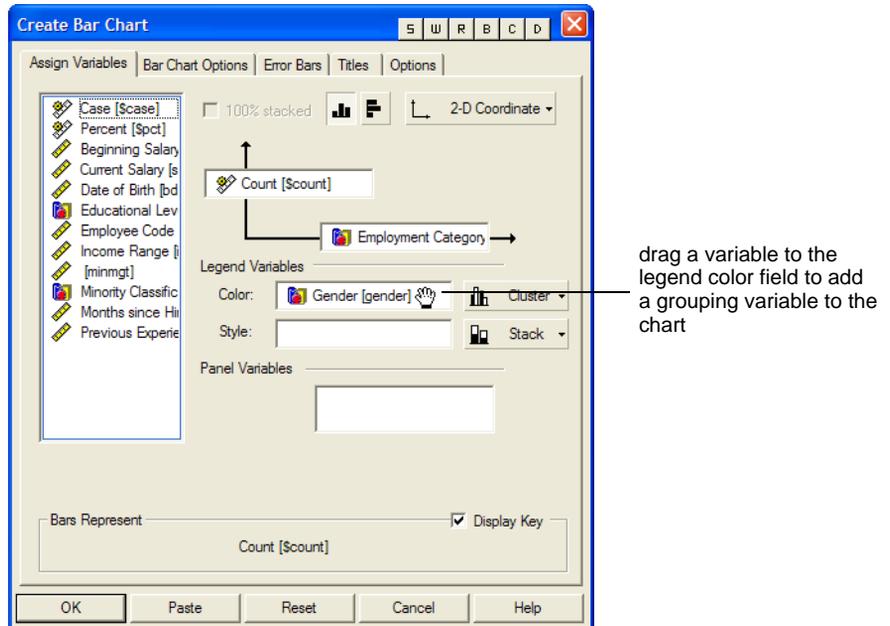
1. From the menu, select **Graphs > Interactive > Bar**.
2. Click once on **Employment Category** and move it to the horizontal axis, as in Figure 28.

FIGURE 28. Moving a variable to the horizontal axis



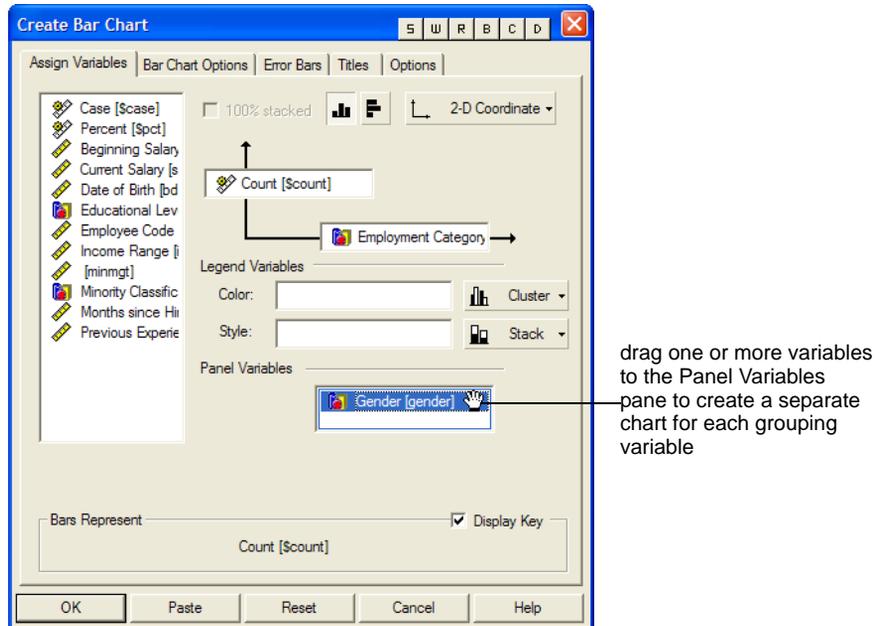
3. Click **OK**. Now you have a simple bar chart, more or less like the first one you created. So let's add some information.
4. From the menu, select **Graphs > Interactive > Bar**. Notice that when the window opens, it still contains the information from your last chart.
5. This time, drag **Gender** to the field under **Legend Variables** called **Color** (Figure 29).

FIGURE 29. Dragging a variable to the legend color field



6. Click **OK**. Now you have a chart with employment category broken out by gender. Now you could add several more grouping variables to the chart, but it would begin to look pretty cluttered. To solve this problem, Tufte came up with the idea of *small multiples*, the practice of displaying smaller charts with each chart displaying only one of the grouping variables. In SPSS, such charts are called *panel charts*.
7. From the menu, select **Graphs > Interactive > Bar**.
8. Notice that your previous choices are still displayed. This time, however, drag Gender from the Legend Variables field down to the field for Panel Variables as in Figure 30.

FIGURE 30. Dragging a variable to the Panel Variables field



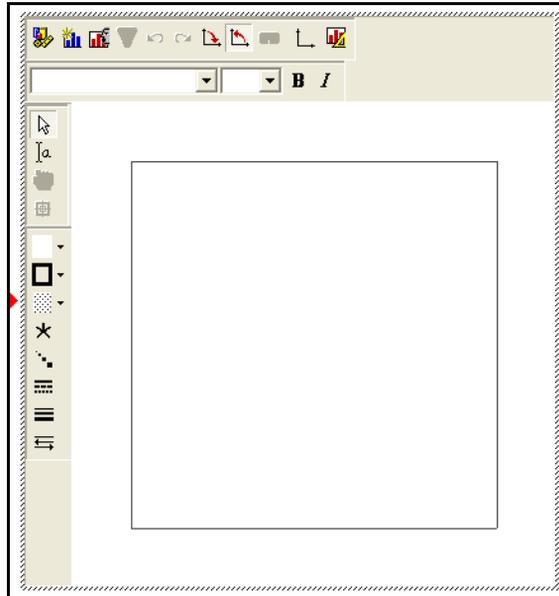
9. Click **OK**. The Draft Output window now displays two charts, one with employment category distributions for women, the other with the same information for men.

Creating a chart from scratch

In the Output window (not the Draft Output window) you can create a chart from scratch, with complete control over every aspect of the graph.

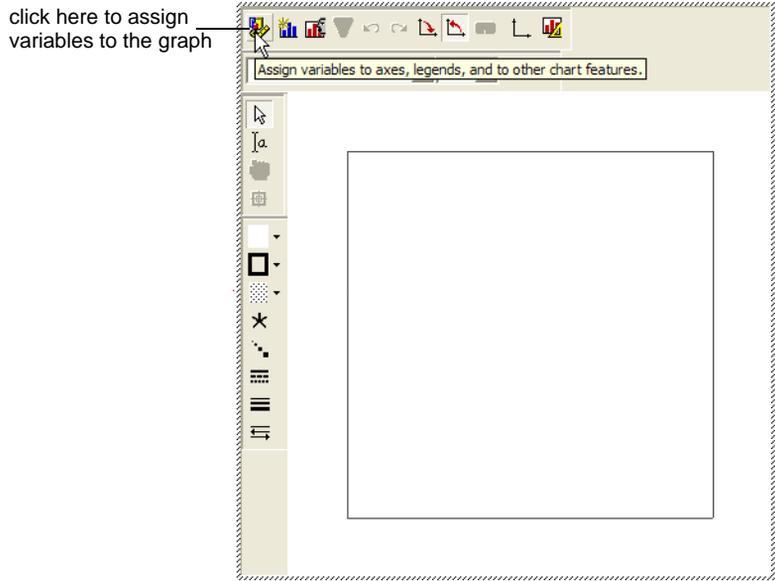
1. From the menu, select File > New > Output.
2. From the menu in the Output window, select Insert > Interactive 2-D Graph. SPSS displays the new, empty graph. (Figure 31)

FIGURE 31. Blank interactive graph



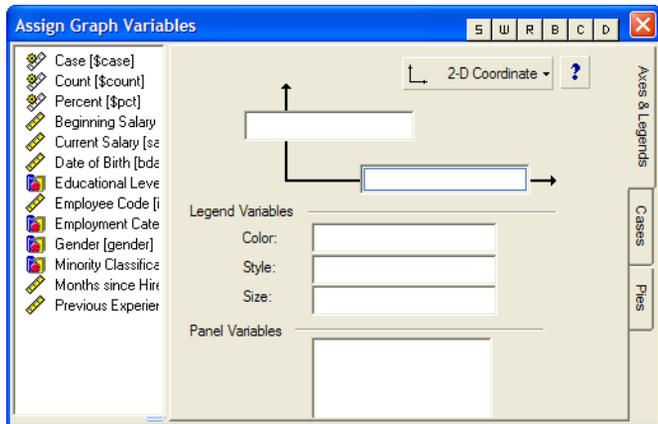
3. Move your cursor slowly over the various icons displayed on the graph. Notice that SPSS displays a description of each icon.
4. Click the Assign Graph Variables icon (Figure 32).

FIGURE 32. Selecting the Assign Graph Variables tool



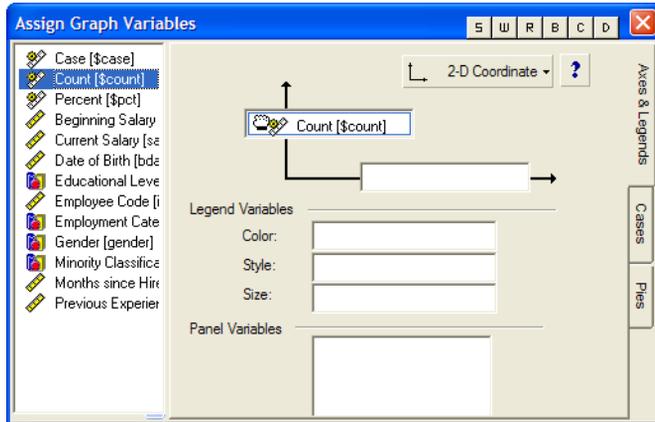
5. SPSS opens the Assign Graph Variables window in Figure 33.

FIGURE 33. Assign Graph Variables window



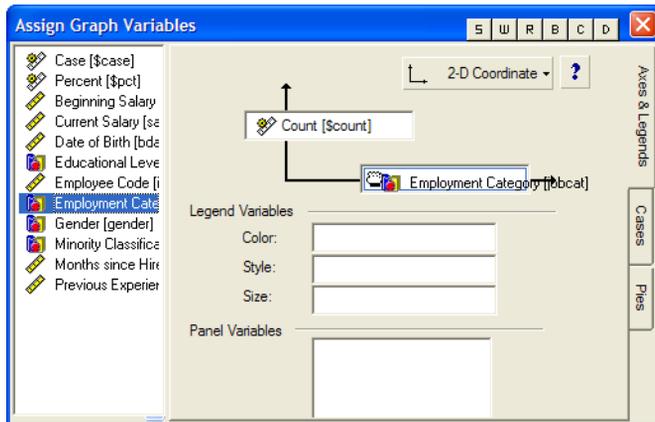
- From the left panel, drag Count [\$count] to the field for the y axis (Figure 34).

FIGURE 34. Dragging a variable to the y axis



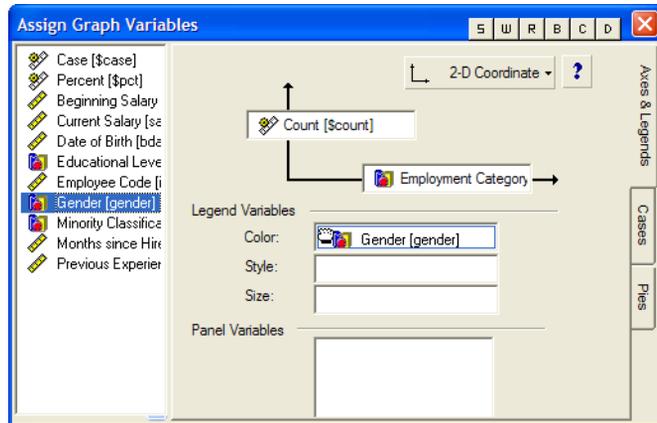
- Now drag Employment Category to the x axis (Figure 35).

FIGURE 35. Dragging a variable to the x axis.



- Finally, drag Gender to the field for Legend Variables Color (Figure 36).

FIGURE 36. Dragging a variable to the Legend Variables Color field



9. Click the X close button. Notice that the chart has the axis variables assigned but still has no data.
10. From the menu, select Insert > Summary > Bar. Your initial chart is now complete.

In fact, SPSS graphing functions go far beyond what we have explored here. You can customize your chart by creating different types of charts (cloud, scatterplots, line graphs, etc.) and by adding elements such as value labels, titles, notes, and much more. Now that you have a general feeling for how graphs work in SPSS, take some time to explore other functions. If you have any questions, be sure to refer to the Help menu.
