



CHAPTER 4

Modifying Data Using the Query Builder

Chapter 4

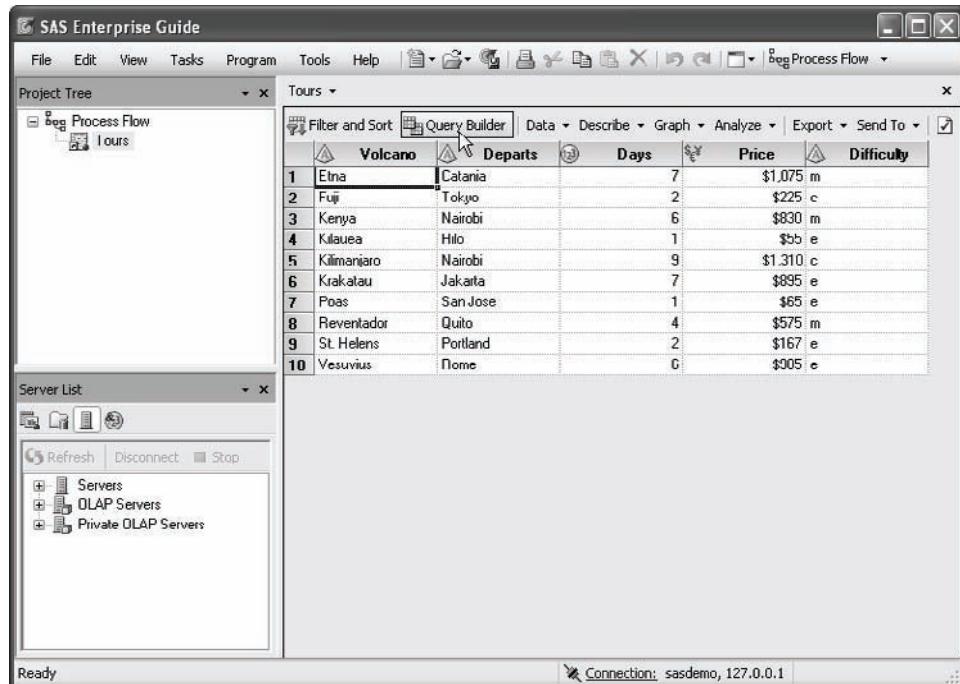
- 4.1 Introducing the Query Builder 202
- 4.2 Selecting Columns in a Query 204
- 4.3 Creating Columns Using Mathematical Operators 206
- 4.4 Creating Columns Using Functions 208
- 4.5 Selected Functions 210
- 4.6 Adding a Grand Total to a Data Table 212
- 4.7 Adding Subtotals to a Data Table 214
- 4.8 Creating Summary Data Tables in a Query 216
- 4.9 Recoding Values in a Query 218
- 4.10 Changing the Result Type of Queries 220

4.1 Introducing the Query Builder

The Query Builder is a tool for manipulating your data. The Query Builder takes a table (or tables), performs some type of data manipulation, and produces a new table. The set of data manipulation instructions defined in the Query Builder is called a query. There are several types of data manipulation you can perform in the Query Builder including:

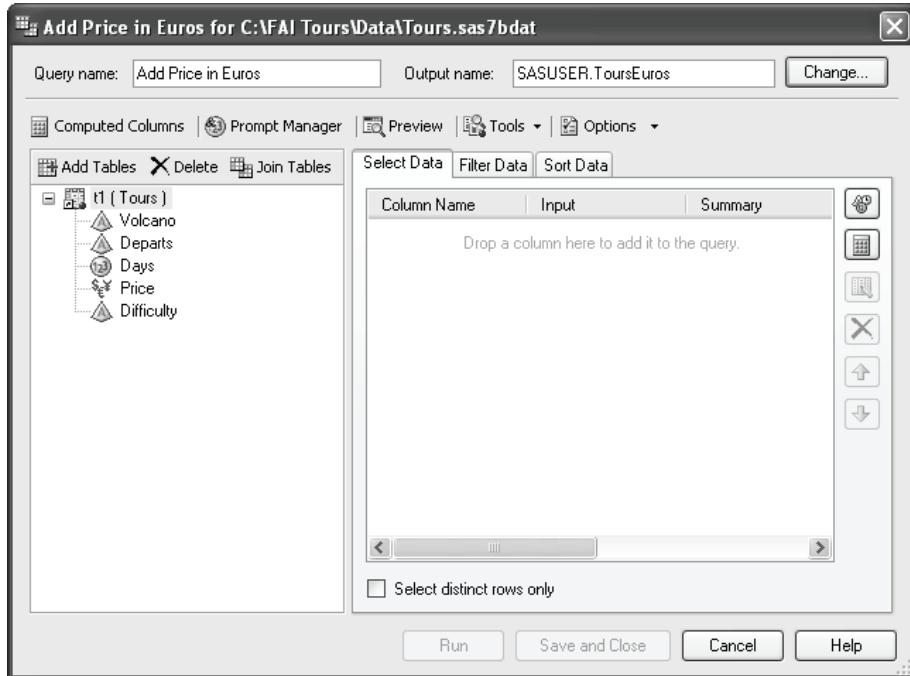
- Creating new columns based on values of existing columns
- Summarizing data
- Sorting data
- Creating a subset of your data
- Joining data tables

Opening the Query Builder There are several ways to open the Query Builder. Perhaps the easiest way is to first open the data table that you want to use for your query, then click the **Query Builder** button on the workspace toolbar.



You can also open the Query Builder by clicking the data icon in the Project Tree or Process Flow to make it active, and selecting **Tasks ▶ Data ▶ Query Builder** from the menu bar. Alternatively, you can right-click the data icon in the Project Tree or Process flow and select **Query Builder** from the pop-up menu.

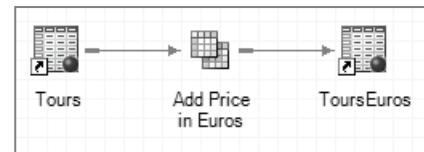
Giving the query a name Queries created in the Query Builder have names that are used to label the query icon in the Process Flow and Project Tree. By default, the first query in your project is named “Query Builder.” Any additional queries will be named “Query Builder 1,” “Query Builder 2,” and so forth. To change the name, enter the new name in the box labeled **Query name** located at the top left. In this example, the query name is “Add Price in Euros.”



Changing the output table name Data tables produced by the Query Builder are stored in a default location and given a name starting with the letters QUERY. To change either the storage location or the table name, click the **Change** button located next to the Output name. This opens a Save File window where you can navigate to the desired storage location and give the table an appropriate name. In this example, the data table will be stored in the SASUSER library and will have the name ToursEuros.

Results The result of a query can be a data table (the default), a data view, or a report. Data views do not contain data. Instead, views contain the instructions needed to create a data table. Data tables and data views produced by the Query Builder can be used in tasks just like any other data table. Reports are for viewing or printing only and cannot be used in tasks. Changing the query result type is discussed in section 4.10.

In order to run the query, you must select the columns you want in your result (discussed in the next section), and you will probably want to perform some other data manipulations. After making the desired selections in the Query Builder and clicking **Run**, a new data table (view or report) will be added to your project. Here is what the Process Flow looks like after running the Query Builder. In this example, the Tours data table is the input to a query named “Add Price in Euros” and the result is a data table named ToursEuros.



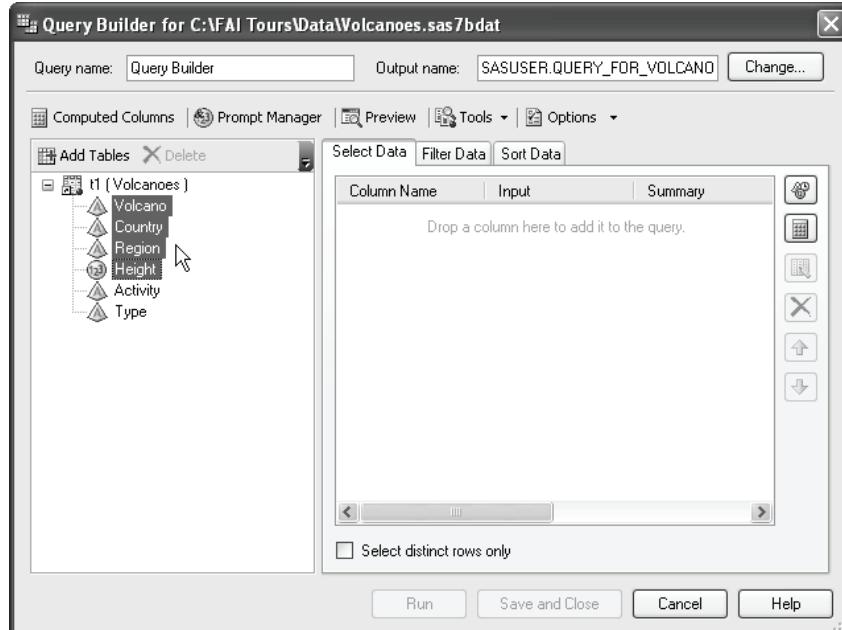
4.2 Selecting Columns in a Query

To run a query, you must tell SAS Enterprise Guide which columns to include in the result. You select columns on the **Select Data** tab of the Query Builder where you can also set properties for columns.

Open the Query Builder by clicking the data icon in the Project Tree or Process Flow to make it active, and selecting **Tasks ▶ Data ▶ Query Builder** from the menu bar. Here is a sample of the Volcanoes data table used in this example.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seal	UK	Eu	251	Extinct	
3	Bamien Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Garibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvoth	Iceland	Eu	1725	Active	Caldera
10	Illimani	Bolivia	SA	6250	Extinct	Stratovolcano

Selecting the data When you open the Query Builder window, the **Select Data** tab is on top, and no columns are selected. To select a column for the query, click the column name in the box on the left and drag it to the **Select Data** tab on the right. You can also right-click the column and choose **Select Column** from the pop-up menu. To add more than one column at a time, hold down the control (CTRL) key (or Shift if you want to select a whole group) when you click the column names. You can also add all the columns in a table to the query, by clicking the table name and dragging it to the **Select Data** tab. In this example, the columns Volcano, Country, Region, and Height have been selected but not yet dragged to the Select Data tab.

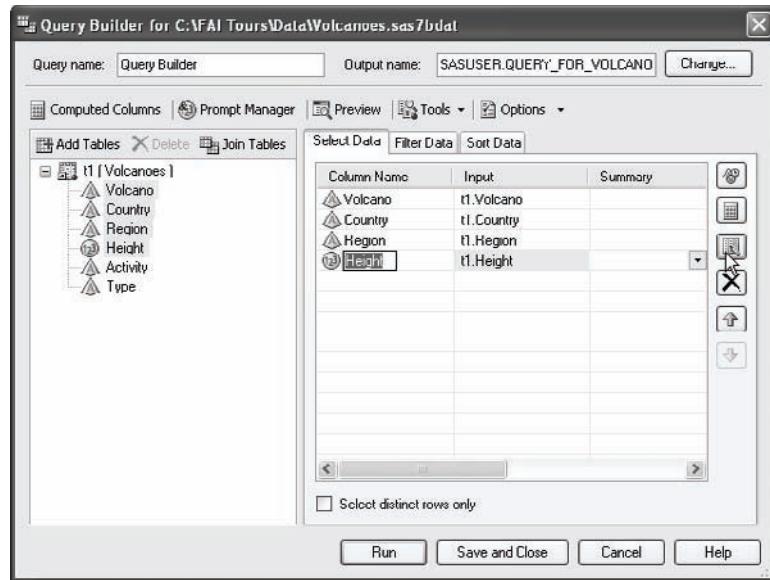


If you want to remove a column from the **Select Data** tab, click the column name on the **Select Data** tab and click the delete button on the right side of the window. You can also change the order of the columns using the up and down arrow buttons .

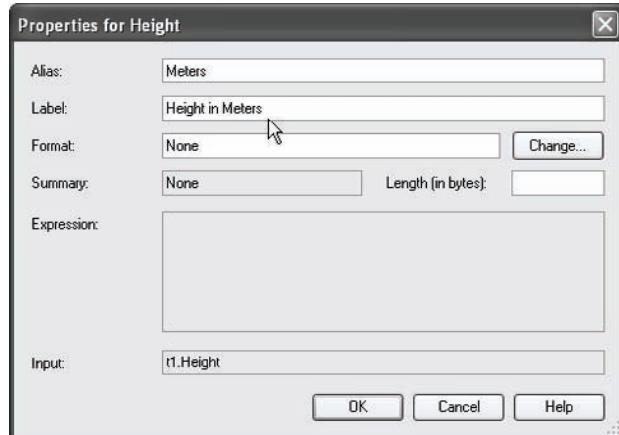
Setting properties for columns You can change the properties of a column in a query by clicking the column name on the **Select Data** tab and clicking the



Properties button located on the right side of the window.



This opens the Properties window for the column. If you want the column name for the output data table to be different than the input data table, then enter the new name in the Alias field. To change the format associated with the column, click the **Change** button. This opens the Formats window where you can choose the desired format. You can also specify a label and length for the column. Here the **Height** column is given the alias **Meters** and label **Height in Meters**. Click **OK** to return to the Query Builder window. After selecting columns and setting properties, click **Run** in the Query Builder window.



Results Here is the data table created by the query. In this result, notice that the column that was named Height in the original table is now named Meters in the output table. Also, the Activity and Type columns from the original table are not included here.

	Volcano	Country	Region	Meters
1	Altar	Ecuador	SA	5321
2	Arthur's Seat	UK	Eu	251
3	Barren Island	India	As	354
4	Elbrus	Russia	Eu	5633
5	Erebus		An	3794
6	Etna	Italy	Eu	3350
7	Fuji	Japan	As	3776
8	Garibaldi	Canada	NA	2678
9	Grimsvotn	Iceland	Eu	1725
10	Illimani	Bolivia	SA	6458

4.3 Creating Columns Using Mathematical Operators

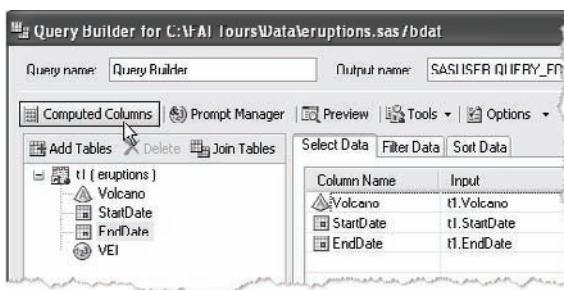
Sometimes you need to create a new column based on data values in other columns. You could do this in a Data Grid, but then if you add new rows, they will not automatically have the computed values. If you use the Query Builder, then every time you run the query it will recompute the new column for all rows including any new ones.

Here is a portion of the Eruptions data table that is used for this example. Open the Query Builder by clicking the data icon in the Project Tree or Process Flow to make it active, and selecting **Tasks ▶ Data ▶ Query Builder** from the menu bar.

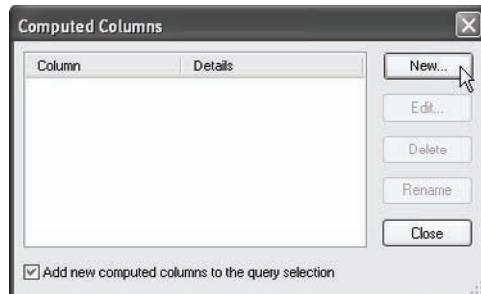
The Query Builder opens with the **Select Data** tab on top. Select the columns for the query. For this example, the Volcano, StartDate, and EndDate columns are selected.

	Volcano	StartDate	EndDate	VEI
1	Barren Island	12/20/1795	12/21/1795	2
2	Barren Island	12/20/1994	06/05/1995	2
3	Erebus	12/12/1912	.	2
4	Erebus	01/03/1972	.	1
5	Etna	02/06/1610	08/15/1610	2
6	Etna	06/04/1787	08/11/1787	4
7	Etna	01/30/1865	06/28/1865	2
8	Etna	12/16/2005	12/22/2005	1
9	Fuji	12/16/1707	02/24/1708	5
10	Grimsvotn	10/31/1603	11/01/1603	2

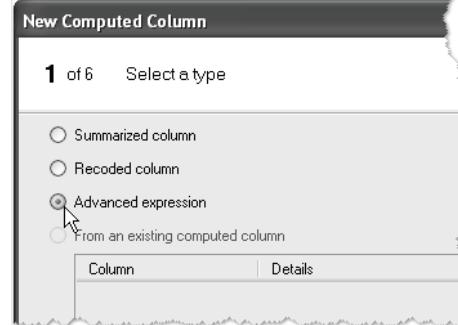
Creating a new column To create a new column, click the **Computed Columns** button located near the top of the window. This opens the Computed Columns window.



Click **New** to open the New Computed Column wizard.



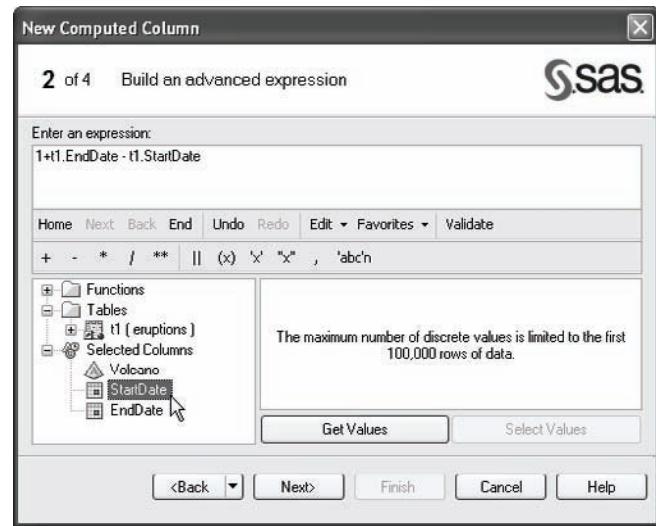
The New Computed Column wizard has up to six windows depending on which type of column you are creating. In the first window, choose **Advanced Expression**, and click **Next**.



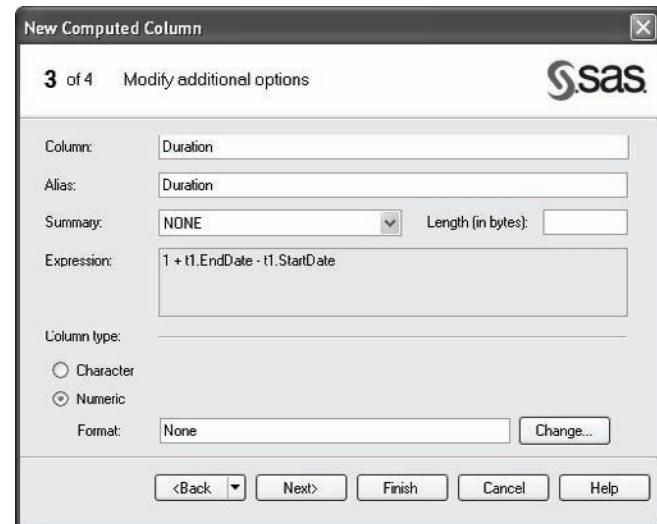
Building the expression

At the top of the second window is a box where you can type your expression. If you like, you can let SAS Enterprise Guide help you build the expression. The bottom left part of the window shows nodes for **Functions**, **Tables**, and **Selected Columns**. To add columns to the expression, expand either the **Tables** or **Selected Columns** node to locate the desired column. Then double-click the column and it will be added to the expression text box. You can use the various operator buttons that appear below the expression text box to build your expression.

In this example, the length of the eruption is computed in days. Because the StartDate and EndDate are both SAS date values (the number of days since January 1, 1960), you can simply subtract the start date from the end date and add one. When you are finished building the expression, click **Next**.



In this example, the length of the eruption is computed in days. Because the StartDate and EndDate are both SAS date values (the number of days since January 1, 1960), you can simply subtract the start date from the end date and add one. When you are finished building the expression, click **Next**.



In the third window you can set the properties of the new column. In this example the column is given a meaningful name and alias. At this point, you can click **Next** to see a summary of your new column in the fourth window (not shown), or click **Finish**. Click **Close** in the Computed Columns window and then click **Run** in the Query Builder.

Results Here are the results of the query including the new column Duration. Notice that the values for Duration are missing for rows where the values for EndDate are missing. If you have missing values for columns that are part of the expression, then the results will be missing.

	Volcano	StartDate	EndDate	Duration
1	Barren Island	12/20/1795	12/21/1795	2
2	Barren Island	12/20/1994	06/05/1995	168
3	Erebus	12/12/1912	.	.
4	Erebus	01/03/1972	.	.
5	Etna	02/06/1610	08/15/1610	191
6	Etna	06/04/1787	08/11/1787	69
7	Etna	01/30/1865	06/28/1865	150
8	Etna	12/16/2005	12/22/2005	7
9	Fuji	12/16/1707	02/24/1708	71
10	Guernsey	10/31/1602	11/01/1602	2

4.4 Creating Columns Using Functions

SAS Enterprise Guide has many built-in functions you can use to build expressions when creating new columns. A function takes a value and turns it into another related value. For example, the MONTH function will take a date and return just the month. The LOG function will return the natural log of a number. There are many functions to choose from in over 20 different categories including character, date and time, descriptive statistics, financial, mathematical, trigonometric, and truncation. Some of the commonly used functions are listed in the next section. You can use functions when you create new columns in the Query Builder.

Here is a portion of the data table that contains data on individual tours including the departure date. To create a column that has the day of month the tour departs, you can use the Day function. To open the Query Builder, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar. The Query Builder opens with the **Select Data** tab on top. Select the columns for the query. In this example Tour, Volcano, and DepartureDate are selected.

	Tour	Volcano	DepartureDate	Guide
1	PS27	Poas	08/05/2011	Carlos
2	SH40	St. Helens	06/19/2011	Casey
3	SH41	St. Helens	07/05/2011	Casey
4	SH42	St. Helens	07/23/2011	Casey
5	SH43	St. Helens	08/15/2011	Kelly
6	FJ12	Fuji	09/12/2011	Cooper
7	ET01	Etna	08/05/2011	Cooper
	KF05	Kenia	05/31/2011	Kelly

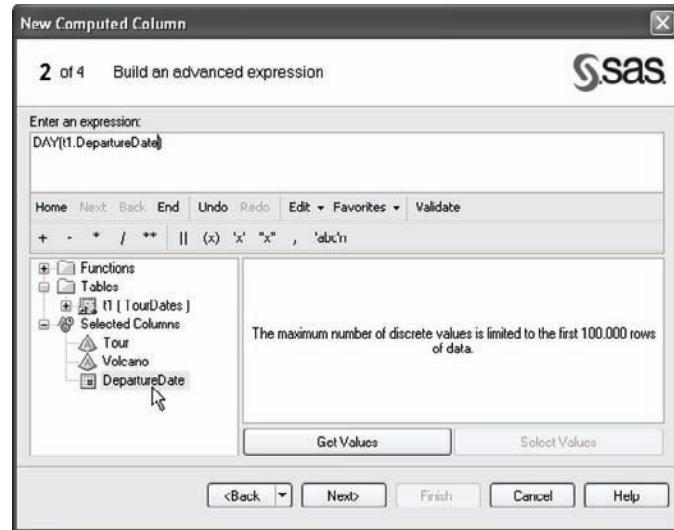
Creating a new column As discussed in the previous section, to create a new column, click the **Computed Columns** button in the Query Builder window. This opens the Computed Columns window. Click **New**, select **Advanced Expression** in the first window of the New Computed Column wizard, and click **Next**.

Choosing a function In the second window, expand the **Functions** node located in the lower left portion of the window to list all available functions. The functions are listed alphabetically and also grouped by category. Expand the **Categories** node to show all the categories of functions. Then expand the node for the desired function category. For this example, choose the **Date and Time** category and scroll down to locate the **DAY** function. Clicking the function name will display information about the function in the lower right portion of the window. Double-clicking the function name will add it to the expression text box located at the top of the window.



Defining arguments for functions Most functions take some sort of argument. When the function is inserted into the expression, a placeholder for the argument appears in the expression. You must replace the placeholder in the function with a valid argument. If the function calls for a character value, that value can be a character column or a character string enclosed in quotation marks. If the function calls for a numeric value, that value can be a numeric column or a number. The DAY function requires a SAS date value that is numeric.

Collapse the Functions node to display the available tables and selected columns. After adding a function to the expression, the placeholder for the function—`<DateValue>` for this example—will be highlighted. Double-click the desired column name, and the placeholder will be replaced by the column you selected, in this case the DepartureDate column from the TourDates table (t1). If you want to use a constant value in the function instead of a column, you can either type the value directly in the expression box in the appropriate location, or use the **Get Values** button to list discrete values for particular columns. Note that it may take a long time to get all the values if your data table is large or if there are many possible values for the column. After listing discrete values, either double-click the value to enter it into the expression, or highlight the value and click **Select Values**.



After you define your expression, you can check it by clicking **Validate**. The Validate window will open telling you whether your expression is valid. Close the Validate window, make any necessary changes, and click **Next**. In the third window of the wizard (not shown), you can give the new column a meaningful name and alias, and also assign the column a format if desired. For this example, the new column is named Day. Click **Next** to display a summary of your new computed column in the fourth window (not shown), then click **Finish** to return to the Computed Columns window. Click **Close** in the Computed Columns window, and then click **Run** in the Query Builder.

	Tour	Volcano	DepartureDate	Day
1	PS27	Poas	08/05/2011	5
2	SH40	St. Helens	06/19/2011	19
3	SH41	St. Helens	07/05/2011	5
4	SH42	St. Helens	07/23/2011	23
5	SH43	St. Helens	08/15/2011	15
6	FJ12	Fuji	09/12/2011	12
7	ET01	Etna	08/05/2011	5
8	KF05	Kenya	05/31/2011	31

Results Here are the query results, which now include the new column Day. Day is the day of the month the tour departs.

4.5 Selected Functions

The following table lists the definition and form of commonly used functions.

Function name	Form of function	Definition
Mathematical		
LOG	LOG(numValue)	Natural logarithm
LOG10	LOG10(numValue)	Logarithm to the base 10
Descriptive Statistics		
MAX	MAX(numValue,numValue,...)	Largest non-missing value
MEAN	MEAN(numValue,numValue,...)	Arithmetic mean of non-missing values
MIN	MIN(numValue,numValue,...)	Smallest non-missing value
SUM	SUM(numValue,numValue,...)	Sum of non-missing values
Character		
LENGTH	LENGTH(charValue)	Returns the position of the last non-blank character (missing values have a length of 1)
SUBSTR{Extract}	SUBSTR(charValue,position,n)	Extracts a substring from a character value starting at 'position' for 'n' characters or until end if no 'n'
TRANSLATE	TRANSLATE(charValue,to-1,from-1,...to-n,from-n)	Replaces 'from' characters in character value with 'to' characters (one-to-one replacement only—you cannot replace one character with two, for example)
UPCASE	UPCASE(charValue)	Converts all letters in character value to uppercase
Date and Datetime ¹		
DATEPART	DATEPART(SAS-datetime-value)	Converts a datetime to a SAS date value
DAY	DAY(SAS-date-value)	Returns the day of the month from a SAS date value
MDY	MDY(month,day,year)	Returns a SAS date value from month, day, and year values
MONTH	MONTH(SAS-date-value)	Returns the month (1–12) from a SAS date value
QTR	QTR(SAS-date-value)	Returns the yearly quarter (1–4) from a SAS date value
TODAY	TODAY()	Returns the current date as a SAS date value

¹A SAS date value is the number of days since January 1, 1960. A SAS datetime is the number of seconds since midnight January 1, 1960.

Here are examples using the selected functions.

Function name	Example	Result	Example	Result
Mathematical				
LOG	LOG(1)	0.0	LOG(10)	2.30259
LOG10	LOG10(1)	0.0	LOG10(10)	1.0
Descriptive Statistics				
MAX	MAX(9.3,8,7.5)	9.3	MAX(-3,,5)	5
MEAN	MEAN(1,4,7,2)	3.5	MEAN(2,,3)	2.5
MIN	MIN(9.3,8,7.5)	7.5	MIN(-3,,5)	-3
SUM	SUM(3,5,1)	9.0	SUM(4,7,.)	11
Character				
LENGTH	LENGTH('hot lava')	8	LENGTH('eruption')	8
SUBSTR{Extract}	SUBSTR('916)734-6281',2,3)	'916'	SUBSTR('Tour12',5)	'12'
TRANSLATE	TRANSLATE ('6/16/2004','-', '/')	'6-16-2004'	TRANSLATE ('hot lava', 'j', 't')	'hot java'
UPCASE	UPCASE('St. Helens')	'ST. HELENS'	UPCASE('Fuji')	'FUJI'
Date and Datetime				
DATEPART	DATEPART(86400)	1	DATEPART(31536000)	365
DAY	DAY(0)	1	DAY(290)	17
MDY	MDY(1,1,1960)	0	MDY(10,17,1960)	290
MONTH	MONTH(0)	1	MONTH(290)	10
QTR	QTR(0)	1	QTR(290)	4
TODAY	TODAY()	<i>today's date</i>	TODAY()-1	<i>yesterday's date</i>

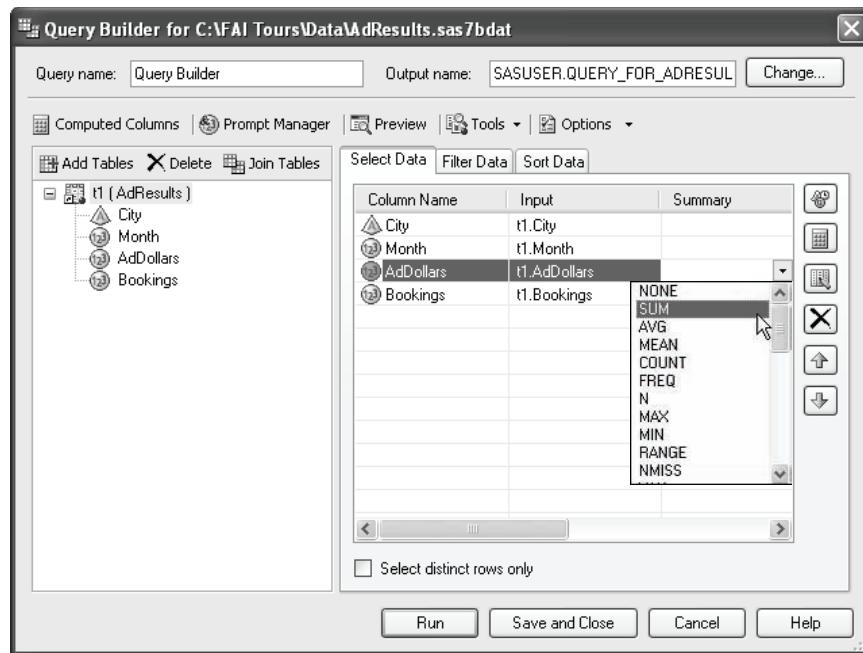
4.6 Adding a Grand Total to a Data Table

Using the Query Builder, you can create columns that contain summary statistics for existing columns. For example, you may want to calculate a grand total over all the rows of data and put the result in a new column. Then you could, for example, compute the percent of the total for each row because the one grand total value is repeated for each row in the data.

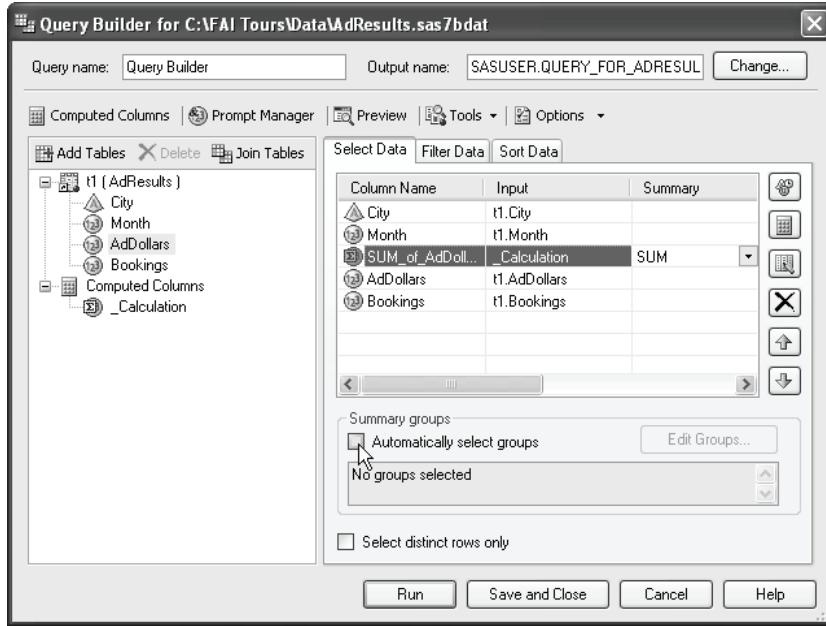
Here is a sample of the AdResults data, which contains the amounts spent on advertising for the Fire and Ice Tours company for both its Seattle and Portland offices. This example creates a new column that has the total amount spent for both offices for the time period. To open the Query Builder, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar.

	City	Month	AdDollars	Bookings
8	Seattle	8	250	17
9	Seattle	9	250	22
10	Seattle	10	325	20
11	Seattle	11	400	25
12	Seattle	12	500	31
13	Portland	1	325	25
14	Portland	2	290	19
15	Portland	3	250	17
16	Portland	4	300	18

Summarizing the data First select the columns for the query, in this case all the columns in the AdResults table. Initially no summary statistics are listed for the columns on the **Select Data** tab. To summarize data in a column, click the column name on the **Select Data** tab, then click the down arrow in the Summary cell. Choose the summary statistic you want to use from the drop-down list. For this example, choose **SUM**.



Selecting summary groups After you choose a summary statistic, the original column will be replaced by the newly summarized column. By default, all selected columns are used for the summary groups. To create a grand total, you don't want any summary groups. Uncheck **Automatically select groups** in the Summary Groups section of the Query Builder so that no groups are selected.



Setting properties for the new column The alias for the new column combines the summary statistic and the old column name. The alias will become the column name in the SAS data table generated by the query. If you don't like the alias, you can change it in the Properties

window. Click the column name, then click the Properties button  on the right side of the Query Builder window to open the Properties window.

Adding back the original column Because the newly computed column replaces the original column, the original column will not be in the output data table unless you add it back. If you want to keep the original column, as well as the newly computed column, then click the original column name in the box on the left and drag it to the **Select Data** tab. Use the up- or down-arrow buttons on the right side of the window to position the original column where you want. In this example, the original column, AdDollars, has been added back in the query and positioned above the Bookings column. When you are satisfied, click **Run**.

Results Here are the results of the query including the new column SUM_OF_AdDollars. Notice that the new column has the same value for all rows.

	City	Month	SUM_OF_AdDollars	AdDollars	Bookings
8	Seattle	8	7845	250	17
9	Seattle	9	7845	250	22
10	Seattle	10	7845	325	20
11	Seattle	11	7845	400	25
12	Seattle	12	7845	500	31
13	Portland	1	7845	325	25
14	Portland	2	7845	290	19
15	Portland	3	7845	250	17
16	Portland	4	7845	300	18

4.7 Adding Subtotals to a Data Table

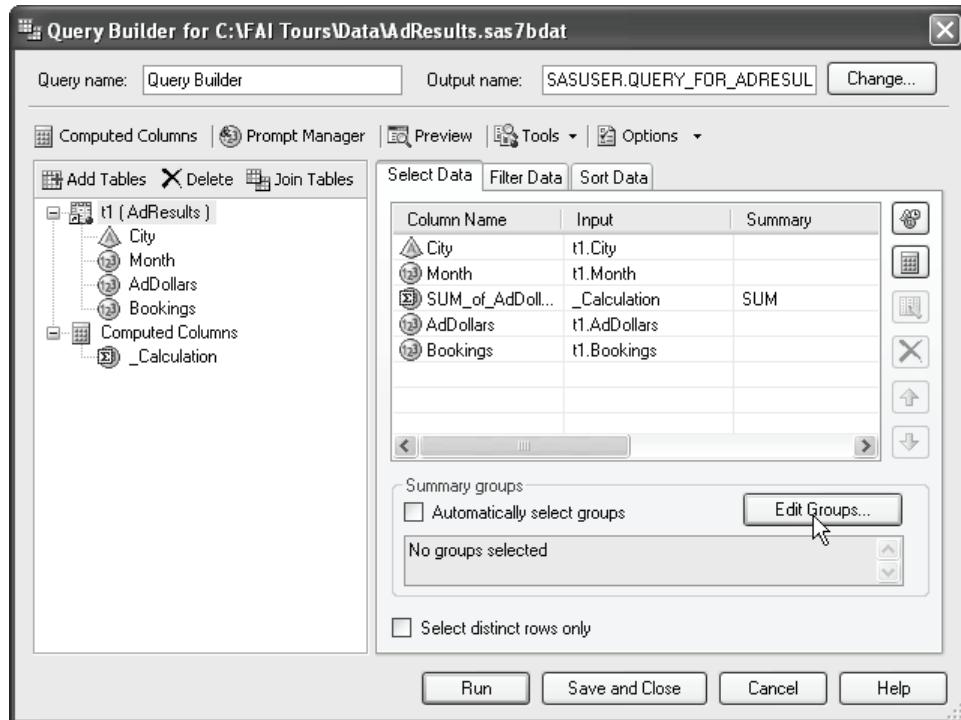
The previous section showed how you can create new columns that summarize all the rows in a data table. This section shows how to summarize all the rows that belong to a group. The steps are the same as adding a grand total, with the additional step of selecting a group column.

Here is a sample of the results from the previous section where the SUM_OF_AdDollars column contains the grand total of the AdDollars column. In this example, the total amount spent by each city's office will be calculated instead of the grand total. Re-open the query by clicking the **Modify**

	City	Month	SUM_of_AdDollars	AdDollars	Bookings
8	Seattle	8	7845	250	17
9	Seattle	9	7845	250	22
10	Seattle	10	7845	325	20
11	Seattle	11	7845	400	25
12	Seattle	12	7845	500	31
13	Portland	1	7845	325	25
14	Portland	2	7845	290	19
15	Portland	3	7845	250	17
16	Portland	4	7845	300	12

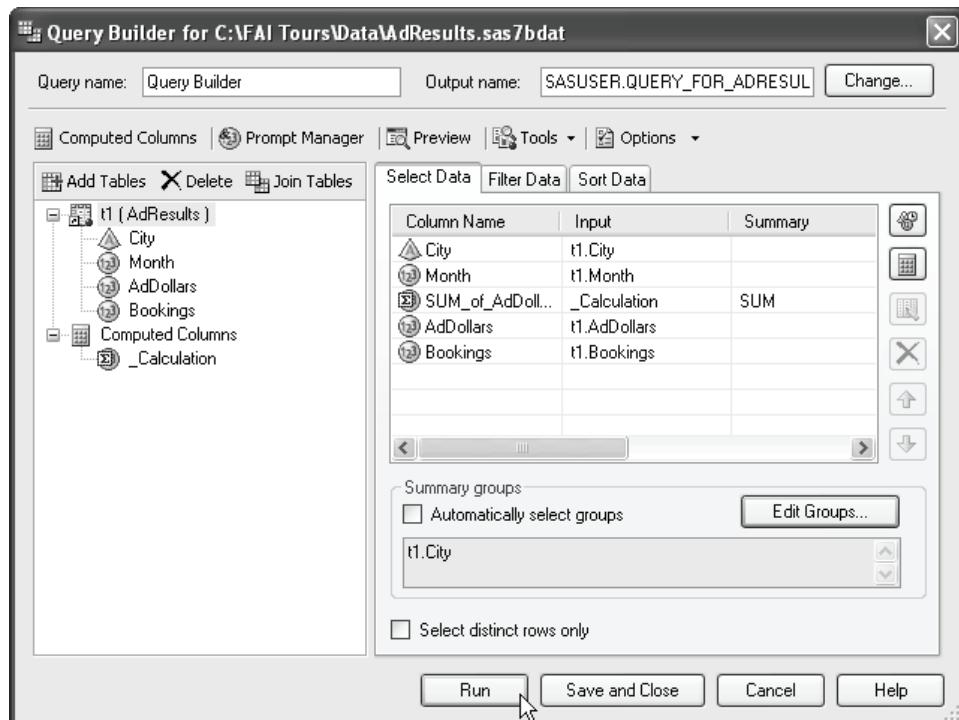
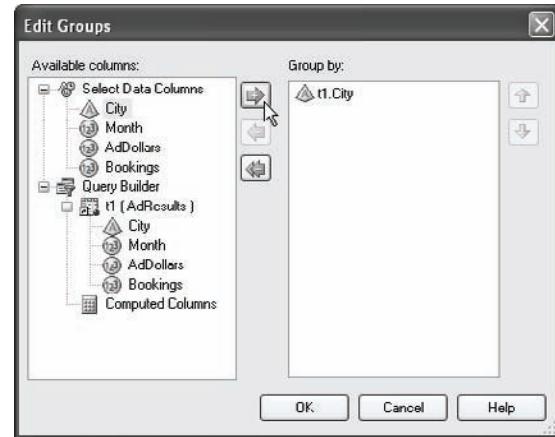
Task button on the workspace toolbar for the query result. This opens the Query Builder window.

Summarizing the data Here is the query from the previous section that contains all the columns from the AdResults table and the SUM_OF_AdDollars column, which is the grand total for the AdDollars column. To change the SUM_OF_AdDollars column to a sub-total by City, click the **Edit Groups** button. This opens the Edit Groups window.



Selecting groups In the Edit Groups window, select the column or columns to use for the groups. Click the column name, City for this example, in the Available columns list and click the plus arrow to add it to the Group by list. Click OK to return to the Query Builder window.

Notice that the name for the summarized column, in this case SUM_OF_AdDollars, does not change. But now the grouping column, City from the AdResults table (t1), appears in the area labeled **Summary Groups** near the bottom of the window. Click Run to run the query.



Results Here are the results of the query. Notice that the rows for Seattle have a different value for SUM_OF_AdDollars than the rows for Portland. Also the resulting data table is now sorted by the grouping column.

	City	Month	SUM_of_AdDollars	AdDollars	Bookings
8	Portland	3	3665	250	17
9	Portland	1	3665	325	25
10	Portland	10	3665	350	24
11	Portland	12	3665	400	33
12	Portland	4	3665	300	18
13	Seattle	7	4180	150	17
14	Seattle	3	4180	525	32
15	Seattle	6	4180	325	18
16	Seattle	12	4180	500	21

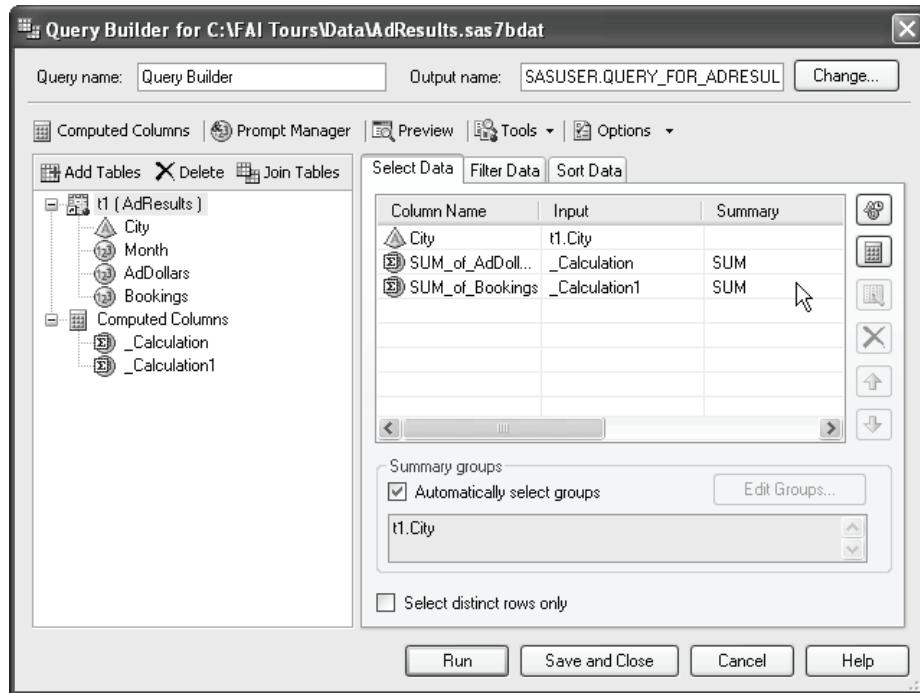
4.8 Creating Summary Data Tables in a Query

The previous section showed how you can create new columns that summarize all the rows belonging to a group. The summarized values were repeated for each row that belonged to the group. But if you want only one row for each group showing just the summarized values, you can do this in a query by eliminating all columns that are not either grouped or summarized.

Here is a sample of the AdResults data that contains the amounts spent on advertising for the Fire and Ice Tours company for its Seattle and Portland offices. This example creates a new table with two rows showing the total number of bookings and amount spent by each office. To open the Query Builder, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar.

	City	Month	AdDollars	Bookings
8	Seattle	8	250	17
9	Seattle	9	250	22
10	Seattle	10	325	20
11	Seattle	11	400	25
12	Seattle	12	500	31
13	Portland	1	325	25
14	Portland	2	290	19
15	Portland	3	250	17
16	Portland	4	300	18

Select the summary and group columns To create a summary data table, select just the columns that will be either summarized or grouped. Choose the type of summarization from the drop-down list that appears when you click the Summary cell for the selected column on the Select Data tab. In this example, the AdDollars and the Bookings columns are summed. Because **Automatically select groups** is checked by default, all the columns that are not summarized are included in the group. For this example, City is the grouping column because it is the only column selected for the query that is not summarized.

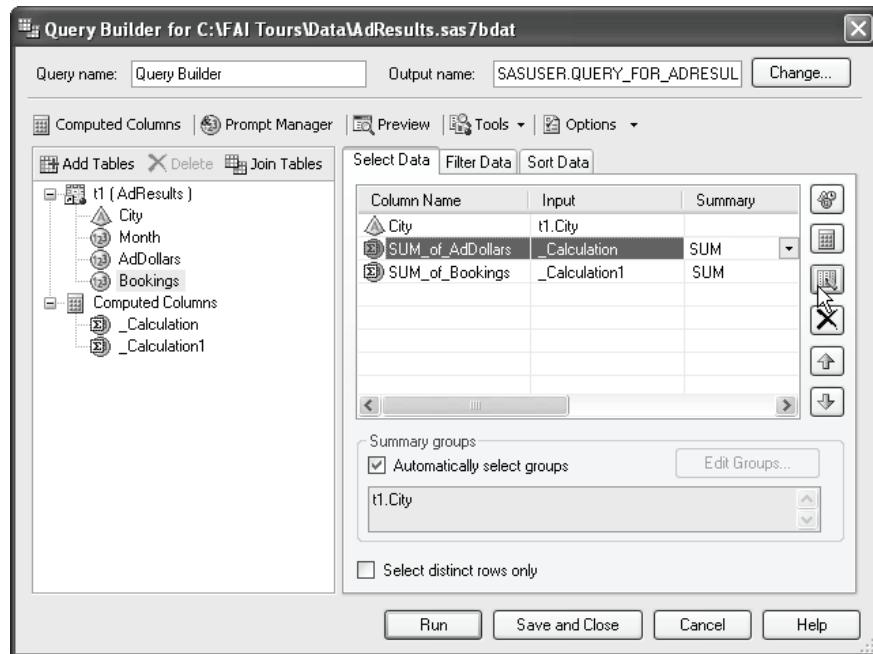


Setting properties for the new columns SAS Enterprise Guide assigns aliases to the summarized columns combining the summary statistic and the old column names. For example, the alias for the sum of the column AdDollars would be SUM_of_AdDollars. The alias will become the column name in the SAS data table generated by the query. If you don't like the alias, you can change it in the

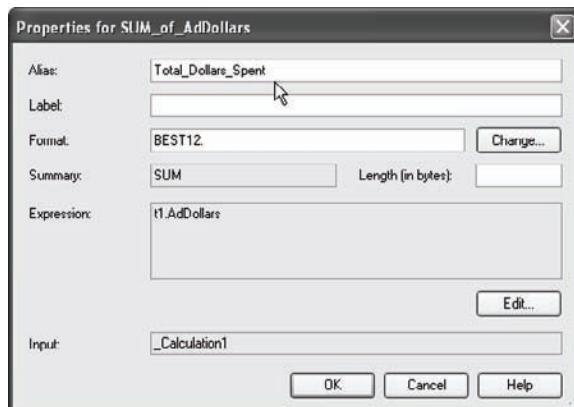
Properties window for the column. Click the column name on the **Select Data** tab of the Query Builder, then click the Properties



button on the right side of the Query Builder to open the Properties window.



In the Properties window for the column, you can specify a new alias, give the column a label, and change the format for the column. After making the desired changes to the properties, click **OK**. Then click **Run** in the Query Builder to produce the result. For this example, the alias for the summarized column **SUM_of_AdDollars** is changed to **Total_Dollars_Spent**.



Results Here is the SAS data table created containing the summarized data. Notice that the new table contains only one row for each value of the grouping column, **City**.

	City	Total_Dollars_Spent	SUM_of_Bookings
1	Portland	3665	254
2	Seattle	4180	270

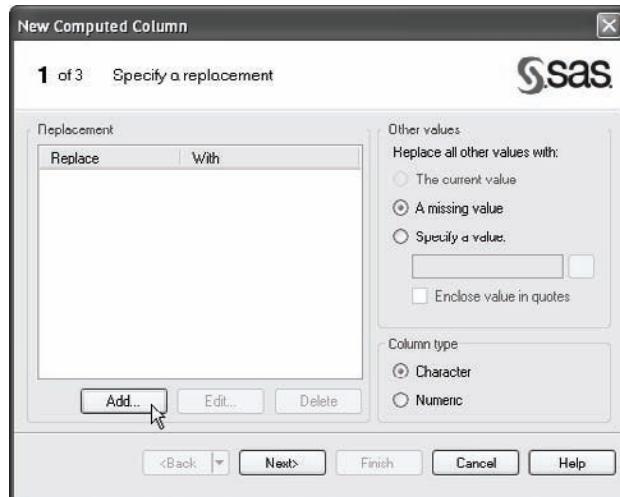
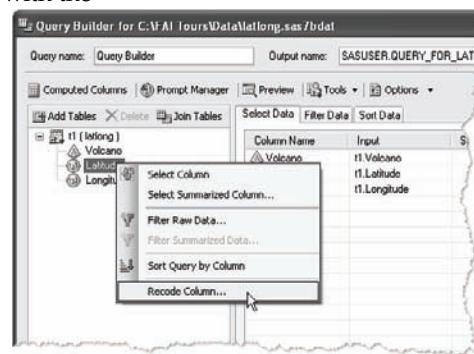
4.9 Recoding Values in a Query

If you want to group data together based on a set of values in a column, you can do this by recoding a column in the Query Builder. For example, if you have sales offices from several different cities, you may want to group them by region. If region is not already defined in the data, then you can define it using the Query Builder. Recoding a column is similar to creating and applying user-defined formats to a column. But, when you recode a column, you create a newly computed column where the data values are actually changed. When you use formats, only the way the data values are displayed is changed.

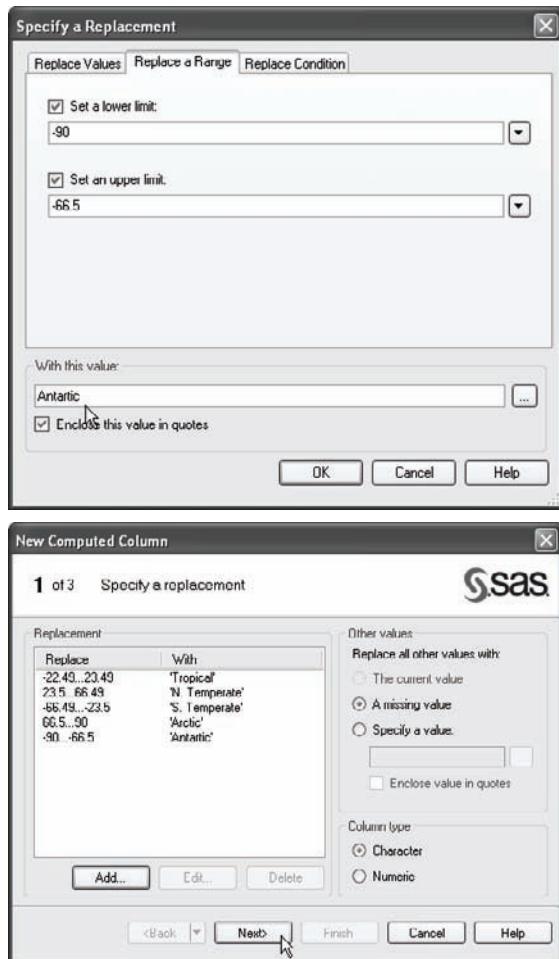
Here are a few rows from the Latlong data table, which gives the latitude and longitude of volcanoes from around the world. Using the recode feature of the Query Builder you can group the volcanoes by zone, according to the value of the column Latitude. To open the Query Builder, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar. This opens the Query Builder window with the **Select Data** tab on top. Select the columns for the query. In this case, all the columns in the table are selected.

Creating the recoded column In the Query Builder window, right-click the column to recode (Latitude for this example) in the list of columns on the left and select **Recode Column**. This opens the New Computed Column wizard.

	Volcano	Latitude	Longitude
1	Altar	-1.67	-78.42
2	Barren Island	12.28	93.52
3	Elbrus	43.33	42.45
4	Erebus	-77.53	167.17
5	Etna	37.73	15
6	Fuji	35.35	138.73
7	Garibaldi	49.85	-123
8	Grimsvotn	64.42	-17.33
9	Illimani	-16.39	-67.47
10	Kenwa	-0.09	37.18



In the first window, specify the type of column to be created (either character or numeric) in the area labeled **Column Type**. For this example, choose **Character**. Then click **Add** to open the Specify a Replacement window.



Defining the replacements In the Specify a Replacement window, there are three tabs: Replace Values, Replace a Range, and Replace Condition. Use Replace Values for one-to-one replacements. It probably makes the most sense to use the Replace a Range feature for numeric data, but it can be used for character data if you want to replace a set of values that fall into a consecutive range alphabetically. Use the Replace Condition feature when you want to use a conditional operator such as "Less Than" or "In a List" to specify your replacements. For this example, click the Replace a Range tab. Enter the lower and upper limits of the range of values to replace in the appropriate boxes. If you want to see what the current values are, then click the down-arrow to the right of the box and a window will open showing all the values. In the box labeled With this value, type the new value that you want to replace the old range of values. When you are satisfied, click OK.

The replacement logic you defined will appear in the first window of the wizard. Click Add again to add another replacement and repeat this procedure for all the replacements you want to make. You can also choose what to do with

values that do not fall into the ranges you specified. For this example, replacing other values with A missing value is fine. When you are finished, click Next.

In the second window (not shown), give the new column a meaningful name and alias and also specify the length and format. For this example, the column name and alias is Zone and the format is None. At this point, you can click Next to see a summary in the third window (not shown), or click Finish. Then click Run in the Query Builder window to run the query.

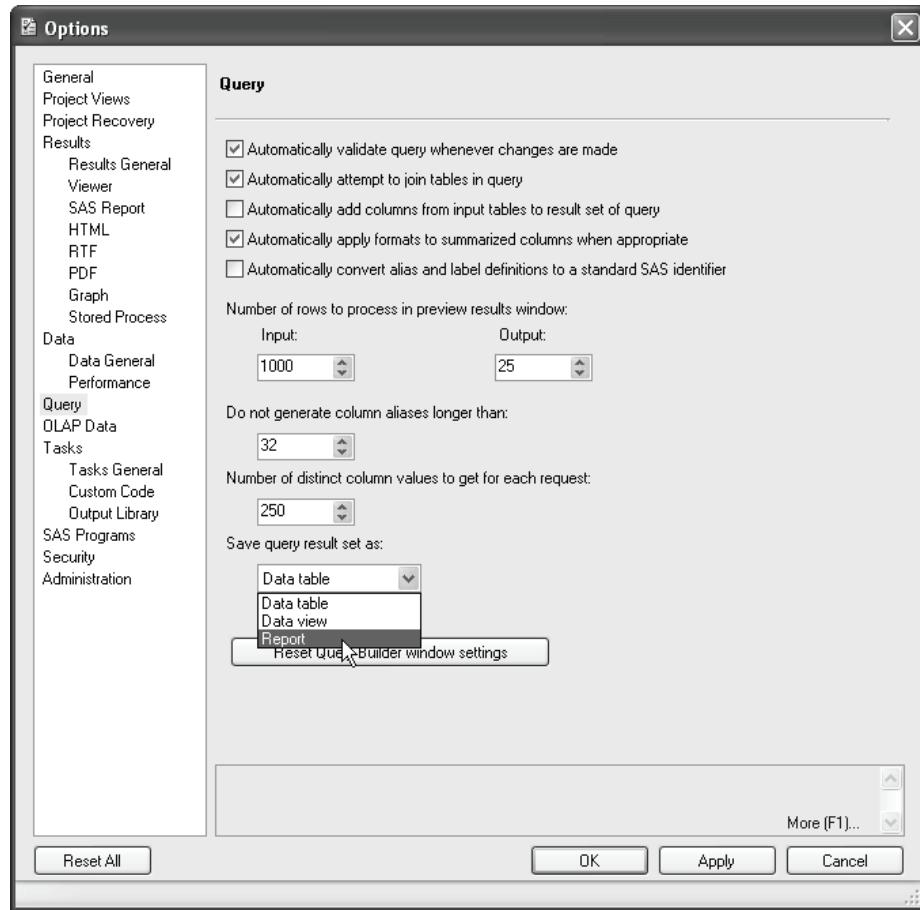
	Volcano	Latitude	Longitude	Zone
1	Altar	-1.67	-78.42	Tropical
2	Barren Island	12.28	93.52	Tropical
3	Elbrus	43.33	42.45	N. Temperate
4	Erebus	-77.53	167.17	Antarctic
5	Etna	37.73	15	N. Temperate
6	Fuji	35.35	138.73	N. Temperate
7	Garibaldi	49.85	-123	N. Temperate
8	Grimsvötn	64.42	-17.33	N. Temperate
9	Illimani	-16.39	-67.47	Tropical
10	Kenia	.0 .09	37.18	Tropical

Results Here is a sample of the query result. Notice that both the Latitude column and the new column, Zone, are part of the query result.

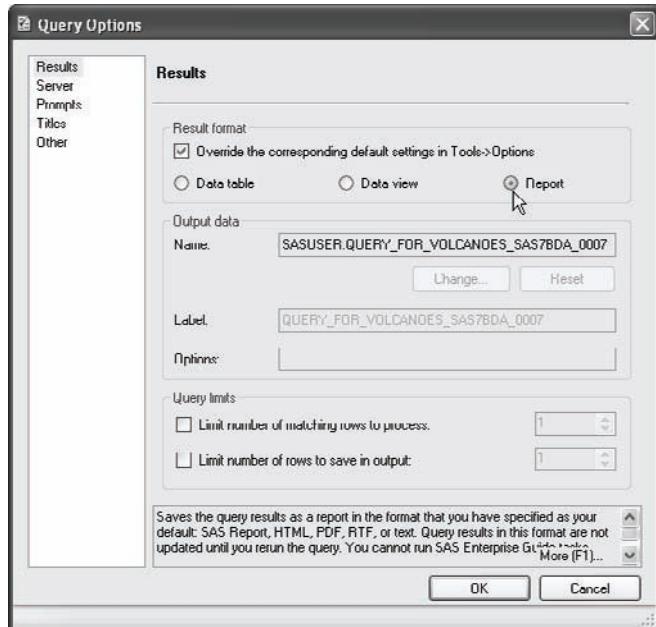
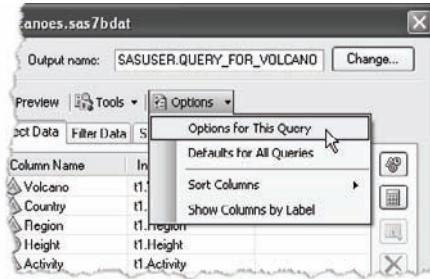
4.10 Changing the Result Type of Queries

When you run a query using the Query Builder, you have a choice about the type of result the query produces. A query can produce a SAS data table, a SAS data view, or a report. A SAS data view is similar to a SAS data table, except that it does not contain any data. Instead, SAS data views contain the instructions required to create a new data table. Data tables and views generated by the Query Builder can be used as a source of data for tasks. Reports are for viewing or printing only.

Setting the default result type The default result type for queries is a data table, but you can change the default. To do this, open the Options window by selecting **Tools ▶ Options** from the menu bar. Click **Query** in the selection pane on the left to open the Query page. Near the bottom of this page is a drop-down list under **Save query result set as**. Select the desired result type. All subsequent queries you build will use this as the default result type.



Setting the result type for a query If you want to change the result type of an individual query, you can do this in the Results page of the Query Options window for the query. Click the **Options** button in the Query Builder window and select **Options for This Query**. Note that you can also select **Defaults for All Queries** to open the Options window shown on the previous page.



Query report All the query examples so far in this book have shown the result as a data table. Here is a partial view of a report generated from a query of the Volcanoes data table with the title Volcanoes Report. Note that in query reports, the column labels are used instead of the column names.

In the Results page of the Query Options window, check **Override the corresponding default settings in Tools -> Options**, and then select the type of result you want for the query. You can choose a data table, data view, or report. Changes made in the Query Options window affect only the results of the current query.

If you choose the Report type, you may also want to give your report a title or footnote. You can assign titles and footnotes in the Titles page of the Query Options window.

Volcanoes Report			
Volcano	Country	Region	Height in Meters
Altar	Ecuador	SA	5321
Arthur's Seat	UK	Eu	251
Barren Island	India	As	354
Elbrus	Russia	Eu	5633
Erebus		An	3794
Etna	Italy	Eu	3350
Fuji	Japan	As	3776
Garibaldi	Canada	NA	2678
Grimsvötn	Iceland	Eu	1725

5

“ I find that a great part of the information I have was acquired by looking up something and finding something else along the way. ”

FRANKLIN P. ADAMS

Attributed to Franklin P. Adams (1881-1960), newspaper columnist, satirist, and poet.



CHAPTER 5

Sorting and Filtering Data

- 5.1 Filtering Data in a Task 224
- 5.2 Using the Filter and Sort Task 226
- 5.3 Sorting Using the Sort Task 228
- 5.4 Sorting Data in a Query 230
- 5.5 Filtering Data in a Query 232
- 5.6 Creating Compound Filters in a Query 234
- 5.7 Creating a Filter with Advanced Expressions 236

5.1 Filtering Data in a Task

Sometimes you don't want to use all the rows in a data table. There are several ways to subset or filter data in SAS Enterprise Guide. If you want to create a new data table, then use the Filter and Sort task or the Query Builder as shown later in this chapter. However, if you simply want to run a task and do not need to create a new data table, then you can filter data directly in the task.

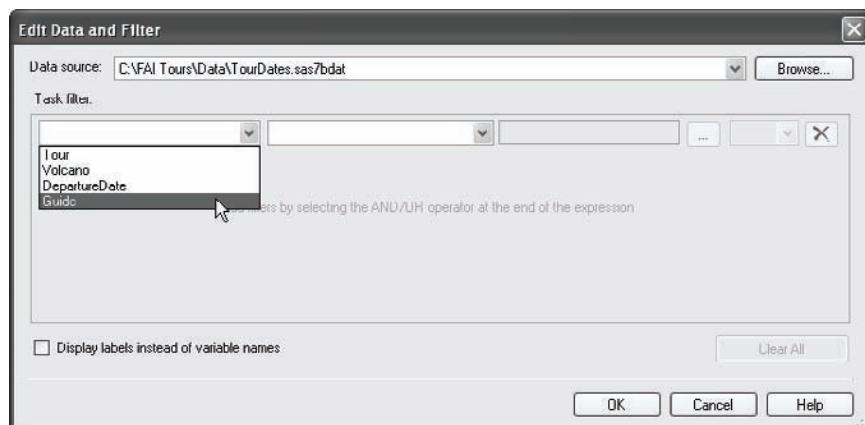
Here is the TourDates data table. This example uses the List Data task to demonstrate how to filter data in a task. To open the List Data task, click the data icon in the Project Tree or Process Flow to make it active. Then select **Tasks ► Describe ► List Data** from the menu bar. The List Data window will open, displaying the Data page. For this example, all the columns in the table are assigned to the List variables role.

	Tour	Volcano	DepartureDate	Guide
1	PS27	Poas	08/05/2011	Carlos
2	SH40	St. Helens	06/19/2011	Casey
3	SH41	St. Helens	07/05/2011	Casey
4	SH42	St. Helens	07/23/2011	Casey
5	SH43	St. Helens	08/15/2011	Kelly
6	FJ12	Fuji	09/12/2011	Cooper
7	ET01	Etna	08/05/2011	Cooper
8	KE05	Kenya	05/31/2011	Kelly
9	KL18	Kilauea	07/08/2011	Malia
10	KL19	Kilauea	07/15/2011	Malia
11	KL20	Kilauea	07/22/2011	Malia
12	RD02	Reventador	07/11/2011	Carlos
13	VS11	Vesuvius	07/21/2011	Cooper
14	VS12	Vesuvius	08/15/2011	Cooper
15	KJ01	Kilimanjaro	06/09/2011	Kelly
16	KK03	Krakatau	07/19/2011	Kelly

Editing the data source The Data source and Task filter are displayed at the top of the Data page in the task window. By default, no filters are applied to data in a task. To make changes to the filter used for the task, click the **Edit** button. This opens the Edit Data and Filter window.

Creating the filter

In the Edit Data and Filter window is an area labeled **Task filter** that contains four empty boxes. Click the down-arrow on the first box, and select the column you want to use for your filter from the drop-down list.

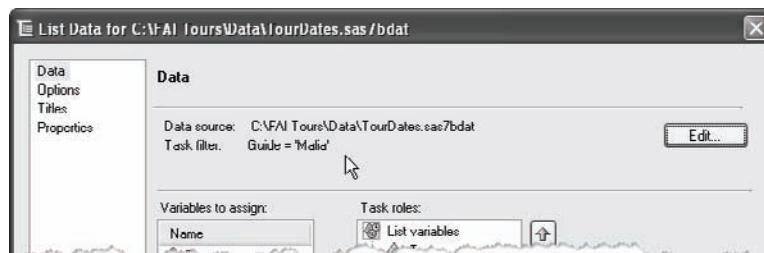
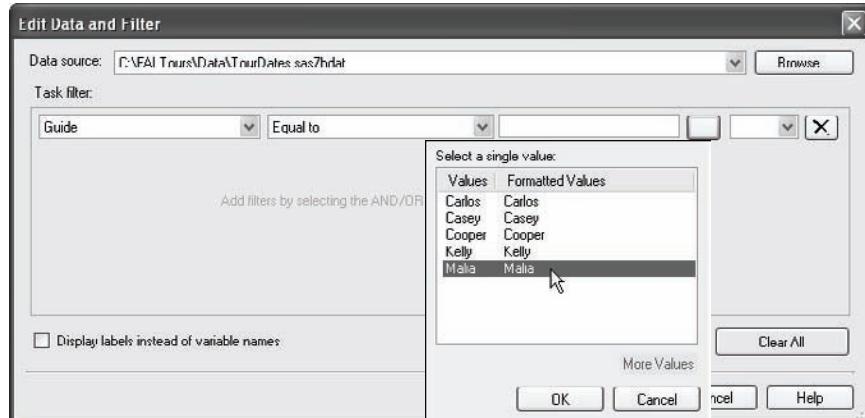
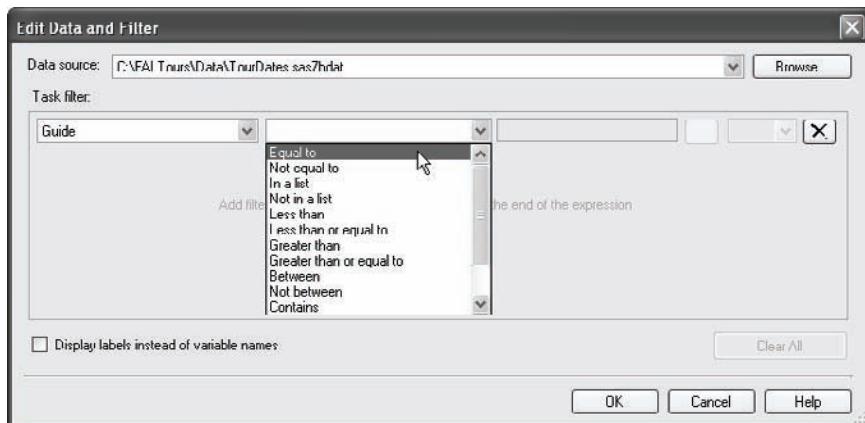


In the second box, select the operator from the drop-down list.

In the third box, either type in a value, or click the ellipsis button [...] to get a list of values for the selected variable. Select the desired value and click OK.

If you want to add another condition to the filter, select AND or OR from the drop-down list in the fourth box. A new row of boxes will appear where you can specify the additional condition. When you are finished specifying the filter conditions, click OK.

The filter you defined will appear in the Task filter area of the Data page in the task window. Click Run to run the task.



Results Here are the results of the List Data task showing only the rows of data that meet the filter condition where Guide is equal to "Malia."

Malia's Tours			
Tour	Volcano	DepartureDate	Guide
KL18	Kilauea	07/08/2011	Malia
KL19	Kilauea	07/15/2011	Malia
KL20	Kilauea	07/22/2011	Malia

5.2 Using the Filter and Sort Task

In the Filter and Sort task you can filter or sort your data, or both. The result of the Filter and Sort task is a data table. If you need to join tables or add any computed columns, then use the Query Builder. If you only want to sort the data, then you may choose to use the Sort Data task instead, which offers more sorting options.

Here is a sample of the Eruptions data table, which is sorted by the volcano name and start date of the eruption. To filter and sort the data, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Filter and Sort** from the menu bar. If you have a Data Grid open, you can also open the Filter and Sort task by clicking **Filter and Sort** on the workspace toolbar. The Filter and Sort window will open.

	Volcano	StartDate	EndDate	VEI
1	Barren Island	12/20/1795	12/21/1795	2
2	Barren Island	12/20/1994	06/05/1995	2
3	Erebus	12/12/1912	.	2
4	Erebus	01/03/1972	.	1
5	Etna	02/06/1610	08/15/1610	2
6	Etna	06/04/1787	08/11/1787	4
7	Etna	01/30/1865	06/28/1865	2
8	Etna	12/16/2005	12/22/2005	1
9	Fuji	12/16/1707	02/24/1708	5
10	Grimsvötn	10/31/1603	11/01/1603	2

Selecting variables The Filter and Sort task has three tabs. On the **Variables** tab, choose the variables you want to keep. Click the desired variable name in the box on the left labeled

Available. Then

click the plus

arrow  to

move the

variable to the

area labeled

Selected. To

move all

variables at once,

click the double

plus arrow .

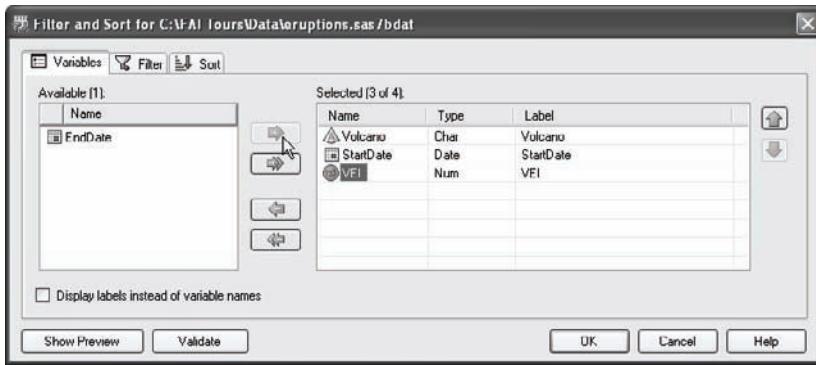
Unselect

variables using

the minus arrow .

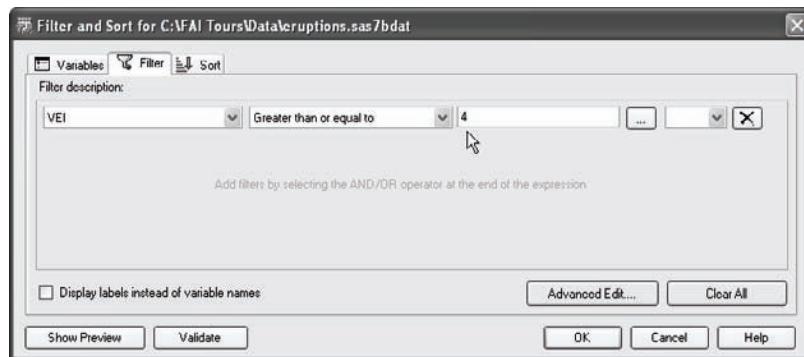
or double minus arrow .

In this example, all variables are selected except EndDate.



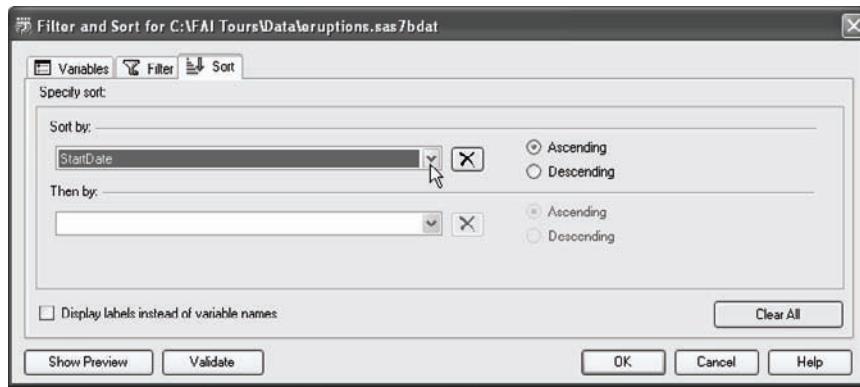
Specifying the filter

Click the **Filter** tab. The area labeled **Filter description** contains four boxes. In the first box, choose the variable for the filter. In the second



box, choose an operator. Then in the third box, enter the value for the filter. To get a list of all possible values for the selected variable, click the ellipsis button . If you want to add another condition to the filter, select AND or OR from the drop-down list in the fourth box. A new row of boxes will appear where you can specify the additional condition. For more complicated filters containing functions or complex logic, click the **Advanced Edit** button to open the Advanced Filter Builder window. In this example the filter selects all eruptions with a volcanic explosivity index (VEI) greater than or equal to 4.

Sorting the data Click the **Sort** tab. Choose a variable from the **Sort by** drop-down list. If you wish, you can choose a second variable from the **Then by** drop-down



list. Each time you choose an additional variable, options will appear allowing you to choose more sorting variables. The default sort order is Ascending. Click **Descending** to reverse the order. In this example, the data will be sorted by StartDate in ascending order. When you are ready to run the task, click **OK**.

	Volcano	StartDate	VEI
1	Vesuvius	12/15/1631	5
2	Santorini	09/27/1650	4
3	Fuji	12/16/1707	5
4	Etna	06/04/1787	4
5	Grimesvötn	01/08/1873	4
6	Vesuvius	12/18/1875	4
7	Krakatau	05/20/1883	6
8	Kliuchevskoi	03/25/1931	4
9	St. Helens	03/27/1980	5
10	Pinatubo	04/02/1991	5
11	Reventador	11/03/2002	4

Results Here is the result of the filter and sort with only the eruptions with VEI greater or equal to 4 and sorted by StartDate. The new data table is given a name starting with the letters FILTER_FOR and is stored in a default location. You cannot change the name or location of data tables created by this task, but you can export the table using any name and location you wish. See section 2.10 for more about exporting data tables.

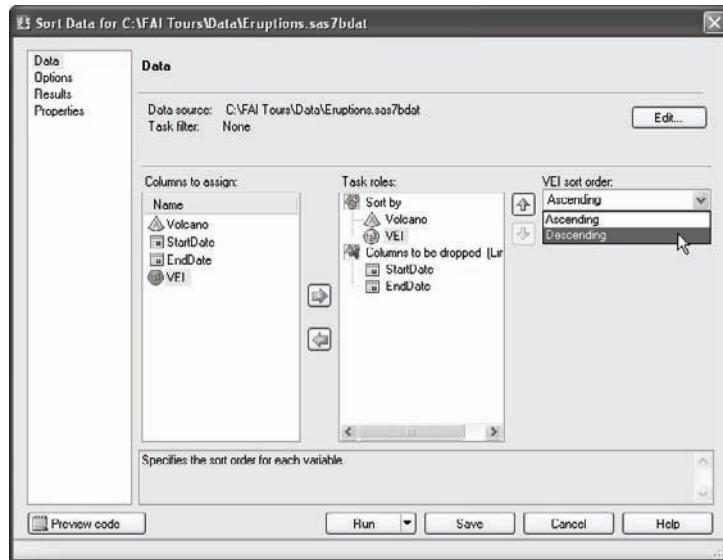
5.3 Sorting Using the Sort Task

There is little need for you to sort data in SAS Enterprise Guide. If a task requires data to be sorted, SAS Enterprise Guide will usually sort the data automatically. However, there may be times when you want to sort the data yourself. If you have a large data table, for example, you may want to store the data in sorted order. With the data presorted, SAS Enterprise Guide will not have to sort the data and your tasks will run more quickly. At other times, you may want to sort data to make it easier to find values in a Data Grid, or to eliminate duplicate rows.

There are three ways to sort data in SAS Enterprise Guide: the Filter and Sort task, the Sort task, or a query. The Sort task gives you the most control over how the data are sorted, while the Filter and Sort task and a query have other functions in addition to sorting. This section discusses the Sort task.

Here is a portion of the Eruptions data table. To sort the data using the Sort Data task, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Sort Data** from the menu bar. The Sort Data window will open.

	Volcano	StartDate	EndDate	VEI
1	Barren Island	12/20/1795	12/21/1795	2
2	Barren Island	12/20/1994	06/05/1995	2
3	Erebus	12/12/1912	.	2
4	Erebus	01/03/1972	.	1
5	Etna	02/06/1610	08/15/1610	2
6	Etna	06/04/1787	08/11/1787	4
7	Etna	01/30/1865	06/28/1865	2
8	Etna	12/16/2005	12/22/2005	1
9	Fiji	12/16/1707	02/24/1708	5
10	Grimsvotn	10/31/1603	11/01/1603	2
11	Grimsvotn	01/08/1873	08/01/1873	4
12	Grimsvotn	12/18/1998	12/28/1998	3
13	Kilauea	05/30/1840	06/25/1840	0
14	Kilauea	05/24/1969	07/22/1974	0
15	Kliuchevskoi	09/25/1737	11/04/1737	2
16	Kliuchevskoi	03/25/1931	03/27/1931	4
17	Kliuchevskoi	01/20/2005	04/07/2005	2
18	Krakatau	10/20/1883	10/21/1883	6



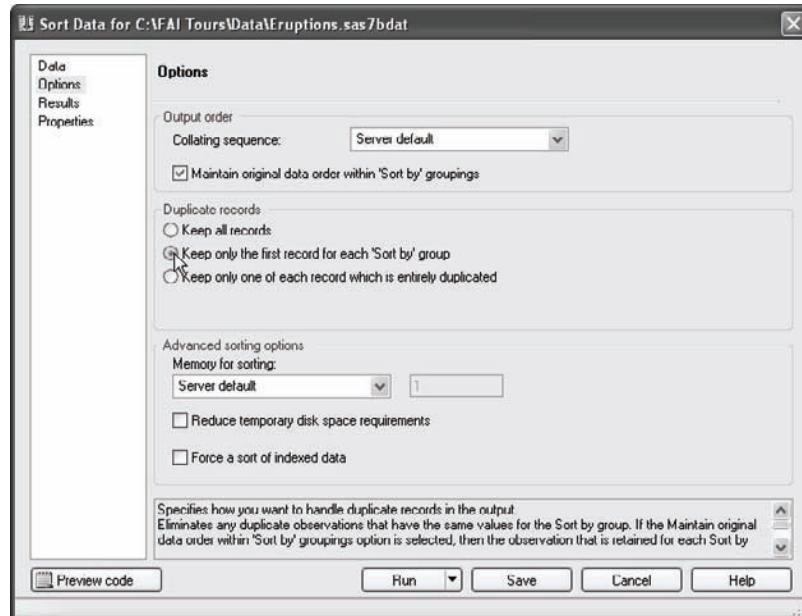
Assigning task roles

Drag the columns you want to sort by to the **Sort by** role. If there is more than one **Sort by** column, SAS Enterprise Guide will sort rows by the first column, then by the second column within values of the first column, and so on. When you assign a column to the **Sort by** role, a box for the sort order will appear on the right. If you click the down-arrow, you can choose to have the data sorted in ascending (the default) or descending order. If you want to exclude columns from the

output, drag those columns to the **Columns to be dropped** role. In this example, the Eruptions data table will be sorted first by Volcano in ascending order, then by VEI in descending order. The StartDate and EndDate columns will be dropped.

Sorting options To open the Options page, click **Options** in the selection pane on the left. In the Options page, you can select the collating sequence for the sort (such as ASCII, EBCDIC, or Server default). The collating sequence determines the sort order, including whether letters come before numerals or numerals before letters. When SAS Enterprise Guide sorts data, missing values are always lowest in the sort order. In the ASCII collating sequence (the default for Windows computers), uppercase letters come before lowercase letters.

You can also choose options for duplicate records. You can keep all records (the default), keep only the first record for each combination of values of the Sort by columns, or keep only one of each record that is entirely duplicated. Duplicate records must be consecutive to be eliminated, so if you choose the last option, it is a good idea to sort by all the columns in the data table. For this example, **Keep only the first record for each 'Sort by' group** is selected so that the result will have only one record for each Volcano and value of VEI.



Results The Sort task creates a new SAS data table and displays it in a Data Grid. Notice that this table is sorted first by Volcano, and then by descending values of VEI within Volcano. The resulting data table has fewer rows than the original because it contains only one row for each combination of the two sort columns. SAS Enterprise Guide stores the table with a default name in a default location. To save the data with a different name or location, use the Results page in the Sort Data task. In the Results page, you can also choose to save in a separate table any duplicates that were eliminated by the sort.

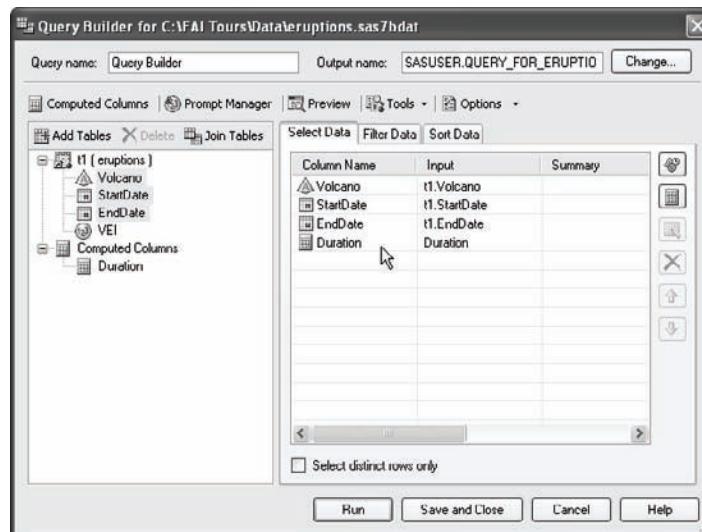
	Volcano	VEI
1	Barren Island	2
2	Erebus	2
3	Erebus	1
4	Etna	4
5	Etna	2
6	Etna	1
7	Fiji	5
8	Grimsvotn	4
9	Grimsvotn	3
10	Grimsvotn	2
11	Kilauea	0
12	Kliuchevskoi	4
13	Kliuchevskoi	2
14	Krakatau	6

5.4 Sorting Data in a Query

If all you want to do is create a sorted version of a data table, then you might want to use the Sort Data, or Filter and Sort task. But, if you also want to create new columns or join tables together, then use the Query Builder.

This example uses the Eruptions data table, which is sorted by the name of the volcano. To change the sort order using a query, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar. The Query Builder window will open with the **Select Data** tab on top.

Selecting the data For all queries you need to select the columns that will be in the result. Click and drag to the **Select Data** tab the columns you want. In this example, all columns in the Eruptions table except VEI have been selected, as well as the computed column Duration. (Section 4.3 shows how to create the Duration column.)



Here is a preview of the data after creating the computed column and before sorting. To preview results, click the **Preview** button in the Query Builder window. Then click the **Results** tab in the Preview window.

Volcano	StartDate	EndDate	Duration
Barron Island	12/20/1795	12/21/1795	2
Barren Island	12/20/1994	06/05/1995	168
Erebus	12/12/1912	.	.
Erebus	01/03/1972	.	.
Etna	02/06/1610	08/15/1610	191
Etna	06/04/1787	08/11/1787	69
Etna	01/30/1865	06/28/1865	150
Etna	12/16/2005	12/22/2005	7
Fuji	12/16/1707	02/24/1708	71
Grimsvotn	10/31/1603	11/01/1603	2

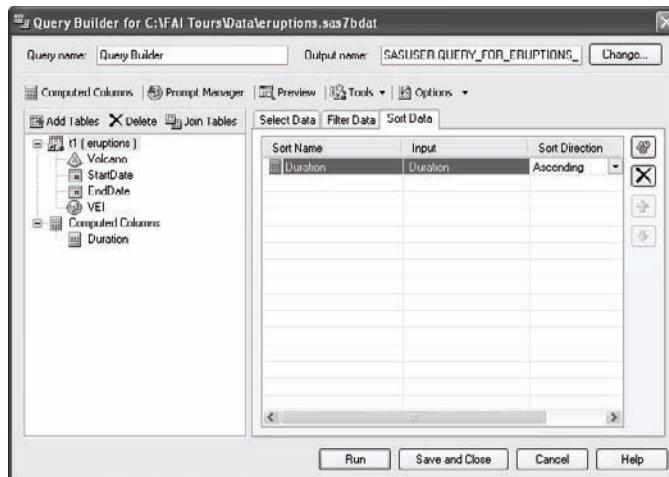
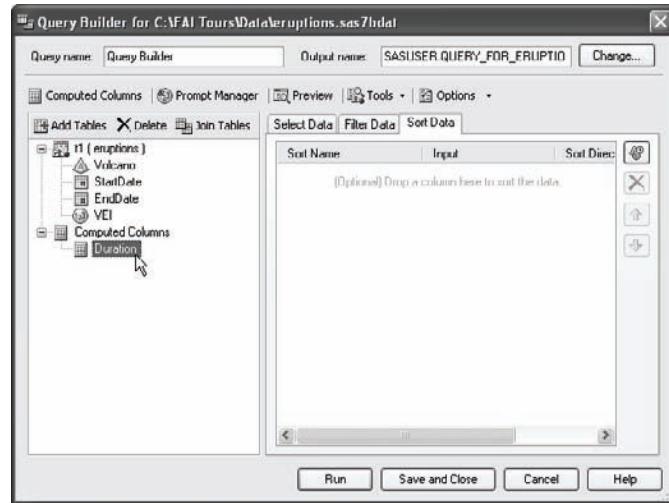
Sorting the data Click the **Sort Data** tab. To sort the data, click the desired column in the list on the left and drag it to the **Sort Data** tab. You can sort by more than one column by dragging multiple columns to the **Sort Data** tab. Notice that even columns not selected for the query (VEI for this example) are listed. You can use columns for sorting even if they don't appear in the result.

If you choose more than one column for sorting, the order of the columns on the **Sort Data** tab will determine how the data are sorted. The data will be sorted by the first column in the list. Then, within unique values of the first column, the data will be sorted by the second column. You can change the sort order by clicking the column name in the **Sort Data** tab and clicking the up- or down-

arrow buttons   to move the columns. To change the sort direction, click the column name on the **Sort Data** tab. Then click the down-arrow ▼ next to the sort direction for that column, and select either Ascending or Descending from the drop-down list. In this query, the data will be sorted by the computed column Duration in ascending order.

Click **Run** in the Query Builder window to create a sorted data table.

Results Here is a portion of the resulting data table sorted by Duration. Missing values are always lowest in the sort order, so they appear first.



	Volcano	StartDate	EndDate	Duration
1	Erebus	01/03/1972	.	.
2	Erebus	12/12/1912	.	.
3	Villarrica	10/26/2008	.	.
4	Poas	04/08/1996	04/08/1996	1
5	Sabancaya	05/01/1997	05/02/1997	2
6	Reventador	12/12/1856	12/13/1856	2
7	Barren Island	12/20/1795	12/21/1795	2
8	Grimsvoth	10/31/1603	11/01/1603	2
9	Kluchevskoi	03/25/1931	03/27/1931	3
10	St. Helens	03/26/1847	03/30/1847	5
11	Reventador	02/24/1944	03/01/1944	7
12	Etna	12/16/2005	12/22/2005	7
13	Grimsvoth	12/18/1998	12/28/1998	11

5.5 Filtering Data in a Query

If you want to create a data table that is a subset of another table, you can use the Filter and Sort task or the Query Builder. The Filter and Sort task only filters and sorts data, but the Query Builder can also create computed or summarized columns and join tables together.

Here is a sample of the Volcanoes data table. To filter the data so that only volcanoes in North America and South America appear in the results, click the data icon in the Project Tree or Process Flow and select **Tasks ▶ Data ▶ Query**

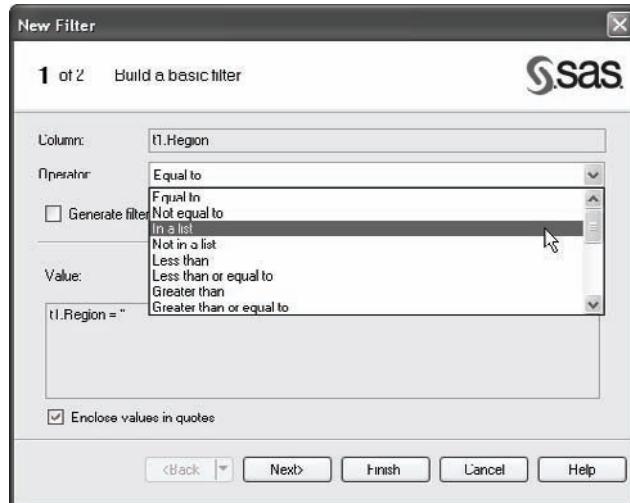
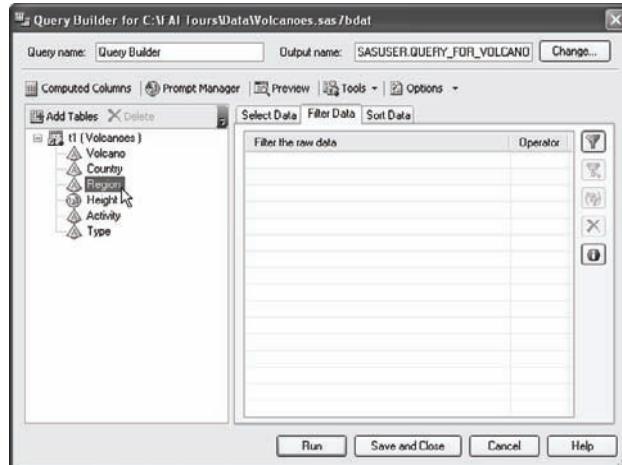
Builder from the menu bar. This opens the Query Builder window, with the **Select Data** tab on top.

Select the columns that you want in the query result (in this example every column except Type). Then click the **Filter Data** tab.

Select the column that you want to use for filtering purposes and drag it to the **Filter Data** tab. If you want to filter based on a summarized column, drag it to the **Filter the summarized data** area of the **Filter Data** tab, which only appears when you have summarized columns in the query. Notice that you can filter by columns that are not even part of the query result. In this example, drag **Region** to the filter area. When you drop the column, the New Filter wizard will open.

Building the filter In the New Filter wizard, you can see the name of the column you dragged to the filter area. Initially, the operator is **Equal to**. To choose a different operator, click the down-arrow to the right of the **Operator** box. In this example, the **In a list** operator is the most useful because it allows you to specify a list of unique values for Region.

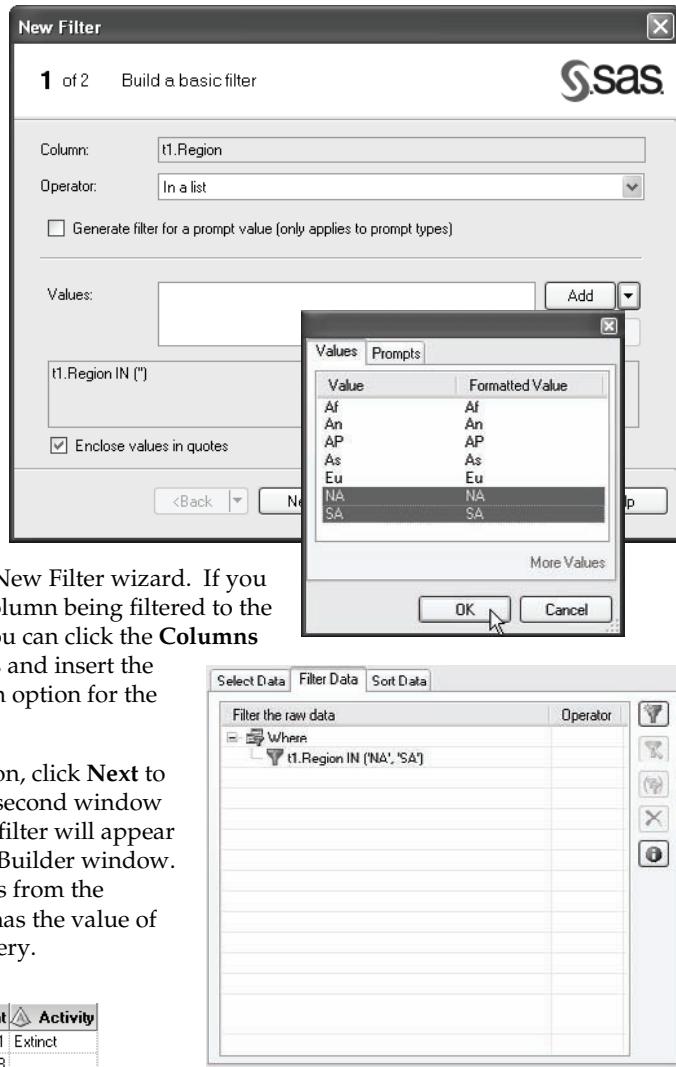
	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	Stratovolcano
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano



After choosing an operator, choose a value or values for your filter condition. You can type the values in the **Values** box or let SAS Enterprise Guide help you select the values. Click the down-arrow next to the **Values** box. In the window that opens, click **Get Values** and you will see a list of all possible values for the column you selected. Note that if you have large data tables, or lots of possible values, it can take a very long time to retrieve the values. Highlight one or more values by holding down the control (CTRL) key as you select values. In this example, select **NA** and **SA**, and then click **OK**. The values will be inserted into the **Values** box of the New Filter wizard. If you want to compare the value of the column being filtered to the value in a different column, then you can click the **Columns** tab to get a list of available columns and insert the column into the filter. This is not an option for the **In a list** operator.

Once you have set the filter condition, click **Next** to see a summary of your filter in the second window (not shown), then click **Finish**. The filter will appear on the **Filter Data** tab of the Query Builder window. In this case, the filter selects all rows from the Volcanoes table (t1) where Region has the value of NA or SA. Click **Run** to run the query.

	Volcano	Country	Region	Height	Activity
1	Altar	Ecuador	SA	5321	Extinct
2	Garibaldi	Canada	NA	2678	
3	Illimani	Bolivia	SA	6458	Extinct
4	Lassen	USA	NA	3187	Active
5	Poas	Costa Rica	NA	2708	Active
6	Popocatepetl	Mexico	NA	5426	Active
7	Reventador	Ecuador	SA	3562	Active
8	Sabancaya	Peru	SA	5976	Active
9	Shishaldin	USA	NA	2857	Active
10	St. Helens	USA	NA	2549	Active
11	Villarica	Chile	SA	2847	Active



Results Here is the data table produced by the query. Only the volcanoes from North and South America are included.

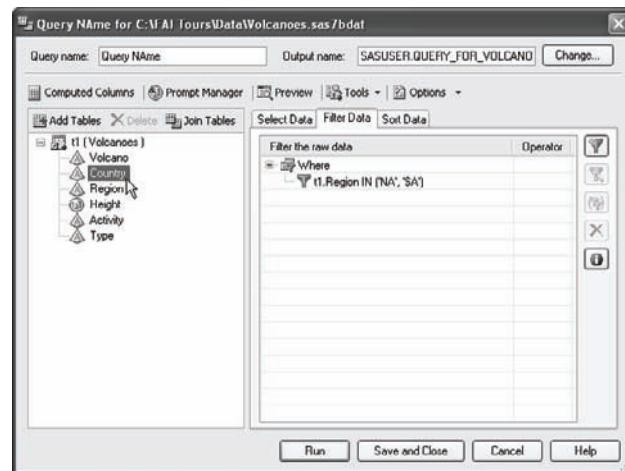
5.6 Creating Compound Filters in a Query

Sometimes you want to base a filter on more than one condition. You can add conditions to a filter using the AND and OR operators. If you use AND, then rows must meet both conditions of the filter to be included in the result. If you use OR, then rows need to meet only one condition or the other. This section shows how to create a compound filter in the Query Builder, but you can create them anywhere you can create a filter.

In the previous section, the Volcanoes data were filtered so that only the volcanoes in North America and South America appeared in the result. Suppose you also want volcanoes that are in Japan. For this result, you need to select rows where the region is North America or South America, or where the country is Japan. To modify the existing query, click **Modify Task** on the workspace toolbar for the query result. In the Query Builder window, click the **Filter Data** tab.

Adding conditions to a filter

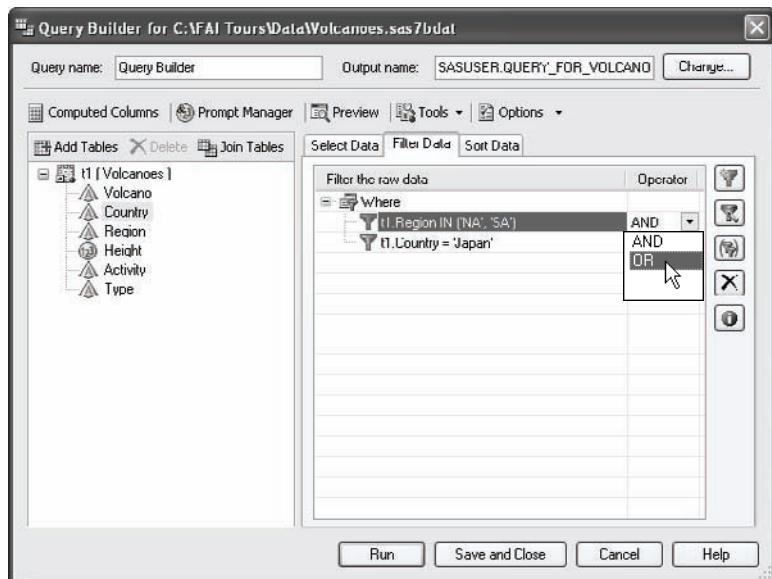
To add a condition to an existing filter, drag the column for the new condition from the column list to the **Filter Data** tab. For this example, drag the **Country** column to the **Filter Data** tab. When you drop the column, the New Filter wizard will open.



Building the Filter The procedure for specifying the additional condition is the same as if you were creating a single condition filter. First, choose an appropriate operator for your condition, and then either type the desired value in the Value box, or click the down-arrow next to the Value box to get a list of values to choose from. In this example, the filter will choose rows where the Country column from the Volcanoes table (t1) is equal to the value **Japan**. When you are satisfied, click **Next** to see a summary of your filter, then click **Finish**.

Setting the logic

After you create the additional condition, it will be added to the existing filter condition on the **Filter Data** tab. The filter now has two conditions. When you add new conditions to your filter, SAS Enterprise Guide automatically chooses the AND operator. To change to OR, click the down-arrow next to the AND operator and choose OR. For this example, set the operator to OR. Click **Run** to run the query.



	Volcano	Country	Region	Height	Activity
1	Altar	Ecuador	SA	5321	Extinct
2	Fuji	Japan	As	3776	Active
3	Garibaldi	Canada	NA	2678	
4	Illimani	Bolivia	SA	6458	Extinct
5	Lassen	USA	NA	3187	Active
6	Poas	Costa Rica	NA	2708	Active
7	Popocatepetl	Mexico	NA	5426	Active
8	Reventador	Ecuador	SA	3562	Active
9	Sabancaya	Peru	SA	5976	Active
10	Shishaldin	USA	NA	2857	Active
11	St. Helens	USA	NA	2549	Active
12	Villarrica	Chile	SA	2847	Active

Results Here is the new data table produced after filtering the Volcanoes data table. Now, all the volcanoes from North America and South America are included, as well as all the volcanoes in Japan.

5.7 Creating a Filter with Advanced Expressions

You can accomplish a lot with basic filters, but sometimes you need more advanced filters containing functions or more complex logic. For these cases, you can create an advanced filter. Advanced filters can be created in the Query Builder or the Filter and Sort task.

Here is a portion of the Bookings data table. This example uses the SUBSTR function in the Query Builder to select rows where the CustomerID starts with the letters "DE." To create a query, click the data icon in the

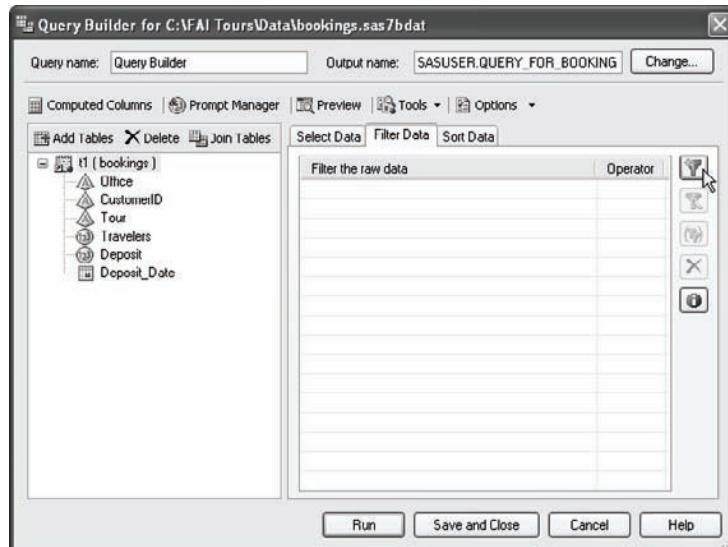
Project Tree or Process Flow and select **Tasks ▶ Data ▶ Query Builder** from the menu bar. This opens the Query Builder with the **Select Data** tab on top. Select the columns that you want in the query result. Then click the **Filter Data** tab. For this example, select all columns.

Opening the advanced filter window

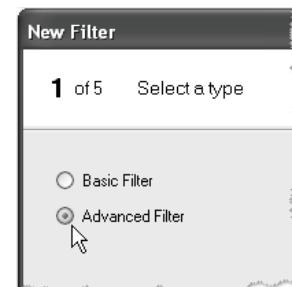
On the Filter Data tab of the Query Builder window, click the New Filter

icon  to open the New Filter wizard.

	Office	CustomerID	Tour	Travelers	Deposit	Deposit_Date
1	Portland	SL28	SH43	10	425	05JUL2011
2	Portland	DE27	PS27	6	75	11JUL2011
3	Portland	SL34	FJ12	4	200	19JUL2011
4	Portland	DI33	SH43	4	150	23JUL2011
5	Portland	BU12	SH43	2	75	23JUL2011
6	Portland	DE31	FJ12	3	175	25JUL2011
7	Portland	WI48	FJ12	2	100	26JUL2011
8	Portland	MG17	PS27	5	65	26JUL2011



The New Filter wizard has up to five windows depending on which type of filter you select. In the first window, select **Advanced Filter** and click **Next**. (In the Filter and Sort task, click the **Advanced Edit** button on the **Filter** tab.)

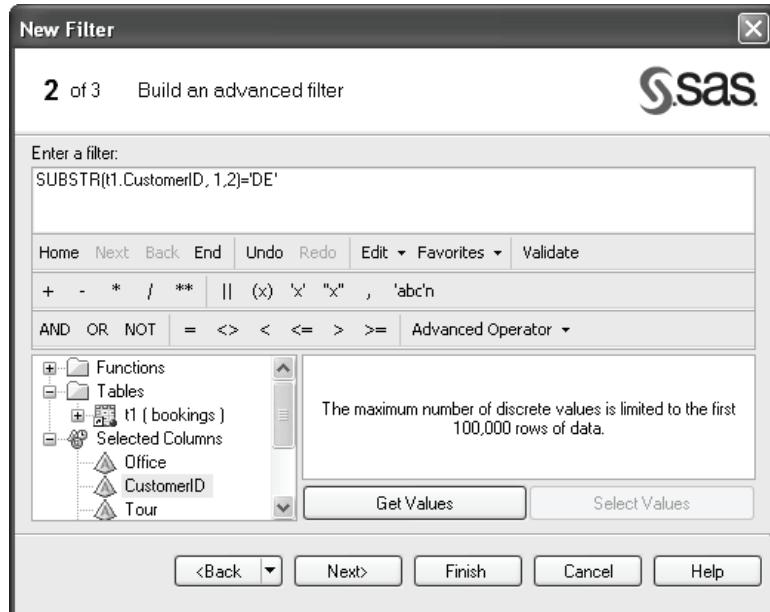


Building the advanced filter

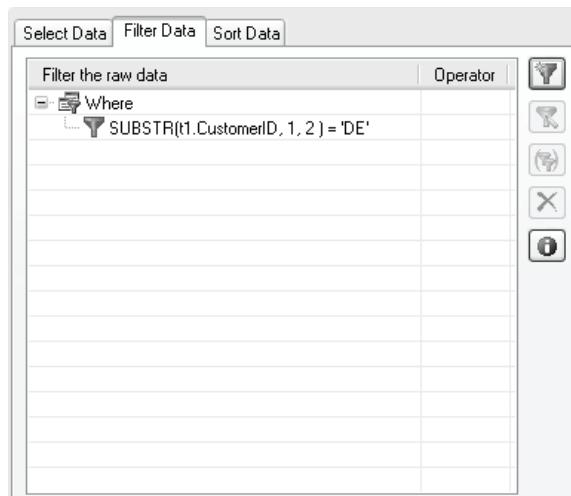
In the second window, type the desired filter expression in the text box labeled **Enter a filter**. If you want to use functions in the expression, you can expand the Functions node in the box on the lower right to view all available functions.

All the columns in the table, along with the selected columns for the query, are also listed in this box. In this example, the SUBSTR function is used to select all rows

where the first two characters of the CustomerID column from the Bookings data table (t1) are equal to "DE." When you are finished building your filter, click **Next** to see a summary of your filter in the third window (not shown), then click **Finish** to return to the Query Builder window.



Your filter will appear on the **Filter Data** tab of the Query Builder. Click **Run** to run the query and view the results.



Results Here is the new data table produced after filtering the Bookings data table. Only the rows with a CustomerID that starts with the letters "DE" appear in the result.

	Office	CustomerID	Tour	Travelers	Deposit	Deposit_Date
1	Portland	DE27	PS27	6	75	11JUL2011
2	Portland	DE31	FJ12	3	175	25JUL2011

6

“ Learning without thought
is labor lost; thought without
learning is perilous. ”

CONFUCIUS



CHAPTER 6

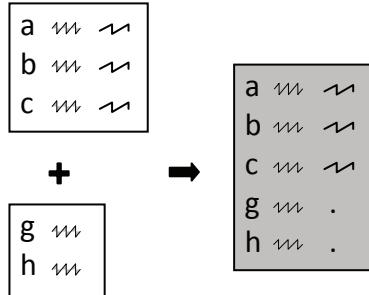
Combining Data Tables

- 6.1 Methods for Combining Tables 240
- 6.2 Appending Tables 242
- 6.3 Joining Tables 244
- 6.4 Setting the Properties of a Join 246

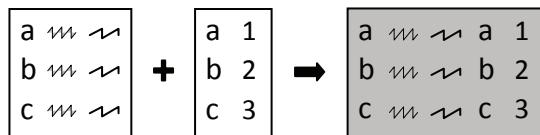
6.1 Methods for Combining Tables

In SAS Enterprise Guide, there are two basic ways to combine data tables: appending and joining. You append when the tables contain the same (or almost the same) columns. You join when the tables contain the same (or almost the same) rows. This section describes, in general terms, what happens when you combine tables.

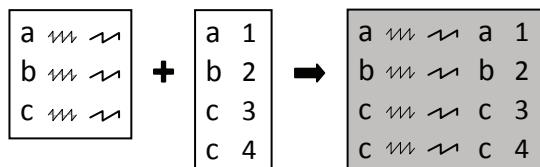
Appending tables Appending tables is like stacking them. It only makes sense to append tables if they have columns in common. Appending tables is done with the Append Table task, which is described in the next section.



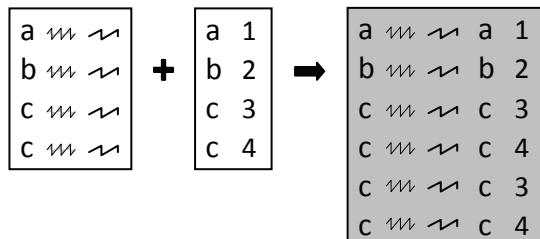
Joining tables To join tables together, the tables need to have a column (or set of columns) that can be used to match rows. You do not need to specify whether the match is one-to-one, one-to-many, or many-to-many—SAS Enterprise Guide determines this automatically. Joining tables is done in the Query Builder, which is described in section 6.3.



One-to-one match In a one-to-one match, one row from the first table is joined with one row from the second table. The values for the common column are unique in both tables.



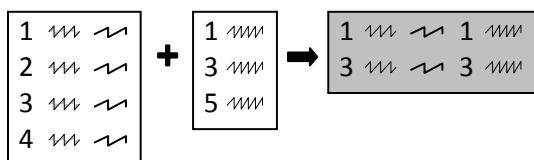
One-to-many match In a one-to-many match, one row from the first table is joined with multiple rows from the second table. The values for the common column are unique in only one table.



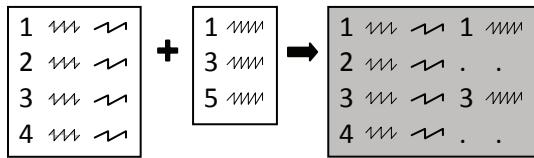
Many-to-many match In a many-to-many match, the values for the common column are not unique in either table. Each row in the first table is joined with all the matching rows from the second table.

When you join tables together, one table may have rows that do not match any rows contained in the other table. By default, any non-matching rows are deleted. In the Query Builder, you can control which rows end up in the new table based on which table they came from. The Query Builder illustrates this with a join indicator (also known as a Venn diagram). The join indicator consists of two circles representing the two tables. The shaded areas show the parts of the tables that will be kept. You can choose different kinds of joins including inner joins, left joins, right joins, and full outer joins. Controlling the type of join is discussed in section 6.4.

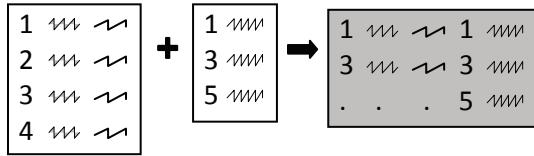
The following graphics all show a one-to-one join, but the join has been modified so that different rows are included in the results.



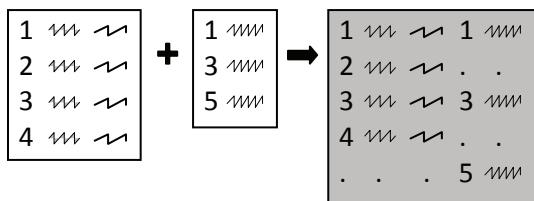
Inner joins An inner join , the default, keeps only rows that match.



Left joins In a left join , all rows from the table on the left are kept even if they do not have a match in the other table.



Right joins In a right join , all rows from the table on the right are kept even if they do not have a match in the other table.



Full outer joins In a full outer join , all rows are kept from both tables.

6.2 Appending Tables

You use the Append Table task to combine tables that contain the same (or almost the same) columns. For example, if you had sales data for January, February, and March in three separate tables, you could append the tables to create one table for the entire quarter.

In this example, a customer living in southern Washington is interested in traveling with the Fire and Ice Tours company. Because this customer lives between Seattle and Portland, she wants to see prices for flights from each city.

Here are two data tables, one showing flights from Portland and the other from Seattle.

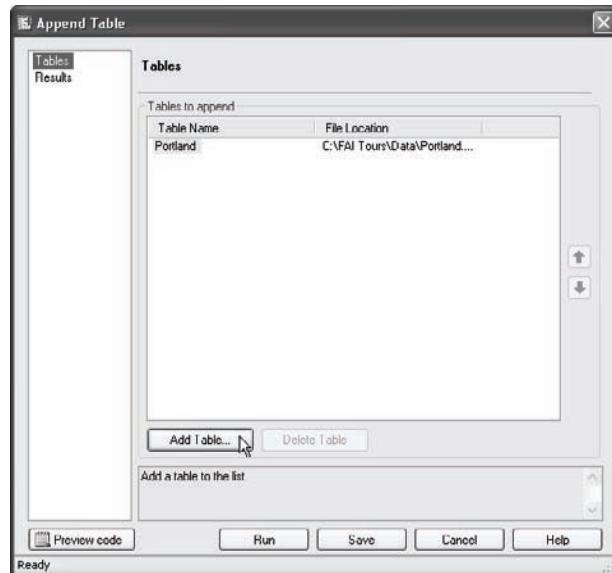
Looking at these two data tables, you can see they contain the same columns, making them good candidates for appending.

	Origin	Destination	FlightNo	FlightPrice
1	Portland	Catania	L469	\$779.00
2	Portland	Hilo	HA25	\$703.00
3	Portland	Nairobi	KLM6034	\$1,833.00
4	Portland	Rome	D1576	\$644.00
5	Portland	San Jose	CA1210	\$494.00
6	Portland	Tokyo	UA383	\$705.00

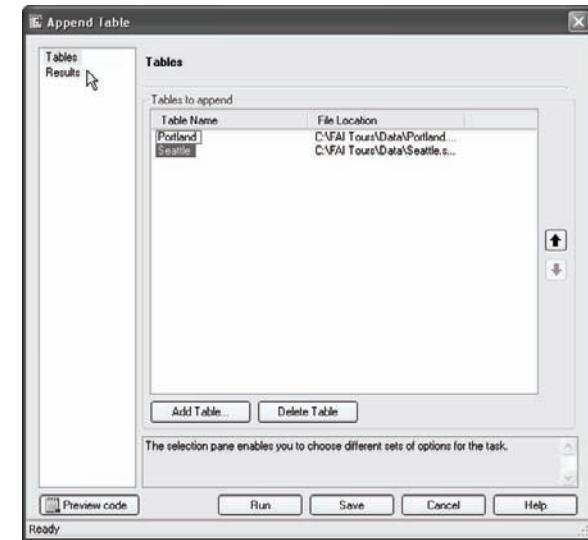
	Origin	Destination	FlightNo	FlightPrice
1	Seattle	Catania	BA48	\$802.00
2	Seattle	Hilo	HA21	\$677.00
3	Seattle	Jakarta	AA119	\$1,815.00
4	Seattle	Nairobi	KLM6034	\$1,761.00
5	Seattle	Quito	CA1086	\$833.00
6	Seattle	Rome	USA6	\$596.00
7	Seattle	San Jose	CA1100	\$480.00
8	Seattle	Tokyo	UA875	\$721.00

To append tables, click one data icon in the Project Tree or Process Flow to make it active, and then select **Tasks ▶ Data ▶ Append Table** from the menu bar. The Append Table window will open.

Adding tables The Append Table window opens showing the active table. To add a table, click **Add Table**, navigate to the table you want to add, and click **OK**. You can append up to 32 tables at once.

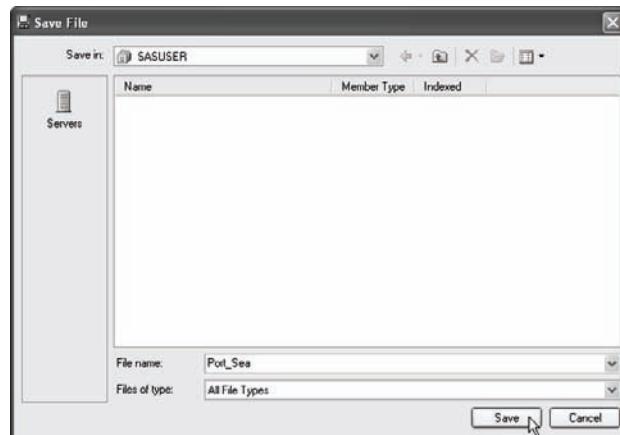


Running the task At this point, you can click **Run** and SAS Enterprise Guide will create the new table, store it in a default location, and give it the name **Append_Table**. (If you have more than one appended table stored in that location, SAS Enterprise Guide will add numbers to the name.) To choose a different name or location, click the **Results** option in the selection pane on the left. Then click **Browse** in the Results page (not shown) to open the Save File window.



In the Save File window, type a name for the new data table in the **File name** box and choose a library. To see the available libraries, click the down arrow in the **Save in** box at the top of the window. In this example, the new data table will be named **Port_Sea** and will be saved in the **SASUSER** library. Once you have specified the library and filename, click **Save**.

In the Append Table window, click **Run**. SAS Enterprise Guide will display the results in a Data Grid.



Results Here is the new data table. Notice that SAS Enterprise Guide concatenated the Portland and Seattle tables by matching the columns.

In this case, the two tables contained exactly the same columns. If there had been a column that existed in one table but not the other, then the data values for that column would be set to missing for rows from the other table.

	Origin	Destination	FlightNo	FlightPrice
1	Portland	Catania	L469	\$779.00
2	Portland	Hilo	HA25	\$703.00
3	Portland	Nairobi	KLM6034	\$1,833.00
4	Portland	Rome	D1576	\$644.00
5	Portland	San Jose	CA1210	\$494.00
6	Portland	Tokyo	UA383	\$705.00
7	Seattle	Catania	B448	\$802.00
8	Seattle	Hilo	HA21	\$677.00
9	Seattle	Jakarta	AA119	\$1,815.00
10	Seattle	Nairobi	KLM6034	\$1,761.00
11	Seattle	Quito	CA1086	\$833.00
12	Seattle	Rome	USA6	\$596.00
13	Seattle	San Jose	CA1100	\$480.00
14	Seattle	Tokyo	UA875	\$721.00

6.3 Joining Tables

When you append tables, you match columns. But often, instead of matching columns, you need to match rows. For example, a teacher might record grades from homework in one table and grades from tests in another. To compute final grades she would need to match the homework and test scores for each student. This is called joining tables, and you do it with a query.

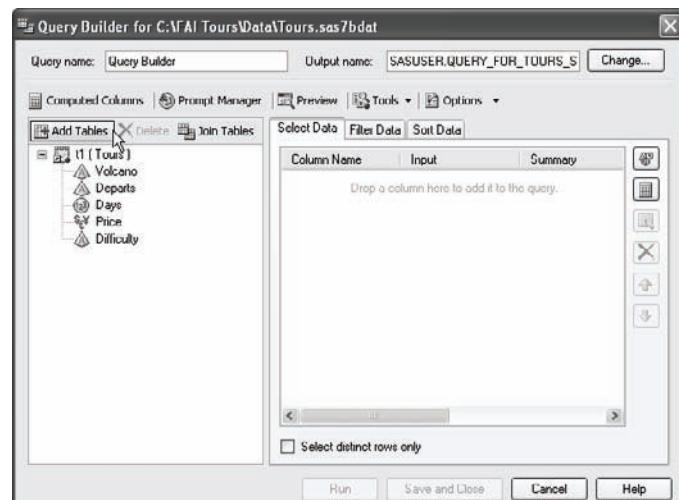
In the preceding section, two tables were appended to create one table containing all the data about flights. Now the data for each tour can be joined with the matching data for flights. To open the Query Builder, click a data icon in the Project Tree or Process Flow, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar.

	Volcano	Departs	Days	Price	Difficulty
1	Etna	Catania	7	\$1,075	m
2	Fuji	Tokyo	2	\$225	c
3	Kenya	Nairobi	6	\$830	m
4	Kilauea	Hilo	1	\$55	e
5	Kilimanjaro	Nairobi	9	\$1,310	c
6	Krakatau	Jakarta	7	\$895	e
7	Poas	San Jose	1	\$65	e
8	Reventador	Quito	4	\$575	m
9	St. Helens	Portland	2	\$167	e
10	Vesuvius	Rome	6	\$985	e

	Origin	Destination	FlightNo	FlightPrice
1	Portland	Catania	L469	\$779.00
2	Portland	Hilo	HA25	\$703.00
3	Portland	Nairobi	KLM6034	\$1,833.00
4	Portland	Rome	D1576	\$644.00
5	Portland	San Jose	CA1210	\$494.00
6	Portland	Tokyo	UA383	\$705.00
7	Seattle	Catania	BA48	\$802.00
8	Seattle	Hilo	HA21	\$677.00
9	Seattle	Jakarta	AA119	\$1,815.00
10	Seattle	Nairobi	KLM6034	\$1,761.00
11	Seattle	Quito	CA1086	\$833.00
12	Seattle	Rome	USA6	\$596.00
13	Seattle	San Jose	CA1100	\$480.00
14	Seattle	Tokyo	UA875	\$721.00

Adding tables When you open the Query Builder, it will show the active table. To open another table, click the **Add Tables** button, and navigate to the table you want to add. For this example, start with the Tours table, and add the Port_Sea table created in the previous section. You can join up to 32 tables at once.

When you add tables, SAS Enterprise Guide will automatically look for columns with the same name and type. If SAS Enterprise Guide does not find any columns with the same name and type, then a warning message will appear telling you to join the columns manually. Click **OK**.



To join two tables manually, click the first table, then right-click the name of one column and select the name of the matching column from the pop-up menu. SAS Enterprise Guide will draw a line from one column to the other. To correctly match the Tours data table and the Port_Sea table, the destination of a flight must match the city from which a tour departs. To join the tables in this example, click the Tours table, and then right-click the column Departs, and select t2 and Destination from the pop-up menu. A Join Properties window will open (not shown). For this example, accept the default settings by clicking OK. In the Tables and Joins window, click Close.

You can use more than one column for matching in a join.

For example, a teacher might combine homework and test scores based on both class and student ID number. To specify additional columns for matching, select them manually as described above.

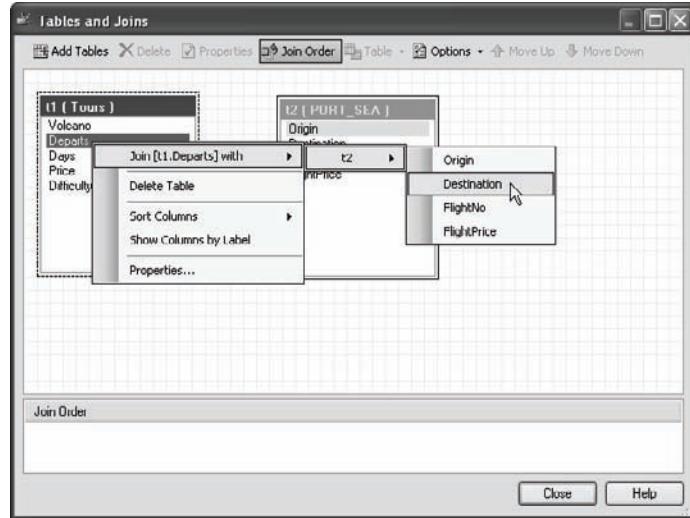
Running the query If you have not already selected the columns to be included in the results, then drag those columns to the **Select Data** tab in the Query Builder. When you are satisfied, click **Run**. SAS Enterprise Guide will display the results in a Data Grid.

Results This Data Grid shows the result of joining the two data tables. Notice that both of the columns used for matching (Departs and Destination) appear in the results. If these columns had the same name (such as City), then a number would have been automatically

added to the name of the second column (resulting in City and City1).

Because there are two tours departing from Nairobi, and two flights with a destination of Nairobi, this is a many-to-many join.

SAS Enterprise Guide kept only the rows that matched. This is an inner join, which is the default type of join. To keep rows that don't match, modify your join as described in the next section.



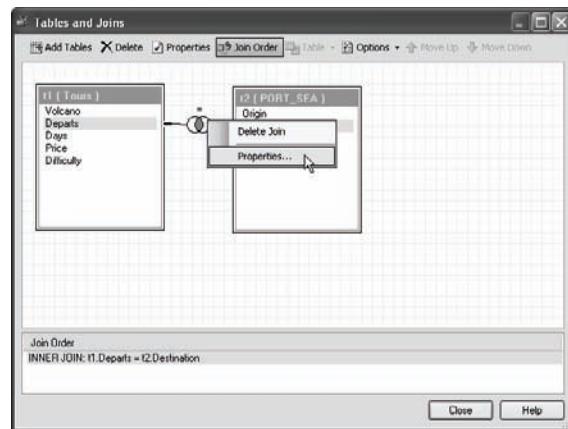
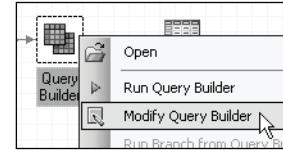
	Volcano	Departs	Days	Price	Difficulty	Origin	Destination	FlightNo	FlightPrice
1	Etna	Catania	7	\$1,075 m	m	Portland	Catania	L469	\$779.00
2	Etna	Catania	7	\$1,075 m	m	Seattle	Catania	BA48	\$802.00
3	Fuji	Tokyo	2	\$225 c	c	Portland	Tokyo	UA383	\$705.00
4	Fuji	Tokyo	2	\$225 c	c	Seattle	Tokyo	UA875	\$721.00
5	Kenya	Nairobi	6	\$830 m	m	Portland	Nairobi	KLM6034	\$1,833.00
6	Kenya	Nairobi	6	\$830 m	m	Seattle	Nairobi	KLM6034	\$1,761.00
7	Kilauea	Hilo	1	\$55 e	e	Portland	Hilo	HA25	\$703.00
8	Kilauea	Hilo	1	\$55 e	e	Seattle	Hilo	HA21	\$677.00
9	Kilimanjaro	Nairobi	9	\$1,310 c	c	Portland	Nairobi	KLM6034	\$1,833.00
10	Kilimanjaro	Nairobi	9	\$1,310 c	c	Seattle	Nairobi	KLM6034	\$1,761.00
11	Krakatau	Jakarta	7	\$895 e	e	Seattle	Jakarta	AA119	\$1,815.00
12	Poas	San Jose	1	\$65 e	e	Portland	San Jose	CA1210	\$494.00
13	Poas	San Jose	1	\$65 e	e	Seattle	San Jose	CA1100	\$480.00
14	Reventador	Quito	4	\$575 m	m	Seattle	Quito	CA1086	\$833.00
15	Vesuvius	Rome	6	\$985 e	e	Portland	Rome	D1576	\$644.00
16	Vesuvius	Rome	6	\$985 e	e	Seattle	Rome	USA6	\$596.00

6.4 Setting the Properties of a Join

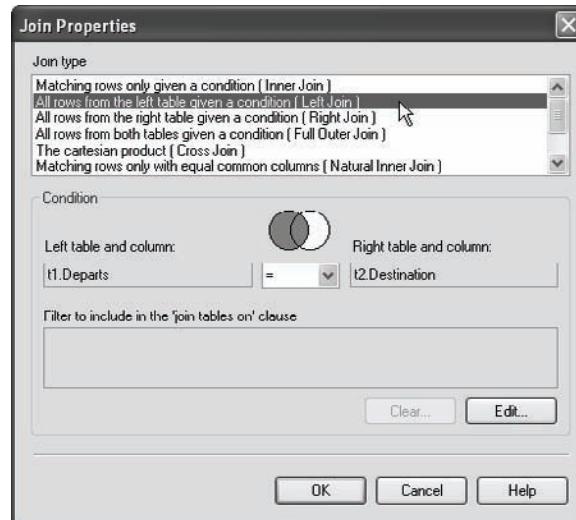
By default, when you join tables, SAS Enterprise Guide keeps only rows for which a match is found. Sometimes that may be just what you want, but at other times, you may want to keep all the rows regardless of whether they match, or all the rows from one table, but not the other. To do this, change the properties of the join.

Reopening the Query window To change a query that you have already run, right-click the query icon in the Project Tree or Process Flow and select **Modify Query Builder** from the pop-up menu. The Query Builder will open. Click the **Join Tables** button to open the Tables and Joins window.

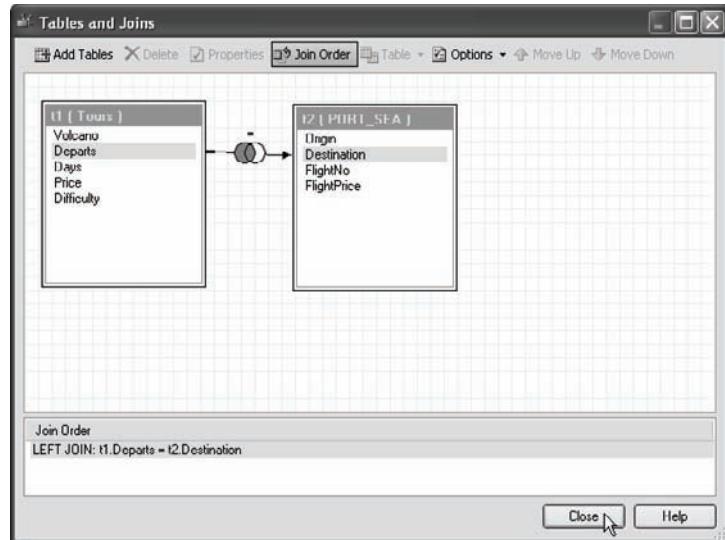
To modify a join, right-click the join indicator between the two tables and select **Properties** from the pop-up menu. The Join Properties window will open.



Selecting the type of join In the Join Properties window, you can choose from several types of joins. In this window, **All rows from the left table given a condition** has been selected. Tours is the table on the left, so all rows from Tours will be included regardless of whether there is a matching row in the Port_Sea table. This is called a left join. When you are satisfied with the join condition, click **OK**.



When you return to the Tables and Joins window, you will see that the join indicator between the two tables has changed. In this example, the circle on the left is filled in, indicating that all rows from the Tours data table will be included. When you are satisfied, click **Close**.



Running the query If you have not already selected the columns to be included in the results, then drag those columns to the **Select Data** tab in the Query Builder. When you are satisfied, click **Run**. SAS Enterprise Guide will display the results in a Data Grid.

Results This Data Grid shows the result of the modified join. This table contains all the tours, including the one for St. Helens. Notice that the values of columns from the Port_Sea table are missing for row 10. That is because there were no flights with a Destination of Portland. Because the customer lives near St. Helens, she doesn't need a flight to go on that tour.

	Volcano	Departs	Days	Price	Difficulty	Origin	Destination	FlightNo	FlightPrice
1	Etna	Catania	7	\$1,075	m	Portland	Catania	L469	\$779.00
2	Etna	Catania	7	\$1,075	m	Seattle	Catania	BA48	\$802.00
3	Kilauea	Hilo	1	\$55	e	Portland	Hilo	HA25	\$703.00
4	Kilauea	Hilo	1	\$55	e	Seattle	Hilo	HA21	\$677.00
5	Krakatau	Jakarta	7	\$895	e	Seattle	Jakarta	AA119	\$1,815.00
6	Kenya	Nairobi	6	\$830	m	Seattle	Nairobi	KLM6034	\$1,761.00
7	Kilimanjaro	Nairobi	9	\$1,310	c	Seattle	Nairobi	KLM6034	\$1,761.00
8	Kenya	Nairobi	6	\$830	m	Portland	Nairobi	KLM6034	\$1,833.00
9	Kilimanjaro	Nairobi	9	\$1,310	c	Portland	Nairobi	KLM6034	\$1,833.00
10	St. Helens	Portland	2	\$167	e				-
11	Reventador	Quito	4	\$575	m	Seattle	Quito	CA1086	\$833.00
12	Vesuvius	Rome	6	\$985	e	Seattle	Rome	USA6	\$596.00
13	Vesuvius	Rome	6	\$985	e	Portland	Rome	D1576	\$644.00
14	Poas	San Jose	1	\$65	e	Seattle	San Jose	CA1100	\$480.00
15	Poas	San Jose	1	\$65	e	Portland	San Jose	CA1210	\$494.00
16	Fuji	Tokyo	2	\$225	c	Seattle	Tokyo	UA875	\$721.00
17	Fuji	Tokyo	2	\$225	c	Portland	Tokyo	UA383	\$705.00

7

“ Celui qui a de l'imagination sans érudition a des ailes, et n'a pas de pieds. ”

“ He who has imagination without learning has wings, but no feet. ”

JOSEPH JOUBERT



CHAPTER 7

Producing Simple List and Summary Reports

- 7.1 Creating Simple List Reports 250
- 7.2 Customizing Titles and Footnotes 252
- 7.3 Adding Groups to List Reports 254
- 7.4 Adding Totals to List Reports 256
- 7.5 Creating Frequency Reports 258
- 7.6 Creating Crosstabulations 260
- 7.7 Creating Simple Summary Reports 262
- 7.8 Creating Summary Data Sets in a Task 264
- 7.9 Creating Grouped Reports with User-Defined Formats 266

7.1 Creating Simple List Reports

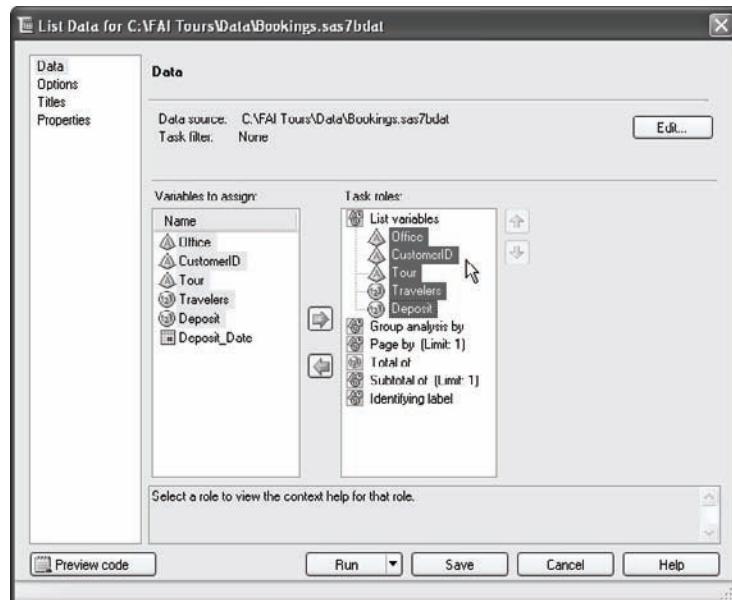
A simple list report has one line for each observation in the data set. It's the kind of report you need when you just want to see your data. You can select some variables and not others, group the data by a particular variable, and insert totals; but as long as you have one line for each observation, it is still a list report. This example creates a list report using the List Data task. You can also use the List Report wizard to create list reports.

Here is a Data Grid showing the Bookings data set used in this example. To open the List Data task, select **Tasks ▶ Describe ▶ List Data** from the menu bar. The List Data window will open, displaying the Data page.

	Office	CustomerID	Tour	Travelers	Deposit	Deposit_Date
1	Portland	SL28	SH43	10	425	05JUL2011
2	Portland	DE27	PS27	6	75	11JUL2011
3	Portland	SL34	FJ12	4	200	19JUL2011
4	Portland	DI33	SH43	4	150	23JUL2011
5	Portland	BU12	SH43	2	75	23JUL2011
6	Portland	DE31	FJ12	3	175	25JUL2011
7	Portland	W148	FJ12	2	100	26JUL2011
8	Portland	NG17	PS27	5	65	26JUL2011
9	Portland	RA28	PS27	2	30	28JUL2011
10	Portland	ME11	PS27	2	30	28JUL2011
11	Portland	GI08	SH43	8	300	31JUL2011
12	Portland	HI15	SH43	4	150	31JUL2011
13	Portland	MA09	SH43	2	75	31JUL2011

Assigning task roles

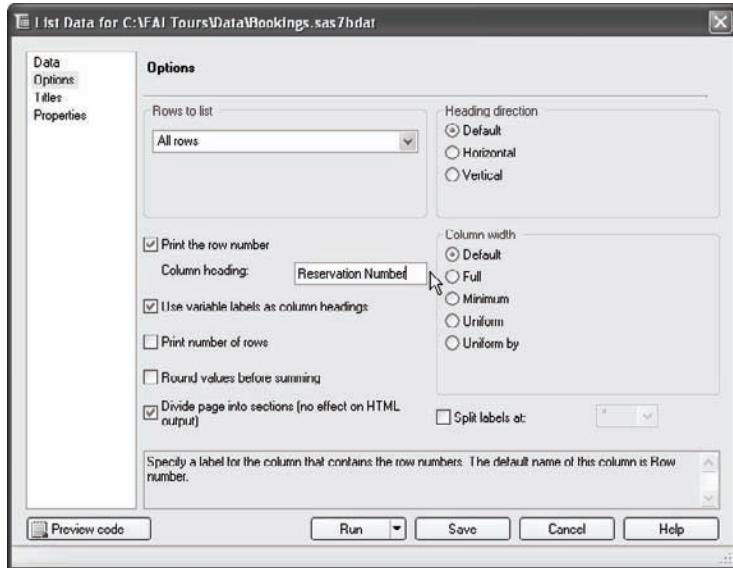
You assign variables to roles by clicking the name of a variable and dragging it to the role you want it to have. For the List Data task, you must assign at least one variable to the **List variables** role. In this window, the variables Office, CustomerID, Tour, Travelers, and Deposit have been assigned to serve as list variables.



Choosing options To further customize your report, click **Options** in the selection pane on the left. By default, the List Data task will print the row number for each line in the report. If you don't want row numbers, then uncheck **Print the row number**.

Print the row number. You can also leave the row number, but change the heading for that particular column. In this example, the heading for the row numbers has been changed to Reservation Number.

When you are satisfied with the options, click **Run**.



Results Here is the report listing the selected variables for the Bookings data set. Notice that row numbers are included and appear under the heading Reservation Number. This report uses the default title and footnote. The next section shows how to customize titles and footnotes.

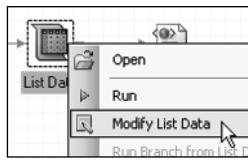
Report Listing

Reservation Number	Office	CustomerID	Tour	Travelers	Deposit
1	Portland	SL28	SH43	10	425
2	Portland	DE27	PS27	6	75
3	Portland	SL34	FJ12	4	200
4	Portland	DI33	SH43	4	150
5	Portland	BU12	SH43	2	75
6	Portland	DE31	FJ12	3	175
7	Portland	WI48	FJ12	2	100
8	Portland	NG17	PS27	5	65
9	Portland	RA28	PS27	2	30
10	Portland	ME11	PS27	2	30
11	Portland	GI08	SH43	8	300
12	Portland	HI15	SH43	4	150
13	Portland	MA09	SH43	2	75

Generated by the SAS System ('Local', XP_PRO) on July 23, 2009 at 06:52:53 PM

7.2 Customizing Titles and Footnotes

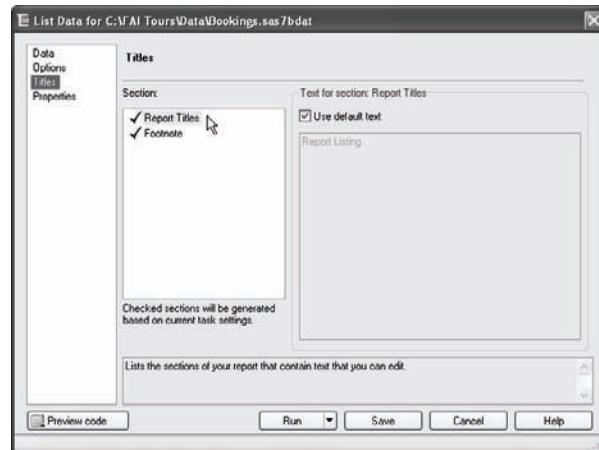
By default, reports in SAS Enterprise Guide have titles that describe the type of report, such as “Summary Statistics” or “Analysis of Variance,” and footnotes that show the date and time the task was run. That’s a good start, but in most cases, you will want titles and footnotes that reflect your unique report. You can easily customize titles and footnotes in any task that produces a report.



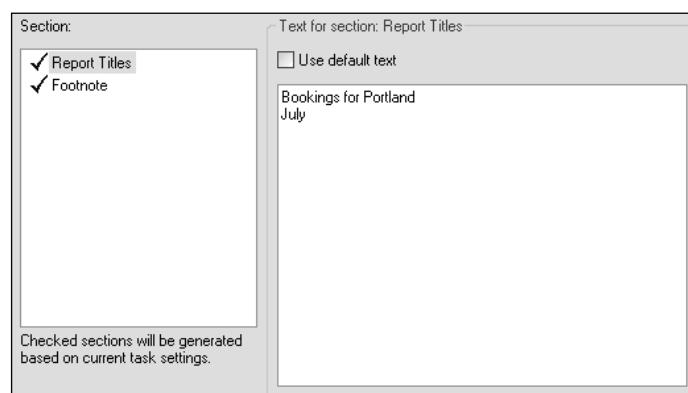
You can change the title or footnote when you first run a task, or re-open the task and modify the titles. This example takes the report that was produced in the previous section and gives it a custom title and footnote. To re-open a task, right-click the task icon in the Project Tree or Process Flow and select **Modify List Data** from the pop-up menu. The task window will open.

In the task window, click **Titles** in the selection pane on the left to display the Titles page.

Titles page The area labeled **Section** lists all the titles and footnotes for that particular task. For the List Data task, you can choose **Report Titles** or **Footnote**. Some tasks have additional titles you can change.

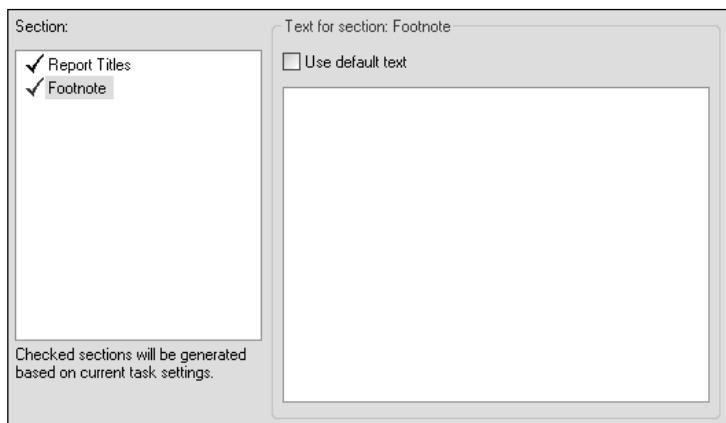


To change a title, click its name in the area labeled **Section**. Then uncheck the **Use default text** option and type up to 10 new titles in the box below. In this example, the title “Report Listing” has been replaced with two titles: “Bookings for Portland” and “July.”



To change a footnote, click **Footnote** in the area labeled **Section**. Then uncheck the **Use default text** option and type up to 10 new footnotes in the box below. In this example, the footnote has simply been deleted.

When you are satisfied with the new titles and footnotes, click **Run** in the task window.



Results Here is the report with the new titles and no footnote.

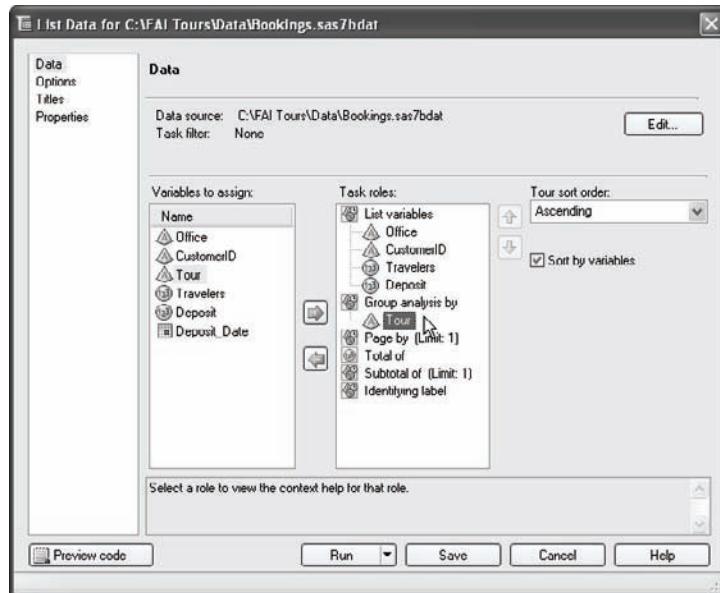
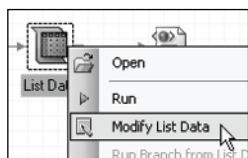
Bookings for Portland						
July						
Reservation Number	Office	CustomerID	Tour	Travelers	Deposit	
1	Portland	SL28	SH43	10	425	
2	Portland	DE27	PS27	6	75	
3	Portland	SL34	FJ12	4	200	
4	Portland	DI33	SH43	4	150	
5	Portland	BU12	SH43	2	75	
6	Portland	DE31	FJ12	3	175	
7	Portland	WI48	FJ12	2	100	
8	Portland	NG17	PS27	5	65	
9	Portland	RA28	PS27	2	30	
10	Portland	ME11	PS27	2	30	
11	Portland	GI08	SH43	8	300	
12	Portland	HI15	SH43	4	150	
13	Portland	MA09	SH43	2	75	

Changing default titles and footnotes If you find yourself changing titles and footnotes a lot, you may want to change the default values. You can do this using the Options window (see section 1.16). To open the Options window, select **Tools ▶ Options** from the menu bar. Then select **Tasks General** from the selection pane on the left. In this page you can specify new titles that will replace the default title for all tasks. You can also set the footnote to blank, or specify new footnotes to replace the default footnote.

7.3 Adding Groups to List Reports

Most tasks that produce reports allow you to assign one or more variables to serve as grouping variables. When you do this, SAS Enterprise Guide divides your data into groups based on the values of the grouping variables, and handles the groups separately. That way, you can get a report for each salesperson, or statistics for each state, or a chart for each quarter.

This example takes the report that was produced in the previous section and adds a grouping variable. To re-open the task, right-click the task icon in the Project Tree or Process Flow and select **Modify List Data** from the pop-up menu. The task window will open.



Assigning a variable to the grouping task role

To produce a report with observations divided into groups, assign one or more variables to the **Group analysis by** role. Each time you drag a variable name to the Group analysis by role, two items will appear on the right: a pull-down list for the sort order, and a check box labeled **Sort by variables**. You use these options to tell SAS Enterprise Guide whether you want the report to be sorted by that grouping variable and, if so, whether in ascending or descending order. In the List Data window above, the variable Tour is a grouping variable, and the variables Office, CustomerID, Travelers, and Deposit are list variables. When you are satisfied with the variables and their task roles, click **Run**.

Results Here is the report with Tour as a grouping variable. Notice that there is a separate section of the report for each Tour.

Bookings for Portland				
July				
Tour=FJ12				
Reservation Number	Office	CustomerID	Travelers	Deposit
1	Portland	SL34	4	200
2	Portland	DE31	3	175
3	Portland	WV48	2	100

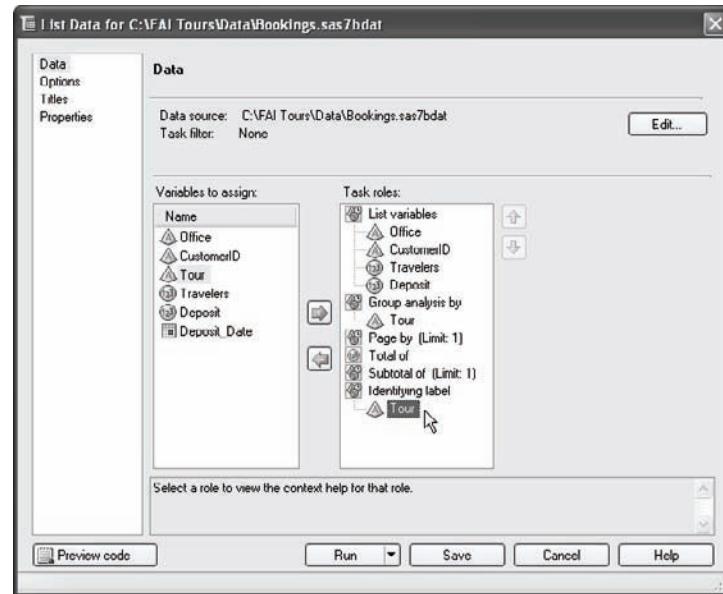
Tour=PS27				
Reservation Number	Office	CustomerID	Travelers	Deposit
4	Portland	DE27	6	75
5	Portland	NG17	5	65
6	Portland	RA28	2	30
7	Portland	ME11	2	30

Tour=SH43				
Reservation Number	Office	CustomerID	Travelers	Deposit
8	Portland	SL28	10	425
9	Portland	DI33	4	150
10	Portland	BU12	2	75
11	Portland	CIO8	8	300
12	Portland	HI15	4	150
13	Portland	MA09	2	75

Assigning a variable to the identifying label role

role When you assign a variable to serve as an identifying label, the values of that variable replace the row numbers in your report. If you assign the same variable to serve as both a grouping variable and an identifying variable, then SAS Enterprise Guide also changes the layout of the report. In this List Data window, the variable Tour has been assigned to both the **Group analysis by** role and the **Identifying label** role. When you are satisfied with the variables and their task roles, click **Run**.

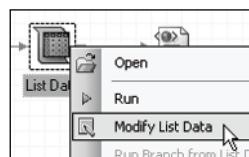
Results Here is the report with Tour serving both as a grouping variable and as an identifying variable. Notice that this report is more compact than the previous one, and that the row numbers have been replaced by the identifying variable.



Bookings for Portland					
July					
Tour	Office	CustomerID	Travelers	Deposit	
FJ12	Portland	SL34	4	200	
	Portland	DE31	3	175	
	Portland	WI48	2	100	
PS27	Portland	DE27	6	75	
	Portland	NG17	5	65	
	Portland	RA28	2	30	
	Portland	ME11	2	30	
SH43	Portland	SL28	10	425	
	Portland	DI33	4	150	
	Portland	BU12	2	75	
	Portland	GI08	8	300	
	Portland	HI15	4	150	
	Portland	MA09	2	75	

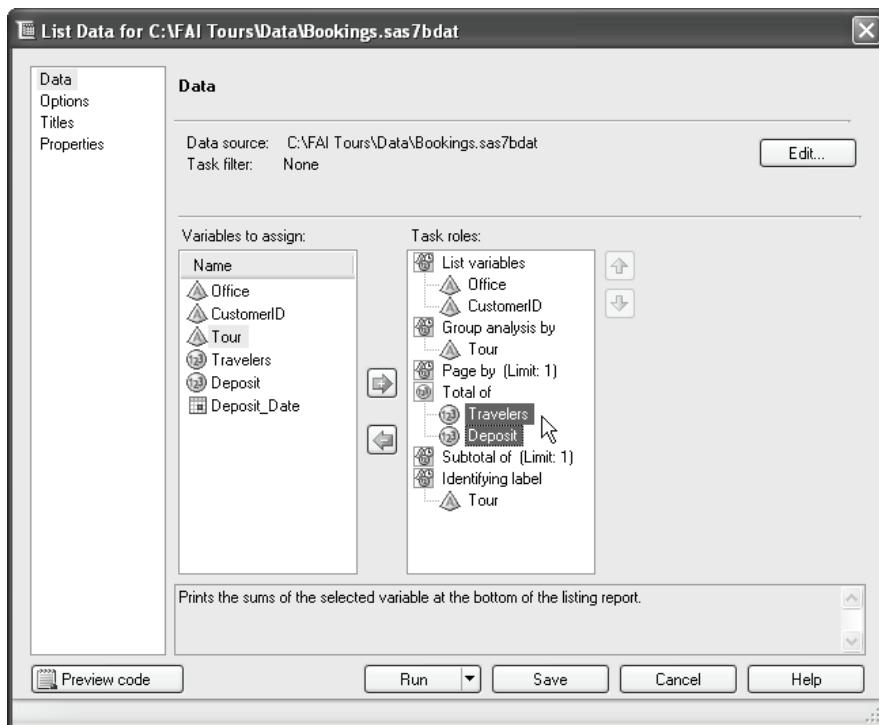
7.4 Adding Totals to List Reports

List reports have one line for each observation in a data set, but it is possible to add some summary data to these reports. Using the Total of task role, you can add subtotals for groups and a grand total at the bottom of the report.



This example takes the report that was produced in the previous section and adds totals. To re-open the task, right-click the task icon in the Project Tree or Process Flow and select **Modify List Data** from the pop-up menu. The task window will open.

Assigning a variable to the Total of task role To produce a list report with totals, assign one or more variables to the Total of task role. To do this, click a variable name and drag it to the **Total of** role. Because this variable will be summed, it must be numeric.



The List Data window above shows the task roles for the Bookings data. The variables *Travelers* and *Deposit* have been assigned to the **Total of** role. In addition, the variables *Office* and *CustomerID* have been assigned to serve as list variables, and *Tour* has been assigned as both a grouping and identifying variable. When you are satisfied with the variables and their roles, click **Run**.

Results Here is the list report of the Bookings data with the variables Travelers and Deposit totaled. Notice that there are totals for each value of the grouping variable, Tour, and a grand total at the bottom of the report. If this report did not have a grouping variable, then only the grand total would appear.

Bookings for Portland				
July				
Tour	Office	CustomerID	Travelers	Deposit
FJ12	Portland	SL34	4	200
	Portland	DE31	3	175
	Portland	WI48	2	100
FJ12			9	475
PS27	Portland	DE27	6	75
	Portland	NG17	5	65
	Portland	RA28	2	30
	Portland	ME11	2	30
PS27			15	200
SH43	Portland	SL28	10	425
	Portland	DI33	4	150
	Portland	BU12	2	75
	Portland	GI08	8	300
	Portland	HI15	4	150
	Portland	MA09	2	75
SH43			30	1175
			54	1850

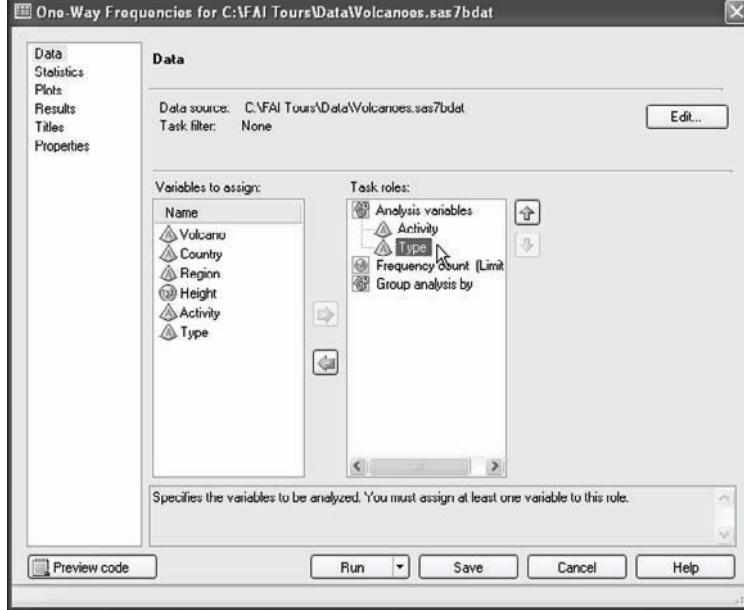
7.5 Creating Frequency Reports

If you have ever wondered exactly how many different values a particular variable has, then a one-way frequency table is what you need. Frequencies are also called counts because you can produce a basic frequency table by simply counting the number of times each data value occurs. To produce frequencies for an individual variable in SAS Enterprise Guide, use the One-Way Frequencies task.

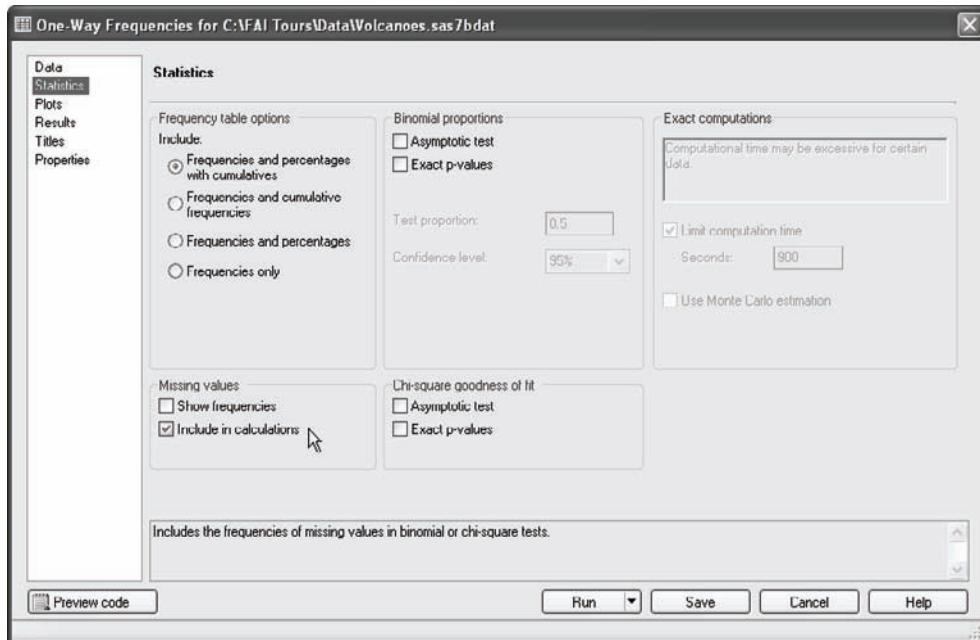
This example uses the Volcanoes data set to produce frequencies for the variables Activity and Type. Open the task by clicking the data icon in the Project Tree or Process Flow and select **Tasks ▶ Describe ▶ One-Way Frequencies** from the menu bar. The One-Way Frequencies window will open, displaying the Data page.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	Stratovolcano
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Garibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvotn	Iceland	Eu	1725	Active	Caldera

Assigning task roles For the One-Way Frequencies task, you must assign at least one variable to serve as an analysis variable. To do this, click a variable name and drag it to the **Analysis variables** role. In this example, the variables Activity and Type will be analysis variables.



Choosing statistics If you click **Statistics** in the selection pane on the left, you will see options for this task. By default, the results will include frequencies and percentages, along with cumulative frequencies and percentages. You can choose among different combinations of statistics. By default, the results exclude any missing values. If you want to include missing values in the resulting table, then check the option **Include in calculations** in the **Missing values** section. When you are satisfied with the options, click **Run**.



Results Here are the one-way frequencies for Activity and Type. Notice that there are two separate tables (one for each analysis variable), and that frequencies for missing values are included.

One-Way Frequencies					
Results					
The FREQ Procedure					
Activity					
Activity	Frequency	Percent	Cumulative Frequency	Cumulative Percent	
Active	23	71.88	25	78.13	
Extinct	7	21.88	32	100.00	
Type					
Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent	
Caldera	2	6.25	4	12.50	
Cinder Cone	1	3.13	5	15.63	
Complex	1	3.13	6	18.75	
Shield	5	15.63	11	34.38	
Stratovolcano	21	65.63	32	100.00	

7.6 Creating Crosstabulations

When you have counts for one variable, they are called one-way frequencies. When you create counts by crossing two or more variables, they are called two-way, three-way, and so on, up to n -way frequencies, or simply crosstabulations. To produce crosstabulations, use the Table Analysis task.

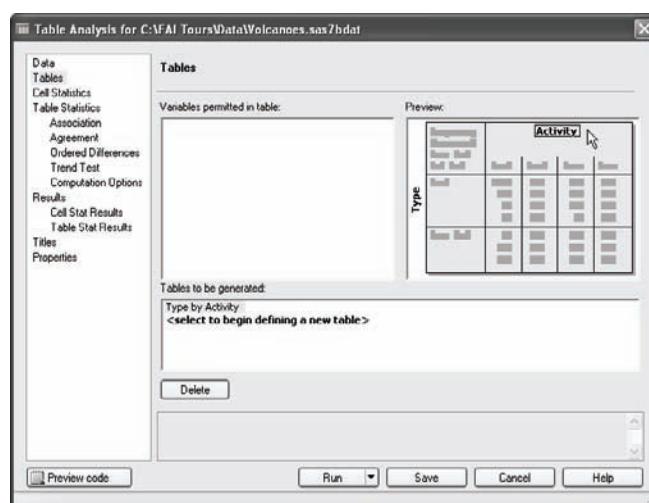
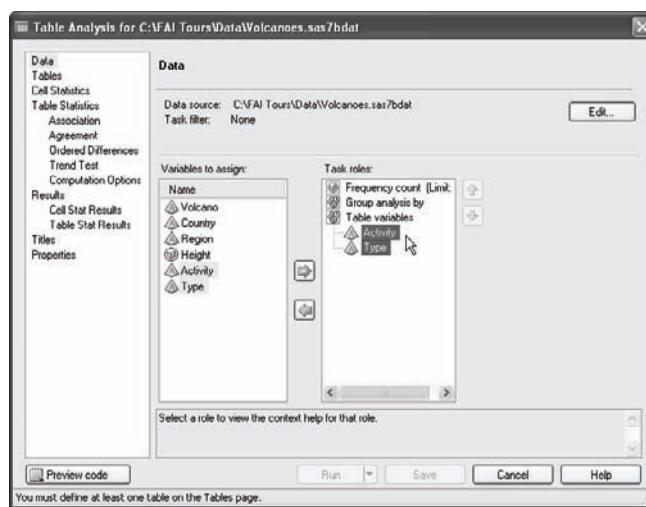
As in the previous section, this example uses the Volcanoes data set, but, instead of having one table for Activity and one for Type, it produces a single table showing Activity crossed with Type. Open the task by clicking the data icon in the Project Tree or Process Flow and select

Tasks ▶ Describe ▶ Table Analysis from the menu bar. The Table Analysis window will open, displaying the Data page.

Assigning task roles In the Table Analysis task, you must assign at least two variables to serve as table variables. In this example, the variables Activity and Type have been dragged to the **Table variables** role.

Arranging tables Click **Tables** in the selection pane on the left to display the Tables page. In this page, you will see an area labeled **Variables permitted in table**, and an area labeled **Preview**. To arrange your table, click the name of a variable and drag it to the Preview area. The values of the first variable dragged over will be used for rows, the second for columns, and the third for pages, but you can switch them around by dragging them within the Preview area. In this window, the variable Type has been assigned to the rows, and Activity to the columns.

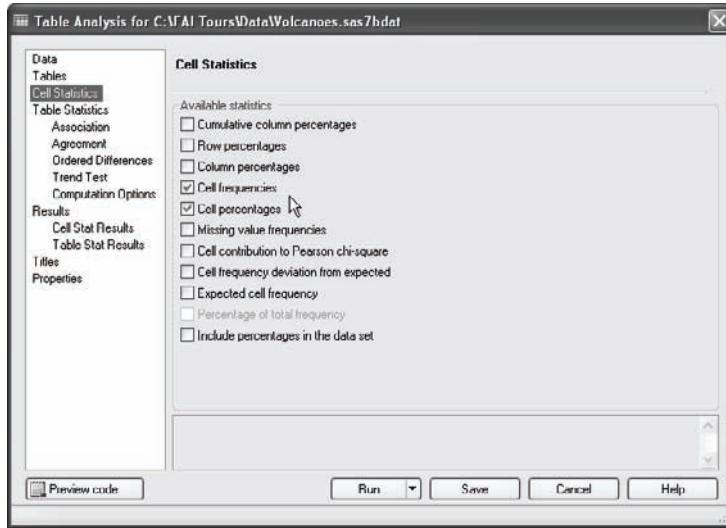
	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Freshus		An	3794	Active	Stratovolcano



When you drag a variable to the Preview area, its name will disappear from the list of variables permitted in the table. (That way you can't accidentally cross a variable with itself.) However, if you click the words **<select to begin defining a new table>** in the area labeled **Tables to be generated**, then the variable names will appear again. This allows you to define multiple tables in a single task.

Choosing cell statistics

If you click **Cell Statistics** in the selection pane on the left, you will see options for basic statistics. By default, the results include column percentages and cell frequencies. The Table Analysis task also offers more advanced statistics, such as the chi-square test. See section 9.3 for a discussion of other options. For this report, **Cell frequencies** and **Cell percentages** have been selected. When you are satisfied with the options, click **Run**.



Results Here is the two-way frequency table for Activity by Type. Because two volcanoes have missing values for Activity and two have missing values for Type, there is a note at the bottom saying that four observations are missing.

Table Analysis Results The FREQ Procedure				
		Table of Type by Activity		
		Activity		
		Active	Extinct	Total
Type	Caldera	Frequency	2	0
		Percent	7.14	0.00
Cinder Cone		Frequency	0	1
		Percent	0.00	3.57
Complex		Frequency	1	0
		Percent	3.57	0.00
Shield		Frequency	4	1
		Percent	14.29	3.57
Stratovolcano		Frequency	16	3
		Percent	57.14	10.71
Total		Frequency	20	5
		Percent	82.14	17.86
Frequency Missing = 4				

7.7 Creating Simple Summary Reports

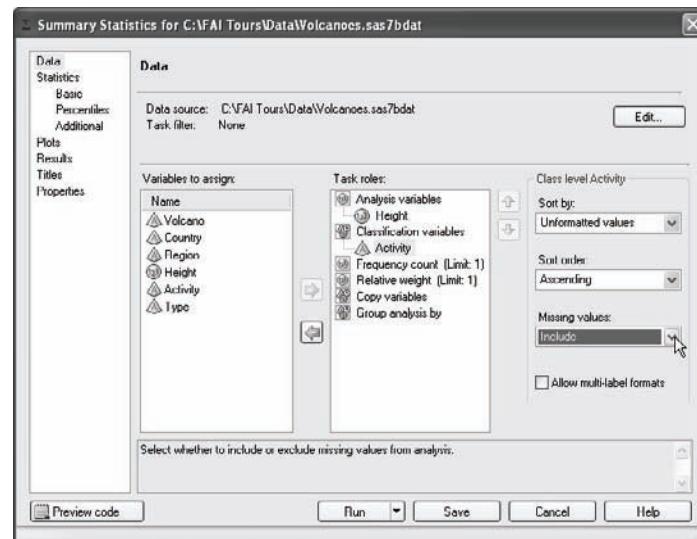
Statistics like the mean, standard deviation, and minimum and maximum values not only give you a feel for your data, but can alert you to unexpected values or errors. You can compute summary statistics with the Summary Statistics task or wizard. This section describes the task.

Here is a sample of the Volcanoes data set. To open the task, click the data icon in the Project Tree or Process Flow and select **Tasks ▶ Describe ▶ Summary Statistics** from the menu bar. The Summary Statistics window will open, displaying the Data page.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Garibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvotn	Iceland	Eu	1725	Active	Caldera

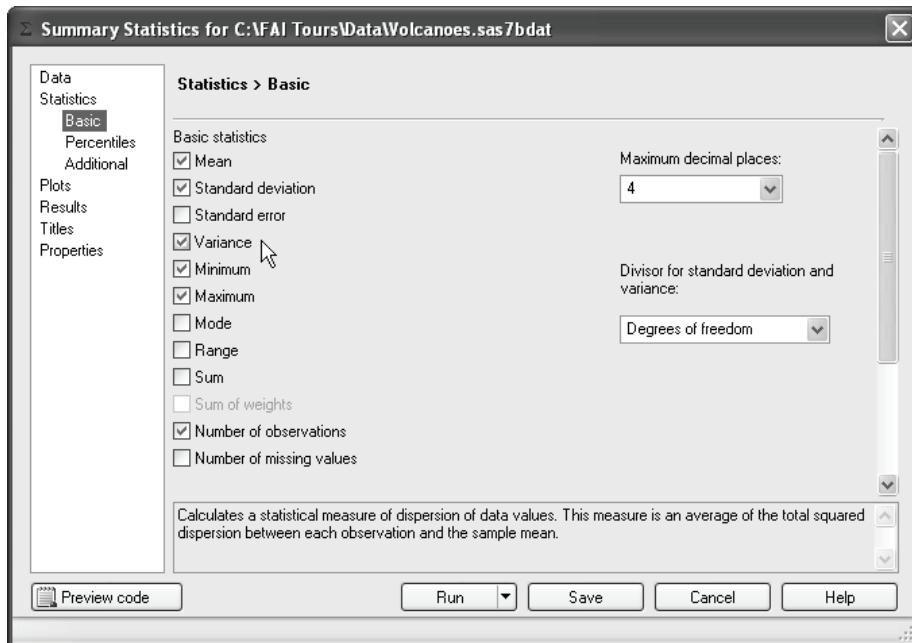
Assigning task roles For the Summary Statistics task, you must assign at least one variable to the **Analysis variables** role, and all analysis variables must be numeric. Classification variables, on the other hand, are optional and may be numeric or character.

If you assign a variable to the **Classification variables** role, then SAS Enterprise Guide will produce separate summary statistics for each combination of the classification variables. When you drag a variable to the classification role, options will appear on the right. You can choose the sort order (Ascending or Descending) and whether to include missing values. The default is to exclude any observations with missing values for the classification variables. The **Group analysis by** role is similar to the classification role, but it produces a separate table for each combination of the grouping variables.



In this example, the variable Height has been designated as an analysis variable, and Activity as a classification variable. The option **Missing values** has been set to **Include**.

Choosing statistics Click **Basic** under the **Statistics** group of options in the selection pane on the left to display the Basic page. Here you can choose statistics and set the number of decimal places to be displayed in the report. The Summary Statistics task offers many other statistics, including the coefficient of variation and percentiles. See section 9.2 for a discussion of more advanced options. In this case, **Variance** has been selected along with the default statistics, and the number of decimal places has been set to four. When you are satisfied with the settings, click **Run**.



Results Here is the report showing summary statistics for Height by Activity. Notice that, because missing values were included, there is a separate line for volcanoes with a missing value for the variable Activity.

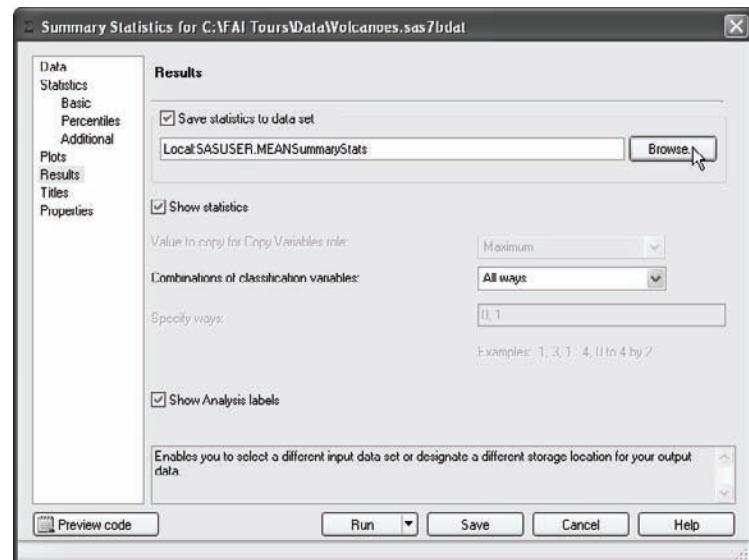
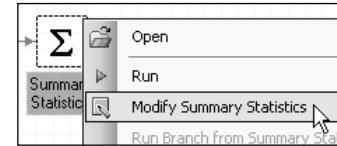
Analysis Variable : Height							
Activity	N Obs	Mean	Std Dev	Variance	Minimum	Maximum	N
	2	4286.5000	2274.7625	5174544.5000	2678.0000	5895.0000	2
Active	23	2852.6087	1508.0614	2274249.1581	354.0000	5976.0000	23
Extinct	7	3635.8571	2572.7682	6619136.1429	251.0000	6458.0000	7

7.8 Creating Summary Data Sets in a Task

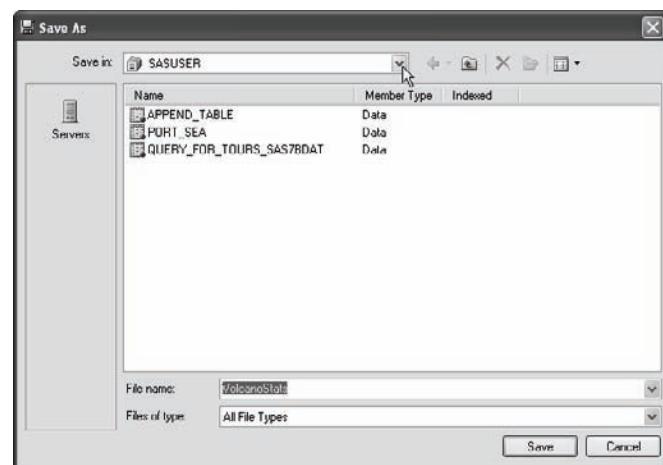
Sometimes you may want to save summary data so you can use it for further analysis or join it with other data. Many tasks can save summary data, including Table Analysis, Summary Tables, and Summary Statistics.

The previous section used the Volcanoes data set to produce a report showing summary statistics. These summary statistics can be saved in a data set. To re-open a task, right-click the task icon in the Project Tree or Process Flow and select **Modify Summary Statistics** from the pop-up menu. The task window will open. Click **Results** in the selection pane on the left to display the Results page.

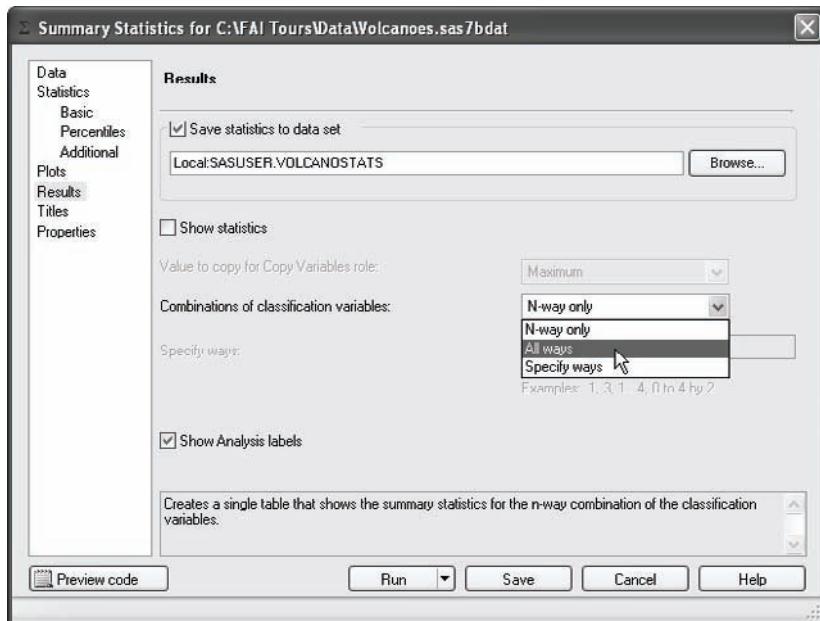
Results page In the Results page, you will see options that affect both your printed report and output data set. To save an output data set, check the **Save statistics to data set** option. SAS Enterprise Guide gives the data set a name beginning with MEAN and stores it in a default location. To specify a different name or location, click **Browse**. This opens the Save As window.



Save As window First, choose a library from the box labeled **Save in**, and then type a name for your file in the **File name** box. In this window, the data set has been named VolcanoStats and will be saved in the SASUSER library. When you are satisfied, click **Save** to return to the Results page.



Choosing options In the Results page, you will see the new data set name. Before you run the report, you may want to make some other changes. Because you are creating an output data set, you may not care about the standard Summary Statistics report. To turn off the report, uncheck the **Show statistics** box.



You can also select the **Combinations of classification variables** that will be included in the results. **N-way only** (the default) includes only the highest interaction of the classification variables. The option **All ways** gives you summary data for all combinations of the classification variables including the grand total. If you choose **Specify ways**, then you can select just the combinations you want. If you have many classification variables, then you will have many possible combinations. In this example, **All ways** is being selected. When you are satisfied with the settings, click **Run**.

Results Here is the output data set displayed in a Data Grid. The first row contains summary statistics for the grand total, while the following rows contain statistics for each level of Activity (missing, Active, and Extinct). Notice that SAS Enterprise Guide has created new variables for the summary statistics (Height_Mean, Height_StdDev, Height_Var, Height_Min, Height_Max, and Height_N). In addition, there are three automatic variables. _FREQ_ tells you how many observations contributed to each group, while _WAY_ and _TYPE_ reflect the type of combination. The _WAY_ and _TYPE_ are the same except that _WAY_ is a numeric variable while _TYPE_ is character.

	Activity	_WAY_	_TYPE_	_FREQ_	Height_Mean	Height_StdDev	Height_Var	Height_Min	Height_Max	Height_N
1		0	0	32	3114	1006	3262900.96	251	6450	32
2		1	1	2	4287	2275	5174544.5	2678	5895	2
3	Active	1	1	23	2853	1508	224249.16	354	5976	23
4	Extinct	1	1	7	3636	2573	6619136.14	251	6458	7

7.9 Creating Grouped Reports with User-Defined Formats

Often you want to summarize your data by groups. For example, if you had a variable for age, you might want to see separate summary statistics for children and adults. If you did not have a variable for age group, you could compute it in a Data Grid, or use the Query Builder to recode your data. Both of these methods add a new variable to your data set. If you don't want to add a new variable, then you can group your data with a user-defined format.

Many tasks allow you to group data using a format. These include List Data, One-Way Frequencies, Table Analysis, Summary Statistics, and Summary Tables. This section groups the Volcanoes data set in a One-Way Frequencies task using the HeightGroup. format created in section 3.4. The HeightGroup. format assigns volcanoes with heights of 0-499 meters to the group Pip-squeak, 500-3999 meters to Middling, and 4000 meters and above to Stupendous.

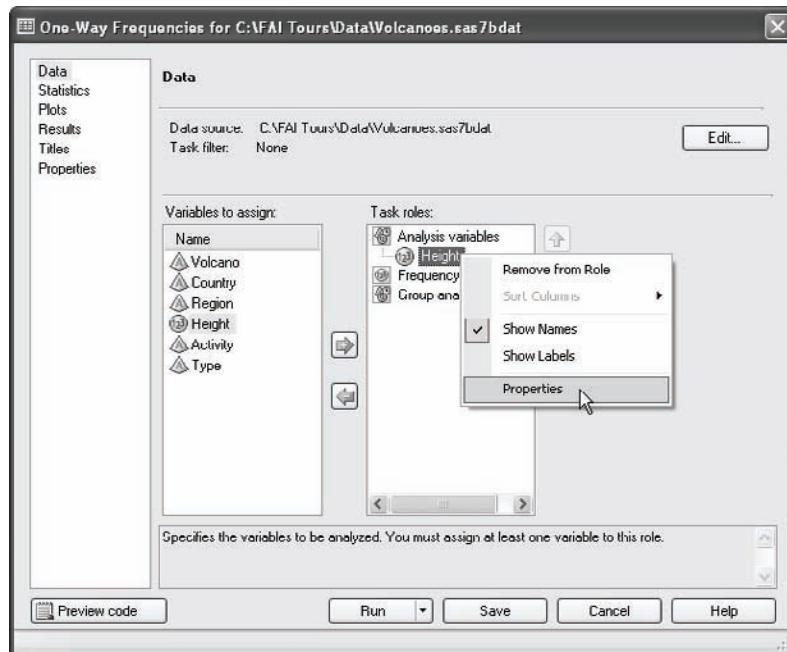
Here is a sample of the Volcanoes data set. To open the task, click the data icon in the Project Tree or Process Flow and select **Tasks ► Describe ► One-Way Frequencies** from the menu bar.

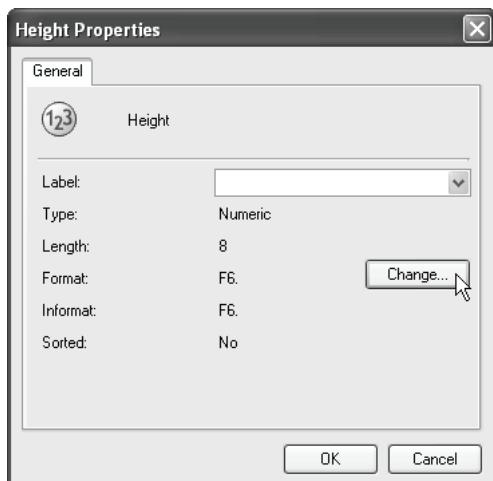
The One-Way Frequencies window will open, displaying the Data page.

Opening the Properties window

Properties To apply a format in a task, right-click the name of the variable you want to be grouped in the Data page (in either the **Variables to assign** area or the **Task roles** area), and select **Properties** from the pop-up menu. In this example, **Properties** is being selected for the variable Height.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Baren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Frethun		An	2794	Active	Stratovolcano

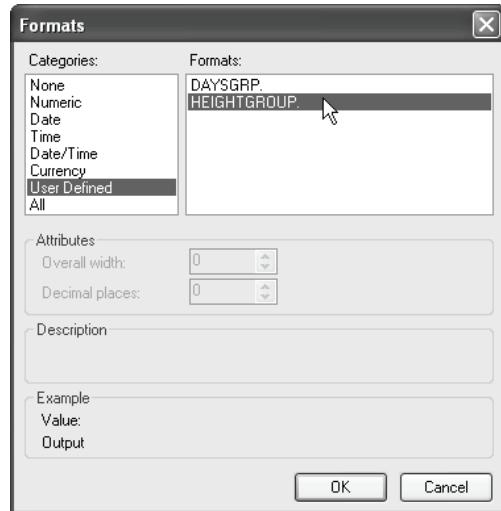




Once you have selected the correct format, click **OK** in the Formats window and click **OK** in the Properties window. Then click **Run** in the task window.

Using custom formats In the Properties window, click **Change** to open the Formats window for that variable.

Then in the Formats window, select the category **User Defined**. All the formats you have created will be listed. Here is the Formats window for the variable Height. Because Height is numeric, only numeric formats are listed. In this example, HEIGHTGROUP. is being selected.



Results Here is the grouped report. Notice that the user-defined format was applied to the variable before it was summarized. Instead of counts for data values like 251 and 5321, the report shows counts for the formatted values, Pip-squeak, Middling, and Stupendous.

One-Way Frequencies				
Results				
The FREQ Procedure				
Height	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Pip-squeak	3	9.38	3	9.38
Middling	20	62.50	23	71.88
Stupendous	9	28.13	32	100.00

8

“ Every now and then a man’s mind is stretched by a new idea or sensation, and never shrinks back to its former dimensions.”

OLIVER WENDELL HOLMES, SR.



CHAPTER 8

Producing Complex Reports in Summary Tables

- 8.1 Creating Summary Tables with Frequencies 270
- 8.2 Adding Statistics to Summary Tables 272
- 8.3 Changing Heading Properties in Summary Tables 274
- 8.4 Changing Class Level Headings and Properties in Summary Tables 276
- 8.5 Changing Table Properties in Summary Tables 278
- 8.6 Changing Data Value Properties in Summary Tables 280

8.1 Creating Summary Tables with Frequencies

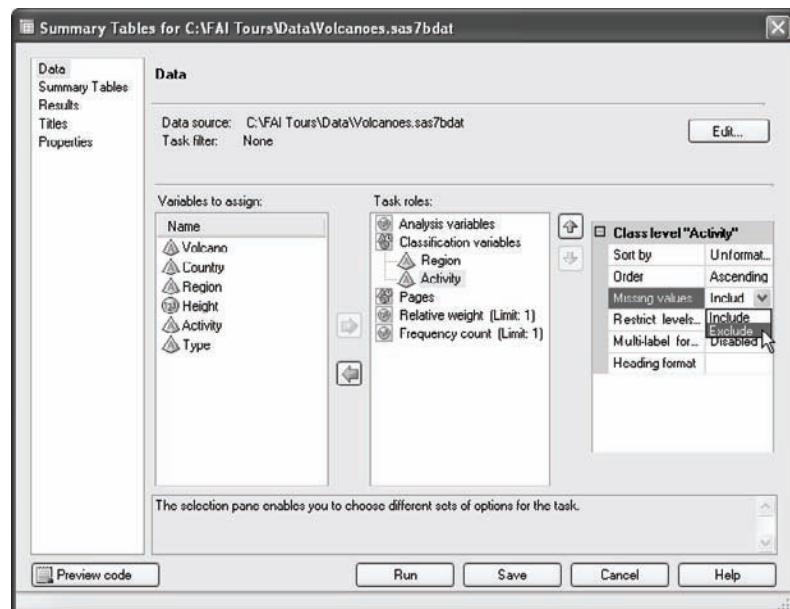
The Summary Tables task is the most powerful and flexible of the reporting tasks in SAS Enterprise Guide. It gives you control not only over which data appear in a report, but also over how data are arranged, summarized, labeled, and even colored. You can access some of these features using the Summary Tables wizard. This chapter shows the task.

This example uses the Volcanoes data set to create a report showing the number of active and extinct volcanoes for each region. To open the task, click the data icon in the Project Tree or Process Flow, and select **Tasks ▶ Describe ▶ Summary Tables** from the menu bar. The Summary Tables window will open, displaying the Data page.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Garibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvotn	Iceland	Eu	1725	Active	Caldera

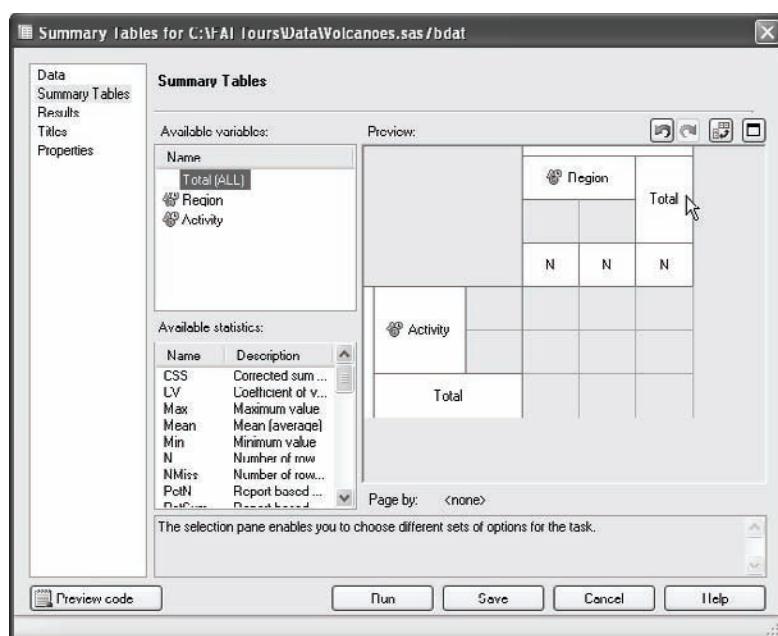
Assigning task roles To produce a summary table showing frequencies, assign one or more variables to the **Classification variables** role. These variables may be character or numeric. SAS Enterprise Guide will divide the data into categories based on the values of the classification variables. The following window shows the Volcanoes data set with the variables Region and Activity serving as classification variables.

When you drag a variable to the classification role, a box appears on the right. In this box, you can select options for that variable, including how to handle missing data. By default, missing values are included as valid rows and columns, which may or may not be what you want. In this case, two volcanoes have a missing value for Activity. To avoid having an entire row devoted to missing values, click **Missing values** and select **Exclude** for the variable Activity.



Once you exclude missing values for a variable, observations with missing values will be excluded from the report even if you decide not to use that particular variable. Because of this, it's a good idea to assign variables to the classification role only if you intend to use them in the current report.

Arranging your table Before you can run a report, you must tell SAS Enterprise Guide how to arrange the report table. Start by clicking the **Summary Tables** option in the selection pane on the left. In the Summary Tables page, you will see areas labeled **Available variables** and **Preview**. To assign a variable to serve as a row or column in your report, drag the variable name from the list of available variables to the Preview area. It may take a little practice to get variables where you want them. The trick is to watch the cursor. If the cursor looks like the universal not-allowed symbol, \otimes , then you cannot drop the variable. When the cursor turns into an arrow, then you can drop the variable.



The undo and redo buttons in the upper-right corner of the Preview area can be quite useful. If you want to switch the row variables with the column variables, use the pivot button. To enlarge the Preview area, click the maximize button . You can drag the **Total(ALL)** variable to the Preview area to tell SAS Enterprise Guide where to

insert totals. In this window, the values of the variable **Activity** form the rows, while the values of **Region** form the columns. When you are satisfied with the arrangement of your table, click **Run**.

Results Here is the report of Activity by Region. Notice that the value in each cell is simply the number of volcanoes in that category. N (the number of non-missing values) is the default statistic for classification variables.

		Region							Total (ALL)
		AP	Af	An	As	Eu	NA	SA	
Activity		N	N	N	N	N	N	N	N
Active		3	2	1	5	4	5	3	23
Extinct		1	1	.	.	3	.	2	7
Total (ALL)		4	3	1	5	7	5	5	30

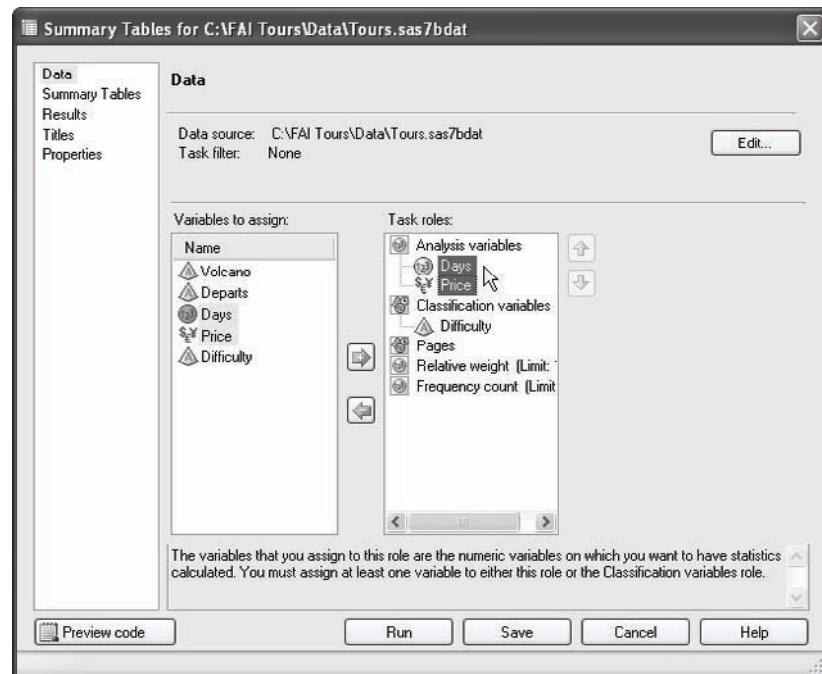
8.2 Adding Statistics to Summary Tables

The previous section showed how to produce a table containing simple counts. Sometimes that's all you need, but often you want more. You might want to know total sales by region, or the mean test score for each class. In summary tables, you can compute sums and means, plus a long list of other statistics, including maximum and minimum values, percentages, medians, quartiles, standard deviations, and variances.

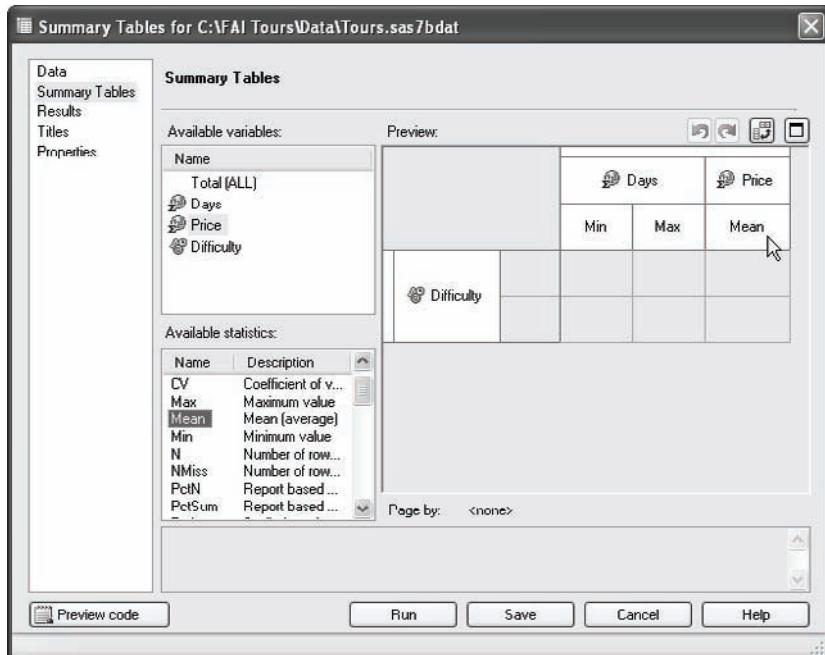
This example uses the Tours data set to create a report with statistics. To open the task, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Describe ▶ Summary Tables** from the menu bar. The Summary Tables window will open, displaying the Data page.

	Volcano	Departs	Days	Price	Difficulty
1	Etna	Catania	7	\$1,075 m	
2	Fuji	Tokyo	2	\$225 c	
3	Kenya	Nairobi	6	\$830 m	
4	Kilauea	Hilo	1	\$55 e	
5	Kilimanjaro	Nairobi	9	\$1,310 c	
6	Krakatau	Jakarta	7	\$895 e	
7	Poas	San Jose	1	\$65 e	
8	Rieventador	Quito	4	\$575 m	
9	St. Helens	Portland	2	\$167 e	
10	Vesuvius	Rome	6	\$985 e	

Assigning task roles There are a few statistics you can compute for classification variables. These include N and PctN (the percentage of frequency). However, most statistics can be computed only for analysis variables. Analysis variables must be numeric. (It's simply not possible to compute a mean using character values like Active and Extinct.) To produce a summary report containing sums and means, assign one or more variables to the analysis role. The following window shows the Tours data set with the variable Difficulty serving as a classification variable, and the variables Days and Price serving as analysis variables.



Arranging your table You arrange analysis variables in your table the same way you arrange classification variables. First, click the **Summary Tables** option in the selection pane on the left. In the Summary Tables page, you will see areas labeled **Available variables** and **Preview**. To assign a variable to serve as a row or column in your report, drag the variable name from the list of available variables to the Preview area. In the following window, the values of Difficulty form the rows, and the values of Days and Price form the columns.



Choosing statistics The default statistic for classification variables is N (the number of non-missing values). The default statistic for analysis variables is Sum. You can choose many other statistics from the box labeled **Available statistics**. To add a statistic to your report, click the name of the statistic and drag it to the **Preview** area. Be sure to watch your cursor carefully. If the cursor looks like the universal not-allowed symbol, \otimes , then you cannot drop the statistic. When the cursor turns into an arrow, then you can drop the statistic.

In the preceding window, the statistics Min and Max have been placed under the variable Days, and the statistic Mean has been placed under the variable Price. When you are satisfied with the arrangement of your table, click **Run**.

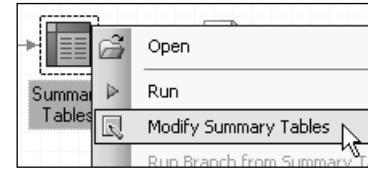
Results Here is the report of Difficulty by Days and Price. Notice that the values in the cells are the minimum and maximum number of Days, and the mean Price.

		Days		Price
		Min	Max	Mean
Difficulty				
c	2.00	9.00	767.50	
e	1.00	7.00	433.40	
m	4.00	7.00	826.67	

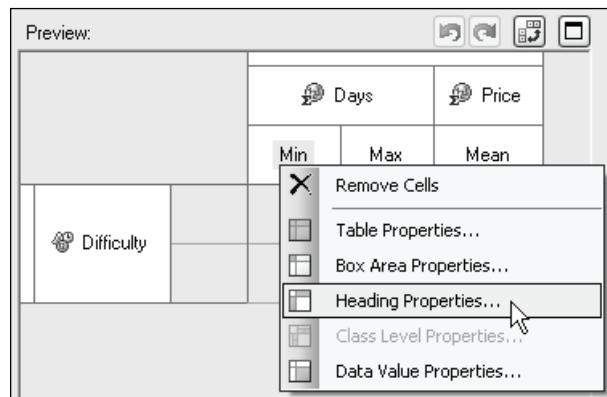
8.3 ▶ Changing Heading Properties in Summary Tables

Once you've constructed a summary table, put each variable in its proper place, and selected statistics, you may want to change the way the table looks. In the Preview area of the Summary Tables window, you can change many properties of headers and data values.

To modify an existing report, right-click the Summary Tables task icon in the Project Tree or Process Flow, and select **Modify Summary Tables** from the pop-up menu. The Summary Tables window will open. To display the Preview area, click the **Summary Tables** option in the selection pane on the left.



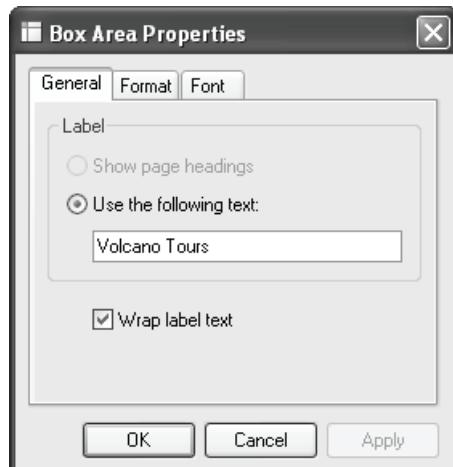
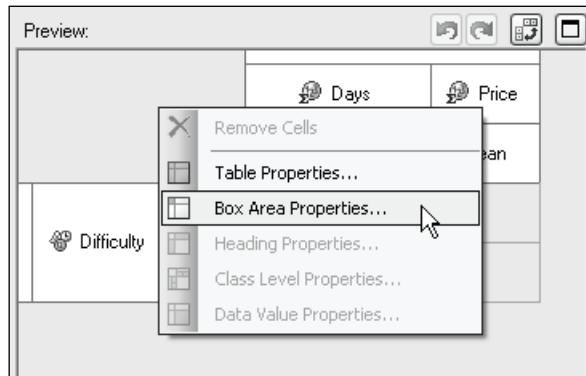
Heading Properties To change headings that are the names of variables or statistics, use the Heading Properties window. For example, to change the heading Min to Minimum, you would right-click **Min** in the **Preview** area and select **Heading Properties** from the pop-up menu. The Heading Properties window will open.



Using the **General** tab of the Heading Properties window, you can type a new label for the variable or statistic you have selected. In this window, the word **Minimum** has been typed in the **Label** box. Using the **Font** tab, you can change the font, font style, size, foreground color, background color, and other attributes of headings. When you are satisfied with the changes, click **OK**.

You can now change other properties. For this example, you should also change the statistic name **Max** to **Maximum** using the Heading Properties window for **Max**.

Box Area Properties Summary Tables reports always contain a box in the upper-left corner. By default, this box is empty. But you can put a label in that box to give your reports a nicely polished look. To do this, right-click anywhere in the Preview area and select **Box Area Properties** from the pop-up menu.



Using the **General** tab of the Box Area Properties window, you can specify the text you want printed in the box area. In this example, the words Volcano Tours have been typed in the text box. Using the **Font** tab, you can change the font, font style, size, foreground color, background color, and other attributes of the text to be printed in the box area.

When you are satisfied with the changes, click **OK**. You can then change other properties, or click **Run** to see the new results.

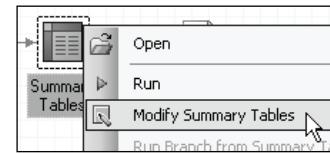
Results Here is the report. Notice that the labels Min and Max have been replaced with Minimum and Maximum, and the phrase Volcano Tours has been inserted in the box area.

Summary Tables				
Volcano Tours	Days			Price
	Minimum	Maximum	Mean	
Difficulty				
c	2.00	9.00	767.50	
e	1.00	7.00	433.40	
m	4.00	7.00	826.67	

8.4 ▶ Changing Class Level Headings and Properties in Summary Tables

The previous section showed how to change headings that are the names of variables or statistics, but classification variables also use data values as headings. These data values are called class level headings. When you change class level headings, you are changing the way those data values are displayed. To change the way data values are displayed, you use a format.

To modify an existing report, right-click the Summary Tables task icon in the Project Tree or Process Flow, and select **Modify Summary Tables** from the pop-up menu. The Summary Tables window will open.

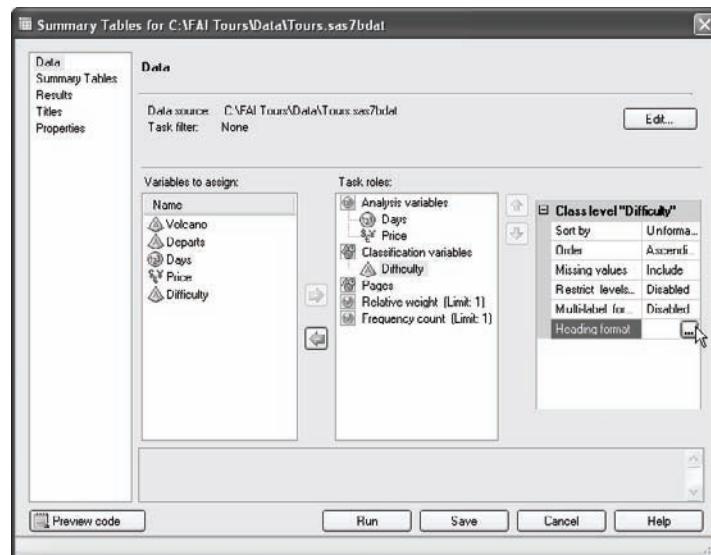


Applying a format to a classification variable

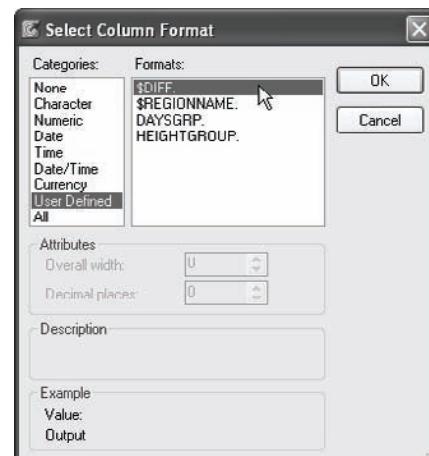
To change headings that are data values, you specify a format in the Data page. Click the name of the classification variable you want to change. A box will open on the right, listing options for that variable. Click the words **Heading format** and the ellipsis button will appear



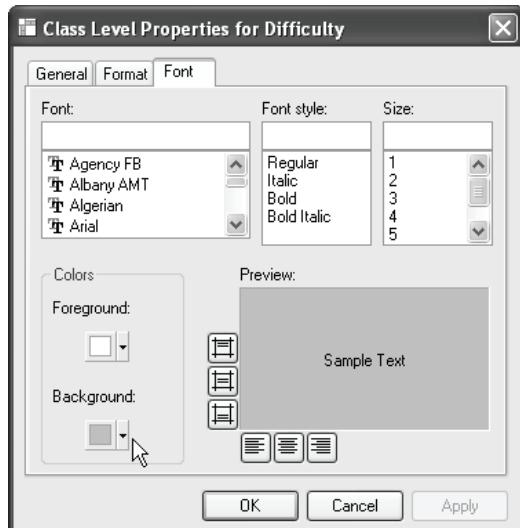
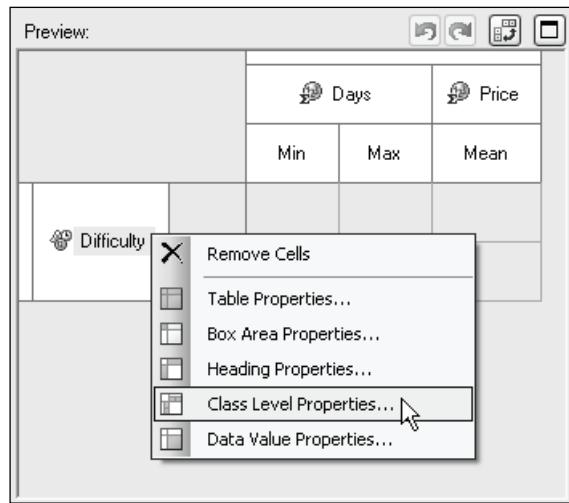
. Click the button to open a Format window for that variable. In this example, the heading format for the variable Difficulty is being selected.



In the Select Column Format window, choose the category of formats you want to see, and then choose the name of the format you want to use. In most cases, to change a class level heading you will need a user-defined format. In this example, the user-defined format \$DIFF. has been selected. The \$DIFF. format was created in Tutorial B. Sections 3.3 and 3.4 also show how to create user-defined formats. Once you are satisfied, click **OK** to return to the Summary Tables window.



Class Level Properties To change other properties of class level headings, click the **Summary Tables** option in the selection pane on the left. Then right-click the name of the classification variable in the **Preview** area and select **Class Level Properties** from the pop-up menu. This example shows Class Level Properties being selected for the variable Difficulty. The Class Level Properties window will open.



Using the **Font** tab, you can change the font, font style, size, foreground color, background color, and other attributes of headings. For this report, the background color has been changed to Gray-25%. When you are satisfied with the changes, click **OK**. Click **Run** to see the new results.

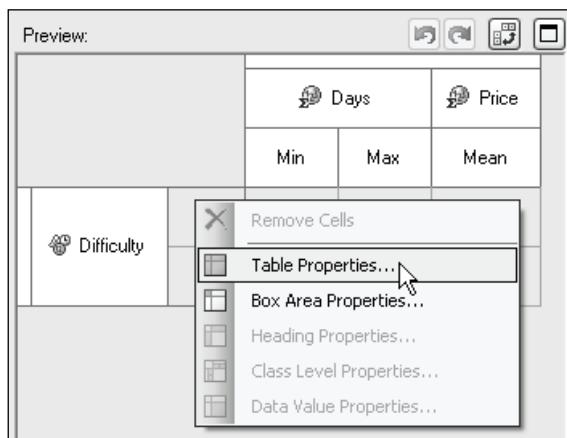
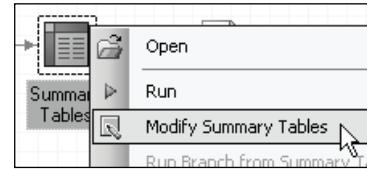
Results Here is the report. Notice that the labels c, e, and m have been replaced with Challenging, Easy, and Moderate (the values of the \$DIFF. format), and have a medium gray background.

Summary Tables				
Volcano Tours	Days		Price	
	Minimum	Maximum	Mean	
Difficulty				
Challenging	2.00	9.00	767.50	
Easy	1.00	7.00	433.40	
Moderate	4.00	7.00	826.67	

8.5 ➤ Changing Table Properties in Summary Tables

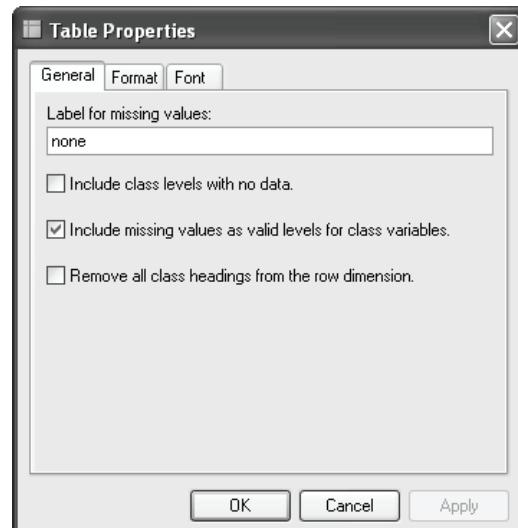
In addition to changing headers and labels, you can make changes to the data cells in a table. To make a change that will apply to all the cells, use the Table Properties window.

To modify an existing report, right-click the Summary Tables task icon in the Project Tree or Process Flow, and select **Modify Summary Tables** from the pop-up menu. The Summary Tables window will open. Click the **Summary Tables** option in the selection pane on the left to display the Preview area.

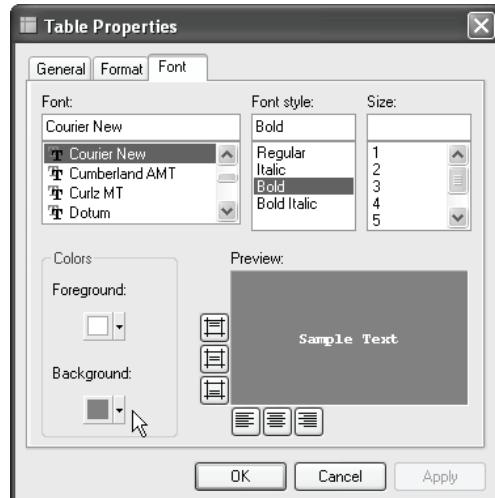
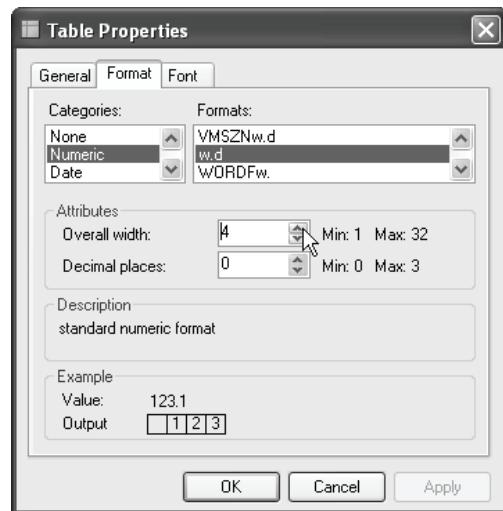


To make changes to all the data cells of a table, right-click anywhere in the **Preview** area and select **Table Properties** from the pop-up menu.

General tab Using the **General** tab of the Table Properties window, you can specify options for the treatment of missing values and class variable levels. By default, missing values are displayed as a period (.). You can specify a more meaningful label. In this example, the label none has been assigned to missing values.



Format tab Using the **Format** tab of the Table Properties window, you can choose a format for the data in the cells of the table. Here the basic numeric format, *w.d*, has been specified with an overall width of 4 characters, and no decimal places.



Font tab Using the **Font** tab of the Table Properties window, you can change the font, font style, size, foreground color, background color, and other attributes of the data cells in your table. In this case, the font has been set to Courier New, the style to Bold, the foreground color to white, and the background color to Gray–50%.

When you are satisfied with the changes, click **OK**. You can then change other properties, or click **Run** to see the new results.

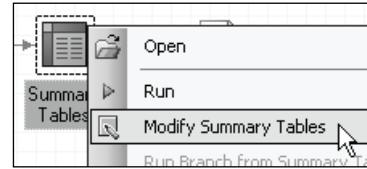
Results Here is the new report. Notice that the data cells have a dark gray background and white foreground. Also, the data are displayed in bold Courier New, and with no decimal places.

Summary Tables			
Volcano Tours	Days		Price
	Minimum	Maximum	
Difficulty			
Challenging	2	9	768
Easy	1	7	433
Moderate	4	7	827

8.6 Changing Data Value Properties in Summary Tables

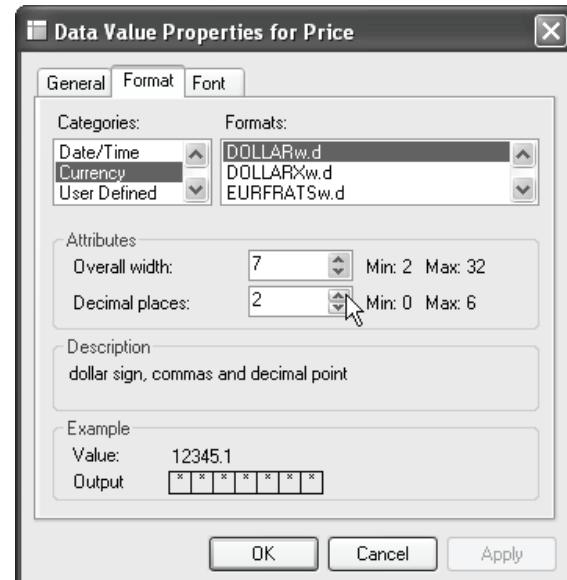
Using the Table Properties window, you can make changes to all the data cells in a report, but sometimes you may want to choose different formats or fonts for different variables or statistics. To do that, use the Data Value Properties window.

To modify an existing report, right-click the Summary Tables task icon in the Project Tree or Process Flow, and select **Modify Summary Tables** from the pop-up menu. The Summary Tables window will open. Click the **Summary Tables** option in the selection pane on the left to display the Preview area.



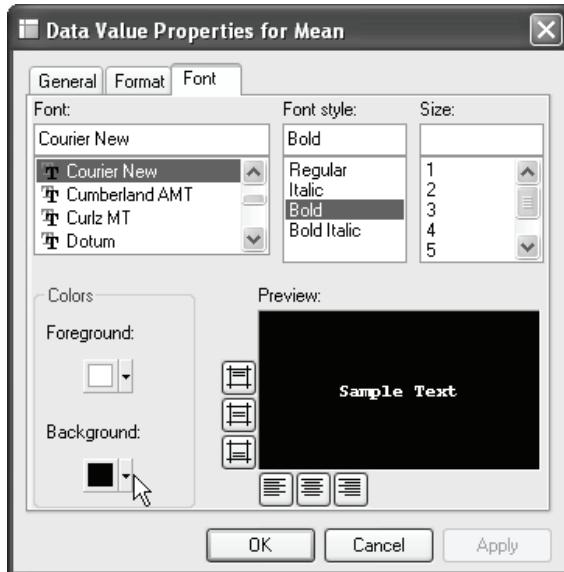
To make changes to a particular row or column, right-click the header for that row or column in the **Preview** area and select **Data Value Properties** from the pop-up menu. In this Preview area, Data Value Properties is being selected for the column Price.

Format tab Using the **Format** tab of the Data Value Properties window, you can choose a format for the data values in the row or column. In this example, the category Currency has been selected, and SAS Enterprise Guide has listed all the available formats for currency data. The format DOLLARw.d is selected, with an overall width of 7 characters, and 2 decimal places.



Font tab Using the **Font** tab of the Data Value Properties window, you can change the font, font style, size, foreground color, background color, and other attributes of the data cells in the row or column. In this example, the font for Price has been set to Courier New, the style to Bold, the foreground color to white, and the background color to black.

When you are satisfied with the changes, click **OK**. You can now change other properties, or click **Run** to see the new results.



Summary Tables			
Volcano Tours	Days		Price
	Minimum	Maximum	Mean
Difficulty			
Challenging	2	9	\$767.50
Easy	1	7	\$433.40
Moderate	4	7	\$826.67

Results Here is the report. Notice that the column for Mean Price looks different from the columns for Minimum and Maximum Days. The background color is black instead of gray, and the numbers have dollar signs in front of them and two decimal places.

9

“ Statistics may be defined as a body of methods for making wise decisions in the face of uncertainty. ”

W. ALLEN WALLIS AND HARRY V. ROBERTS



CHAPTER 9

Basic Statistical Analysis

- 9.1 Distribution Analysis 284
- 9.2 Summary Statistics 286
- 9.3 Table Analysis 288
- 9.4 Correlations 290
- 9.5 Linear Regression 292
- 9.6 Analysis of Variance 294

9.1 Distribution Analysis

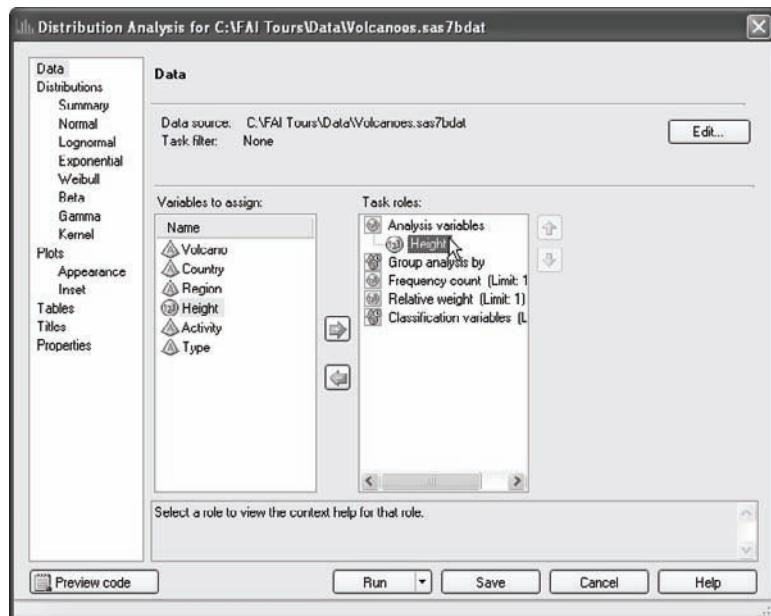
When you are doing statistical analysis, generally your goal is to examine the relationship between two or more variables. You may want to know how length of day affects the growth of plants, or how an advertising campaign influences sales. But before you start testing hypotheses, it's a good idea to pause and do a little exploration. The Distribution Analysis task is a good place to start. Distribution Analysis produces statistics describing the distribution of a single variable.

This example explores the distribution of the variable Height in the Volcanoes data set. In the Project Tree or Process Flow, click the data icon to make it active. Then select **Tasks ▶ Describe ▶ Distribution Analysis** from the menu bar. The Distribution Analysis window will open, displaying the Data page.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Baren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Gairibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvoth	Iceland	Eu	1725	Active	Caldera
10	Hilmaen	Norway	EA	5450	Extinct	Stratovolcano

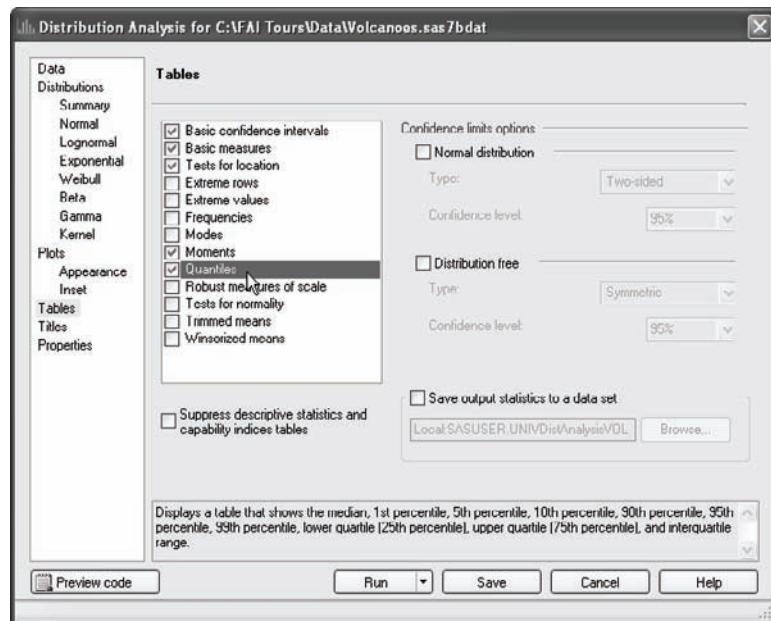
Assigning task roles For Distribution Analysis, you must assign at least one variable to serve as an analysis variable, and that variable must be numeric. In this example, the variable Height has been assigned to the **Analysis variables** role.

Distributions and Plots In the Summary page, choose the distributions (if any) that you want to fit to your data. Then, when you open the pages for the distributions you selected, you can see the available options for the distribution. The plot style (traditional or ODS) is selected in the Summary page while the desired plots are selected in the Appearance page. The Inset page allows you to add insets to some plot types that show statistics you select. You can also control the location and format for the inset.



Choosing statistics

In the Tables page, you can choose sets of statistics. For this example, select **Basic confidence intervals**, **Basic measures**, **Tests for location**, **Moments**, and **Quantiles**. When you are satisfied with your selections, click **Run**.



Distribution analysis of: Height		
The UNIVARIATE Procedure		
Variable: Height		
Moments		
N	32	Sum Weights
Mean	3113.5625	Sum Observations
Std Deviation	1806.35239	Variance
Skewness	0.16096928	Kurtosis
Uncorrected SS	411366864	Corrected SS
Coeff Variation	58.0156137	Std Error Mean
Basic Statistical Measures		
Location		Variability
Mean	3113.563	Std Deviation
Median	2957.500	Variance
Mode	.	Range
		Interquartile Range
Basic Confidence Limits Assuming Normality		
Parameter	Estimate	95% Confidence Limits
Mean	3114	2462 3765
Std Deviation	1806	1448 2402
Variance	3262909	2097164 5767244
Tests for Location: Mu0=0		
Test	Statistic	p Value
Student's t	t 9.750572	Pr > t <.0001
Sign	M 16	Pr >= M <.0001
Signed Rank	S 264	Pr >= S <.0001

Results The resulting report starts with basic information about the distribution of the variable: the number of observations (N), mean, and standard deviation. Skewness indicates how symmetrical the distribution is (whether it is more spread out on one side than the other), while kurtosis indicates how flat or peaked the distribution is. Other sections of the report contain the mean, the median, and the mode (in this case, there is no mode because no two volcanoes had the same value of Height); confidence limits assuming normality; tests of the hypothesis that the mean is zero; and quantiles.

9.2 Summary Statistics

There are many ways to summarize data in SAS Enterprise Guide. The Summary Statistics task gives you basic descriptive statistics like the mean, minimum, and maximum. You can also request more advanced statistics such as the coefficient of variation and quartiles. The Distribution Analysis task produces many of the same statistics, but the Summary Statistics task gives you more control over which specific statistics are produced and formats the results differently.

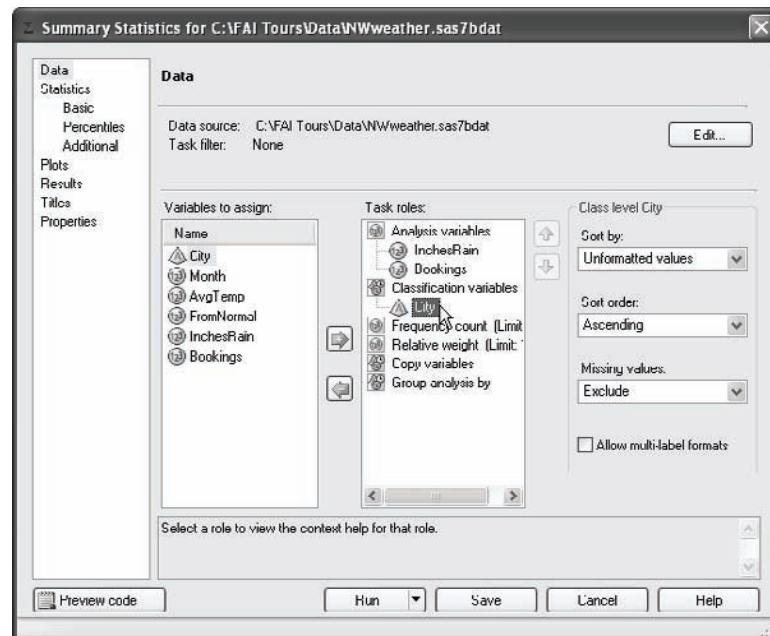
The Fire and Ice Tours company has weather data by month for both its Seattle and Portland offices, along with the number of tour bookings for each month. Here is a sample of the data set, NWweather. To produce summary statistics, click the data icon in the Project Tree or

	City	Month	AvgTemp	FromNormal	InchesRain	Bookings
7	Seattle	7	67.9	2.6	0.06	17
8	Seattle	8	66.4	0.8	0.32	17
9	Seattle	9	62.6	1.5	0.89	22
10	Seattle	10	54.3	1.6	8.96	20
11	Seattle	11	42.8	-2.4	6.77	25
12	Seattle	12	41.8	1.1	3.88	31
13	Portland	1	44.8	4.9	7.64	22
14	Portland	2	44.3	1.2	2.37	19
15	Portland	3	49	1.8	5.75	17
16	Portland	4	50.8	-0.4	4.37	18

Process Flow to make it active. Then select **Tasks ▶ Describe ▶ Summary Statistics** from the menu bar. The Summary Statistics window will open, displaying the Data page.

Assigning task roles

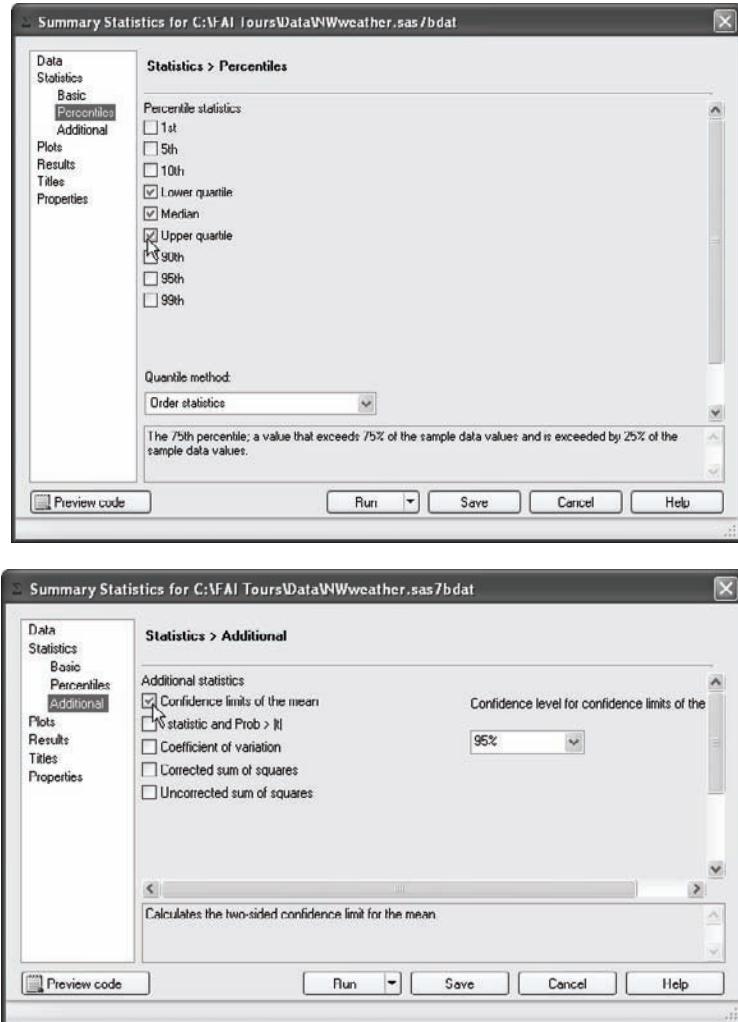
You should assign to the **Analysis variables** role all the numeric variables you want summarized. If you choose a classification variable, then you will get separate analyses for each value of the classification variable. The **Group analysis by** role produces the same result as the **Classification variables** role, but the output is formatted differently. In this example, the variables InchesRain and Bookings have been assigned to the **Analysis variables** role, and the variable City has been assigned to the **Classification variables** role.



Choosing statistics

The statistics are grouped into Basic, Percentiles, and Additional. Several statistics in the Basic page (not shown) are chosen by default: mean, standard deviation, minimum, maximum, and number of observations. In the Basic page you can also choose the maximum number of decimal values used to display the results. In the Percentiles page, you can choose from various percentile statistics, including the median and the upper and lower quartiles. The Additional page has five more statistics, including confidence limits of the mean. In the Plots page (not shown), you can request histograms and box plots of your data. In this example, the mean and number of observations have been selected in the Basic page, the median and quartiles in the Percentiles page, and the confidence limits of the mean in the Additional page. Also, the maximum decimal places have been set to 2 in the Basic page.

Results In the result, each analysis variable will have its own entry in the table, and a separate entry for each level of any classification or grouping variables. In this example, City is a classification variable so there are separate statistics for Portland and Seattle.



Summary Statistics										
Results										
The MEANS Procedure										
City	N Obs	Variable	Mean	N	Lower	Median	Upper	Lower	Upper	CL for 95% Mean
					Quartile			CL for 95% Mean	CL for 95% Mean	
Portland	12	InchesRain	3.13	12	0.58	2.69	5.06	1.38	4.87	24.38
		Bookings	20.92	12	17.50	20.00	23.50	17.45	24.38	
Seattle	12	InchesRain	3.48	12	0.70	2.25	6.56	1.39	5.57	25.16
		Bookings	21.92	12	18.00	20.50	24.00	18.67	25.16	

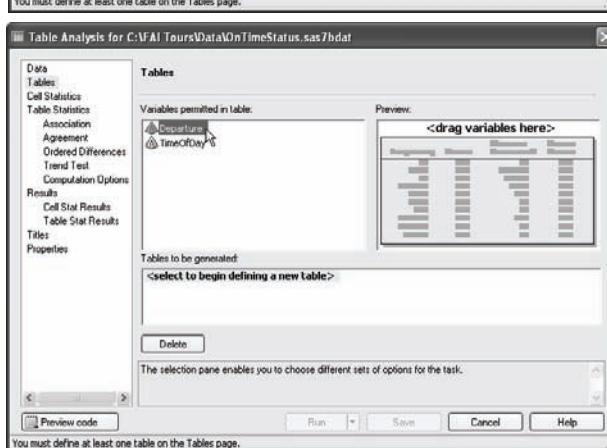
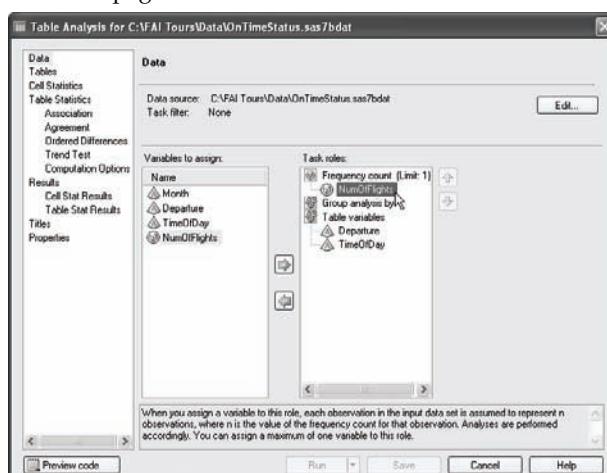
9.3 Table Analysis

The Table Analysis task produces crosstabulations and statistics for categorical data. You can choose measures of association, including chi-square, and you can request additional tests such as trend tests and measures of agreement.

Here is the OnTimeStatus data set, which shows the number of flights between Seattle and Chicago in the winter months. The data are broken down by the time of day, and whether flights had a delayed departure of 15 minutes or more. The objective is to determine if there is an association between the flights' time of day and punctuality. In the Project Tree or Process Flow, click the data icon to make it active. Then select **Tasks ► Describe ► Table Analysis** from the menu bar.

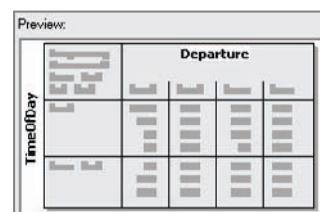
The Table Analysis window will open, displaying the Data page.

	Month	Departure	TimeOfDay	NumOfFlights
1	Dec	OnTime	AfterNoon	25
2	Dec	Late	AfterNoon	33
3	Dec	OnTime	BeforeNoon	80
4	Dec	Late	BeforeNoon	13
5	Jan	OnTime	AfterNoon	15
6	Jan	Late	AfterNoon	12
7	Jan	OnTime	BeforeNoon	30
8	Jan	Late	BeforeNoon	13
9	Feb	OnTime	AfterNoon	20
10	Feb	Late	AfterNoon	8
11	Feb	OnTime	BeforeNoon	43
12	Feb	Late	BeforeNoon	1

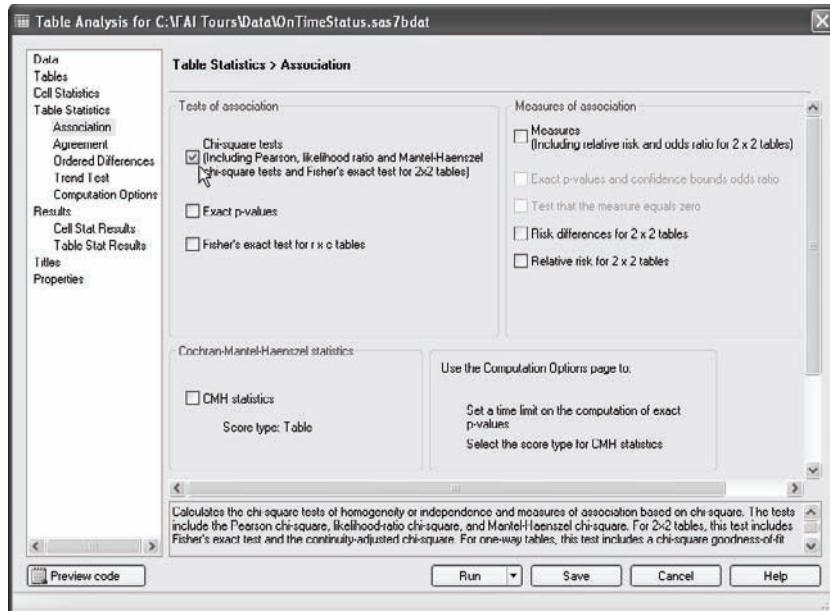


Assigning task roles For a two-way table, you must have two table variables. In this example, the variables TimeOfDay and Departure have been assigned to the **Table variables** role. Because each row in this table represents multiple flights, the variable NumOfFlights has been assigned to the **Frequency count** role. If each row in your data represents one count, then do not use the **Frequency count** role.

Creating the table To arrange your table, click the **Tables** option in the selection pane on the left. The first variable you drag to the **Preview** area will form the columns of the table. The second variable you drag will form the rows. In this example, Departure is on the top, and TimeOfDay on the side.



Choosing statistics You can choose many different statistics in the Table Analysis task. The different categories of statistics are listed under **Table Statistics** in the selection pane on the left. Click **Association** and choose the tests you want. In this example, **Chi-square tests** has been chosen.



Results The output starts with a frequency table and is followed by the tests of association, including a table for Fisher's Exact Test since this is a 2x2 table. In this example, it appears that late departures tend to be more frequent in the afternoon hours. The probability of obtaining a chi-square value this large or larger by chance alone is less than 0.0001.

Table Analysis Results The FREQ Procedure				Statistics for Table of TimeOfDay by Departure																																																																																
Table of TimeOfDay by Departure <table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Departure</th> <th></th> </tr> <tr> <th colspan="2"></th> <th>Late</th> <th>OnTime</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th rowspan="2">TimeOfDay</th> <th>Frequency</th> <td>53</td> <td>60</td> <td>113</td> </tr> <tr> <th>Col Pct</th> <td>66.25</td> <td>28.17</td> <td></td> </tr> <tr> <th rowspan="2">BeforeNoon</th> <th>Frequency</th> <td>27</td> <td>153</td> <td>180</td> </tr> <tr> <th>Col Pct</th> <td>33.75</td> <td>71.83</td> <td></td> </tr> <tr> <th>Total</th> <th>Frequency</th> <td>80</td> <td>213</td> <td>293</td> </tr> </tbody> </table>						Departure					Late	OnTime	Total	TimeOfDay	Frequency	53	60	113	Col Pct	66.25	28.17		BeforeNoon	Frequency	27	153	180	Col Pct	33.75	71.83		Total	Frequency	80	213	293	<table border="1"> <thead> <tr> <th>Statistic</th> <th>DF</th> <th>Value</th> <th>Prob</th> </tr> </thead> <tbody> <tr> <td>Chi-Square</td> <td>1</td> <td>35.5961</td> <td><.0001</td> </tr> <tr> <td>Likelihood Ratio Chi-Square</td> <td>1</td> <td>35.1538</td> <td><.0001</td> </tr> <tr> <td>Continuity Adj. Chi-Square</td> <td>1</td> <td>34.0070</td> <td><.0001</td> </tr> <tr> <td>Mantel-Haenszel Chi-Square</td> <td>1</td> <td>35.4746</td> <td><.0001</td> </tr> <tr> <td>Phi Coefficient</td> <td></td> <td>0.3486</td> <td></td> </tr> <tr> <td>Contingency Coefficient</td> <td></td> <td>0.3291</td> <td></td> </tr> <tr> <td>Cramer's V</td> <td></td> <td>0.3486</td> <td></td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">Fisher's Exact Test</th> </tr> </thead> <tbody> <tr> <td>Cell (1,1) Frequency (F)</td> <td>53</td> </tr> <tr> <td>Left-sided Pr <= P</td> <td>1.0000</td> </tr> <tr> <td>Right-sided Pr >= P</td> <td>3.428E-09</td> </tr> <tr> <td>Table Probability (P)</td> <td>2.772E-09</td> </tr> <tr> <td>Two-sided Pr <= P</td> <td>3.968E-09</td> </tr> </tbody> </table> <p>Sample Size = 293</p>				Statistic	DF	Value	Prob	Chi-Square	1	35.5961	<.0001	Likelihood Ratio Chi-Square	1	35.1538	<.0001	Continuity Adj. Chi-Square	1	34.0070	<.0001	Mantel-Haenszel Chi-Square	1	35.4746	<.0001	Phi Coefficient		0.3486		Contingency Coefficient		0.3291		Cramer's V		0.3486		Fisher's Exact Test		Cell (1,1) Frequency (F)	53	Left-sided Pr <= P	1.0000	Right-sided Pr >= P	3.428E-09	Table Probability (P)	2.772E-09	Two-sided Pr <= P	3.968E-09
		Departure																																																																																		
		Late	OnTime	Total																																																																																
TimeOfDay	Frequency	53	60	113																																																																																
	Col Pct	66.25	28.17																																																																																	
BeforeNoon	Frequency	27	153	180																																																																																
	Col Pct	33.75	71.83																																																																																	
Total	Frequency	80	213	293																																																																																
Statistic	DF	Value	Prob																																																																																	
Chi-Square	1	35.5961	<.0001																																																																																	
Likelihood Ratio Chi-Square	1	35.1538	<.0001																																																																																	
Continuity Adj. Chi-Square	1	34.0070	<.0001																																																																																	
Mantel-Haenszel Chi-Square	1	35.4746	<.0001																																																																																	
Phi Coefficient		0.3486																																																																																		
Contingency Coefficient		0.3291																																																																																		
Cramer's V		0.3486																																																																																		
Fisher's Exact Test																																																																																				
Cell (1,1) Frequency (F)	53																																																																																			
Left-sided Pr <= P	1.0000																																																																																			
Right-sided Pr >= P	3.428E-09																																																																																			
Table Probability (P)	2.772E-09																																																																																			
Two-sided Pr <= P	3.968E-09																																																																																			

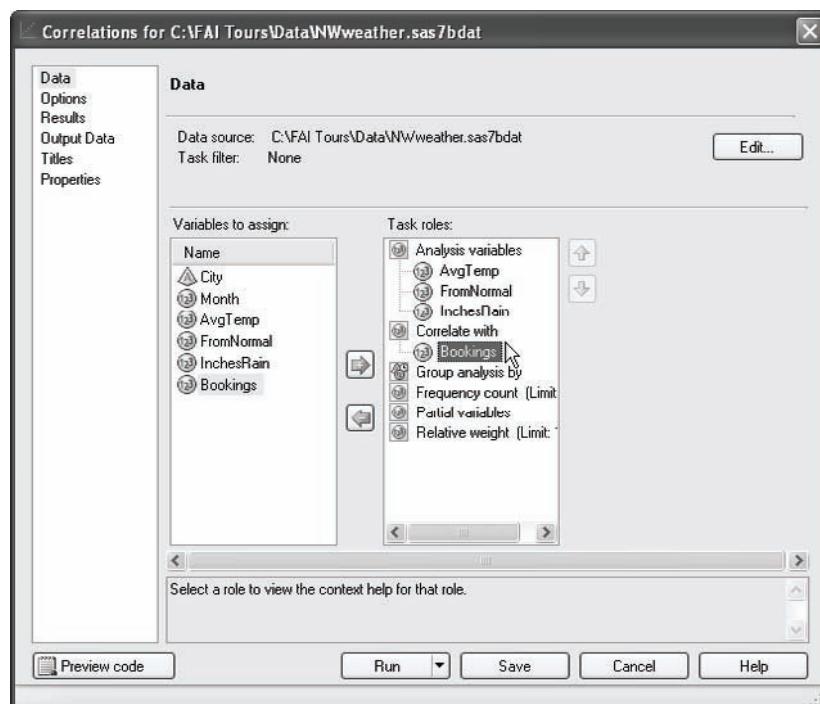
9.4 Correlations

The Correlations task produces correlation coefficients that measure relationships between numeric variables. A correlation coefficient of one means that two variables are perfectly correlated, while a correlation coefficient of zero means that there is no relationship between the two variables.

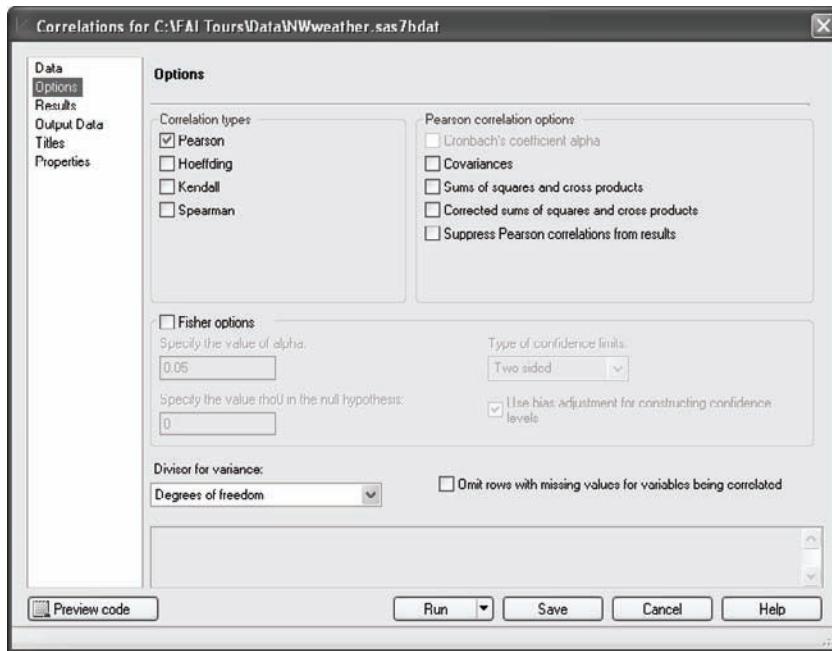
Here is a portion of the NWweather data set showing the average temperature, the deviation from the normal average temperature, the inches of rain, and the number of bookings for the Fire and Ice Tours company for each month for both Seattle and Portland. Using the Correlations task, you can measure the relationship between local weather and the number of tours booked each month. In the Project Tree or Process Flow, click the data icon to make it active. Then select **Tasks ▶ Multivariate ▶ Correlations** from the menu bar. The Correlations window will open, displaying the Data page.

Assigning task roles

For correlations, variables assigned to the **Analysis variables** role will appear across the top of the table, while variables assigned to the **Correlate with** role will appear down the side of the table. If there are no Correlate with variables, then the Analysis variables will appear both across the top and down the side of the table. In this example, the three weather variables AvgTemp, FromNormal, and InchesRain have been assigned to the **Analysis variables** role, and Bookings has been assigned to the **Correlate with** role.



Choosing statistics and plots To run a correlation, all you need to do is assign variables to the task roles. However, you may want to choose some additional statistics. The Options page allows you to choose the type of correlation: Pearson (the default), Hoeffding, Kendall, or Spearman. There are additional options for Pearson correlations. For this example, the type of correlation is set to Pearson.



If you click **Results** in the selection pane on the left, you can request plots and choose the statistics to be included in the results. When you are satisfied with all the settings, click **Run**.

Results The output starts with a list of the analysis variables, followed by simple statistics. Next are the correlation coefficients. In this example, two variables—AvgTemp and InchesRain—are correlated with the number of bookings. AvgTemp is negatively correlated (higher temperatures correspond with lower bookings), while InchesRain is positively correlated (more rain corresponds with more bookings). The FromNormal variable is not significantly correlated with Bookings.

Correlation Analysis						
The CORR Procedure						
1 With Variables:	Bookings					
3 Variables:	AvgTemp FromNormal InchesRain					
Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Bookings	24	21.41667	5.19127	514.00000	13.00000	33.00000
AvgTemp	24	54.30417	10.21284	1303	41.70000	71.60000
FromNormal	24	1.41250	1.99659	33.90000	-2.40000	4.90000
InchesRain	24	3.30417	2.97348	79.30000	0	8.96000
Pearson Correlation Coefficients, N = 24						
	Prob > r under H0: Rho=0					
	AvgTemp	FromNormal	InchesRain			
Bookings	-0.60049	0.09931	0.57642			
	0.0019	0.6443	0.0032			

9.5 Linear Regression

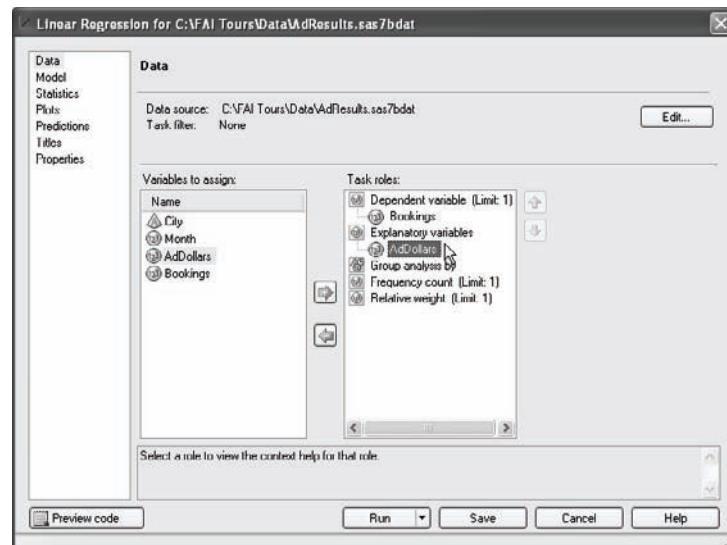
In SAS Enterprise Guide, you can perform many different types of regression analysis, and the models that you build can be quite complex. You can choose linear, nonlinear, logistic, and generalized linear models. In addition, within each type of regression, there are many options for customizing your analysis. This section shows how to do a simple linear regression with one dependent and one explanatory variable. You must have SAS/STAT software installed on your SAS server to perform regression analysis.

The Fire and Ice Tours company started a local advertising campaign in both Seattle and Portland. It wants to see if the money spent on advertising is increasing the number of tour bookings. Here is a sample of the AdResults data set with data for the dollars spent on advertising and the number of bookings for each month and city. A linear regression analysis will show if there is a relationship between dollars spent and bookings. In the Project Tree or Process Flow, click the data icon to make it active. Then select **Tasks ▶ Regression ▶ Linear** from the menu bar. The Linear Regression window will open, displaying the Data page.

	City	Month	AdDollars	Bookings
8	Seattle	8	250	17
9	Seattle	9	250	22
10	Seattle	10	325	20
11	Seattle	11	400	25
12	Seattle	12	500	31
13	Portland	1	325	25
14	Portland	2	290	19
15	Portland	3	250	17
16	Portland	4	300	18

Assigning task roles

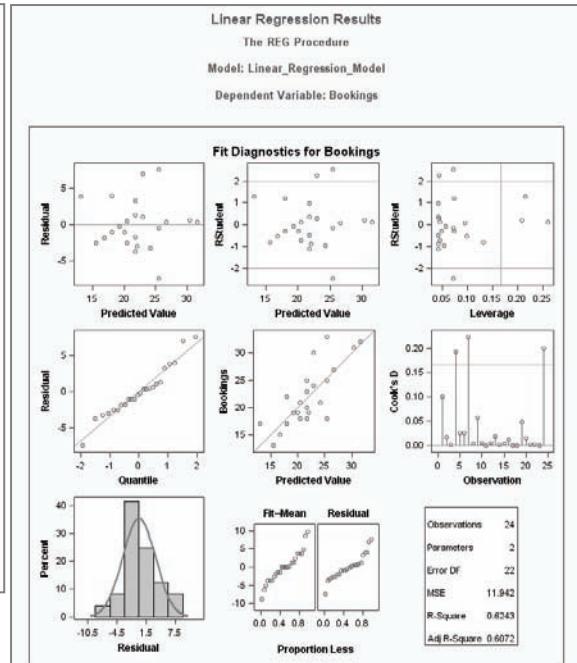
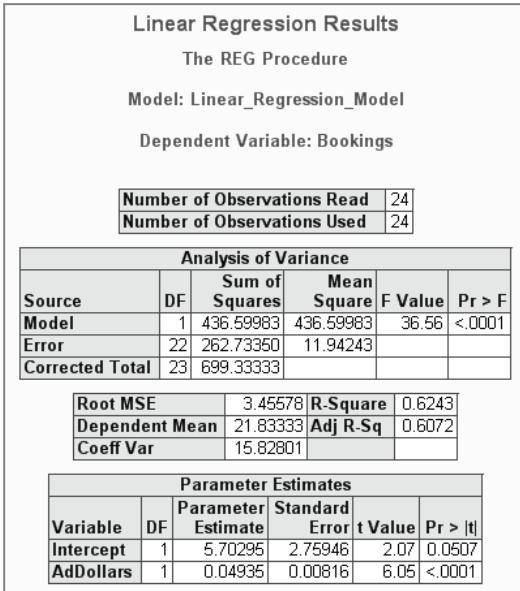
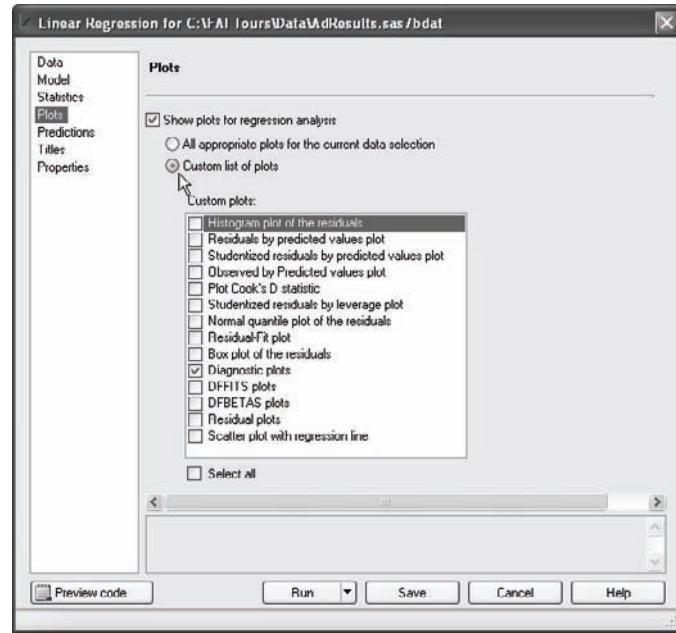
For a simple linear regression, you must assign one variable to the **Dependent variable** role, and one to the **Explanatory variables** role. Both the dependent and the explanatory variables must be numeric. This example tests whether the number of bookings can be explained by the dollars spent on advertising. So, the variable Bookings has been assigned to the **Dependent variable** role, and the variable AdDollars has been assigned to the **Explanatory variables** role.



Choosing statistics Because you can perform many different types of regression analysis using this task, there are a lot of options listed in the selection pane on the left. In the Model page, you can choose the model selection method, including forward, backward, stepwise, and several methods based on R-squared. The Statistics page gives choices for additional statistics, including details on estimates, correlations, and diagnostics. For this simple example, there is no need to change the model type, or to request additional statistics.

Selecting plots By default, all appropriate plots are produced for your data and model. To control which plots are produced, select **Custom list of plots** in the Plots page to display all available plots. Then check the desired plots. Diagnostic plots are selected by default.

Results The results of the regression analysis start with the number of observations used for the analysis, followed by the Analysis of Variance table, statistics, and Parameter Estimates. In this example, the model is significant with a *p*-value of less than 0.0001, which means that the number of dollars spent on advertising can be used to explain some of the variation in Bookings. The diagnostic plots show eight different plots on the same page along with basic regression statistics.



9.6 Analysis of Variance

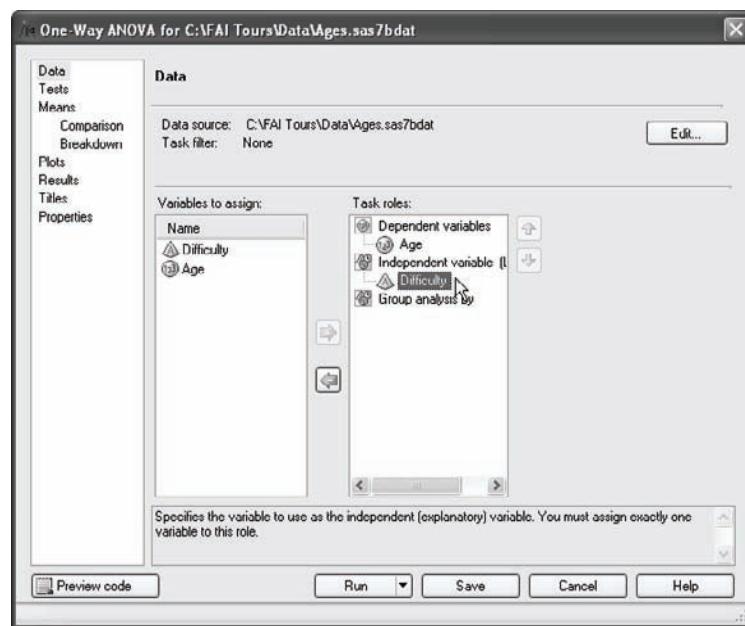
SAS Enterprise Guide can perform several types of analysis of variance, including one-way ANOVA and nonparametric one-way ANOVA, as well as mixed and linear models. This section shows the One-Way ANOVA task, which performs analysis of variance tests, and comparisons of means. You must have SAS/STAT software installed on your SAS server to use any of the ANOVA tasks.

The Fire and Ice Tours company wants to know if the tours it offers, with the three difficulty ratings, attract customers from different age groups. Ten customers were surveyed in each of the three difficulty categories to find their ages. Here is a sample of the resulting data set, Ages. In the Project Tree or Process Flow, click the data icon to make it active. Then select **Tasks ▶ ANOVA ▶ One-Way ANOVA** from the menu bar. The One-Way ANOVA window will open, displaying the Data page.

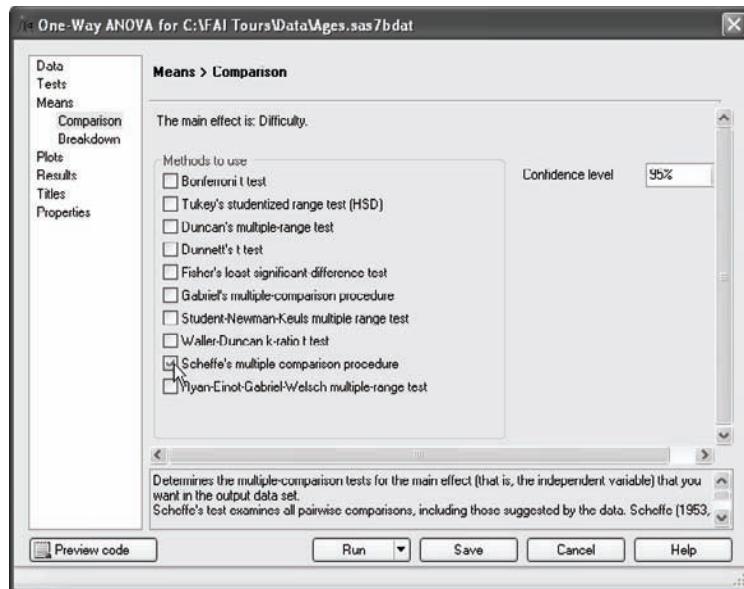
	Difficulty	Age
7	e	66
8	e	57
9	e	39
10	e	33
11	m	26
12	m	37
13	m	42
14	m	27

Assigning task roles

For one-way ANOVA, you must assign one variable to the **Dependent variables** role, and one to the **Independent variable** role. The dependent variable is a numeric variable whose means you want to test. The independent variable determines the different categories. In this example, the variable Age has been assigned to the Dependent variables role. The variable Difficulty has been assigned to the Independent variable role. If you want to test more than one variable at a time, you can assign several variables to the Dependent variables role, but each variable will be analyzed separately.



Choosing statistics The One-Way ANOVA task offers several groups of options in the selection pane on the left. In the Tests page, you can select tests for equal variance. In the Plots page, you can request box-and-whisker or means plots. You can choose descriptive statistics for the dependent variables in the Breakdown page.



One-Way Analysis of Variance

Results

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Difficulty	3	c e m

Number of Observations Read	30
Number of Observations Used	30

Page Break

One-Way Analysis of Variance

Results

The ANOVA Procedure

Dependent Variable: Age

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1681.400000	840.700000	6.41	0.0053
Error	27	3541.800000	131.177778		
Corrected Total	29	5223.200000			

R-Square	Coeff Var	Root MSE	Age Mean
0.321910	24.68381	11.45329	46.40000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Difficulty	2	1681.400000	840.700000	6.41	0.0053

Comparison of means

To do any comparison of means tests, choose them in the Comparison page. In this example, **Scheffe's multiple comparison procedure** has been selected.

Results The output starts with a table giving the number of classes (categories) and the number of observations in the data. Next is the result of the analysis of variance, followed by the results of Scheffe's test. Scheffe's test includes the comparison

of the means between the three levels of difficulty. Letters are used to group the means, where means labeled with different letters are significantly different from each other. In this case, people in the challenging tours are significantly older than people in the moderate tours. However, while people in the challenging tours are also older than people in the easy tours, they are not significantly older. In this example, the *p*-value of 0.0053 shows that the overall model is also significant.

Scheffe's Test for Age					
Note: This test controls the Type I experimentwise error rate.					
Alpha					0.05
Error Degrees of Freedom					27
Error Mean Square					131.1778
Critical Value of F					3.35413
Minimum Significant Difference					13.266
Means with the same letter are not significantly different.					
Scheffe Grouping					
A					54.200 10 c
A					
B					48.700 10 e
B					
B					36.300 10 m

10

“ There is magic in graphs.
The profile of a curve reveals
in a flash a whole situation....
The curve informs the mind,
awakens the imagination,
convinces. ”

HENRY D. HUBBARD



CHAPTER 10

Producing Graphs

- 10.1 Bar Charts 298
- 10.2 Pie Charts 300
- 10.3 Simple Line Plots 302
- 10.4 Multiple Line Plots by Group 304
- 10.5 Scatter Plots 306
- 10.6 Controlling the Axes 308

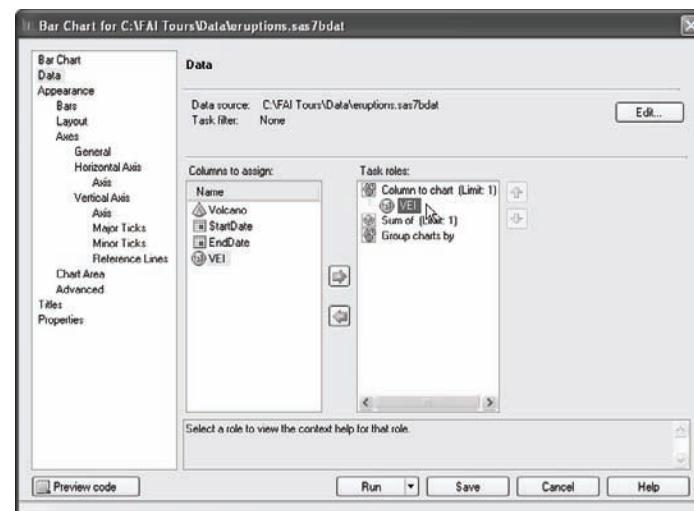
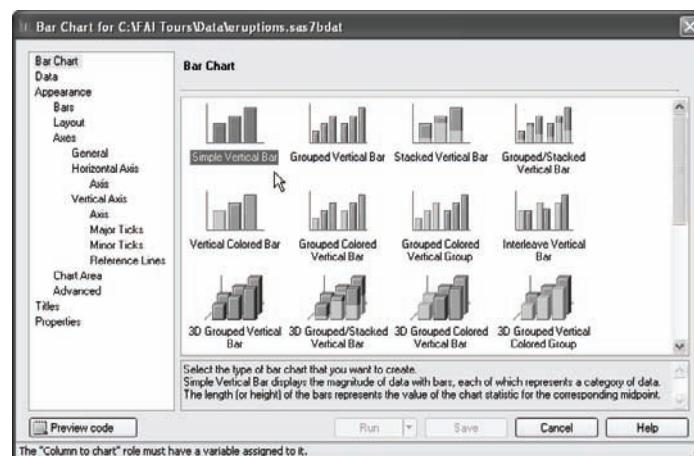
10.1 Bar Charts

Bar charts can be an effective way to present data when you want to show the frequency, percentage, sum, or mean of values in your data, broken into groups. Bar charts can be created using the Bar Chart task or wizard. This section describes the task. You must have SAS/GRAF software installed on your SAS server to create any of the graphs shown in this chapter.

The Eruptions data set contains the volcano name and the Volcanic Explosivity Index (VEI) for selected eruptions. Here is a sample of the data. Notice that the column VEI ranges in value from 0 to 6. From these data, you can create a bar chart showing the number of eruptions for each value of VEI. In the Project Tree or Process Flow, click the data icon to make it active, and then select **Tasks ▶ Graph ▶ Bar Chart**. The Bar Chart window will open, displaying the Bar Chart page.

Choosing the type of bar chart There are many different types of bar charts you can create. Scroll down to see all the available types. For this example, select **Simple Vertical Bar**, then click **Data** in the selection pane on the left to assign columns to roles.

Volcano	StartDate	EndDate	VEI
Grimsvotn	12/18/1998	12/28/1998	3
Kilauea	05/30/1840	06/25/1840	0
Kilauea	05/24/1969	07/22/1974	0
Kliuchevskoi	09/25/1737	11/04/1737	2
Kliuchevskoi	03/25/1931	03/27/1931	4
Kliuchevskoi	01/20/2005	04/07/2005	2
Krakatau	05/20/1883	10/21/1883	6
Krakatau	07/04/1938	07/02/1940	3
	05/29/2000	10/20/2000	1

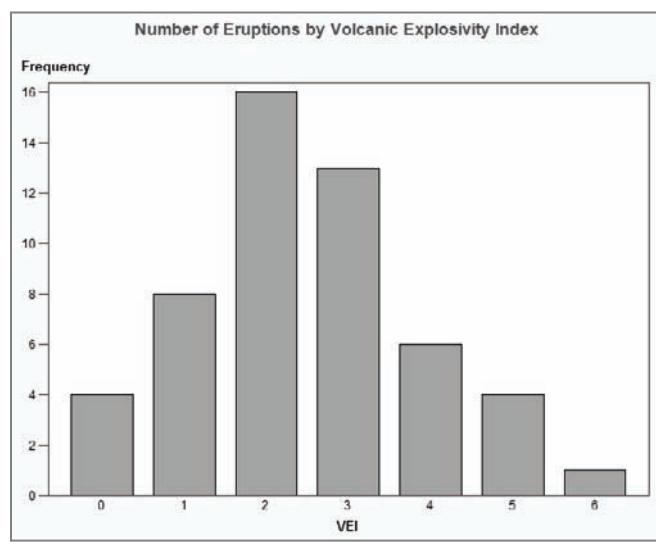
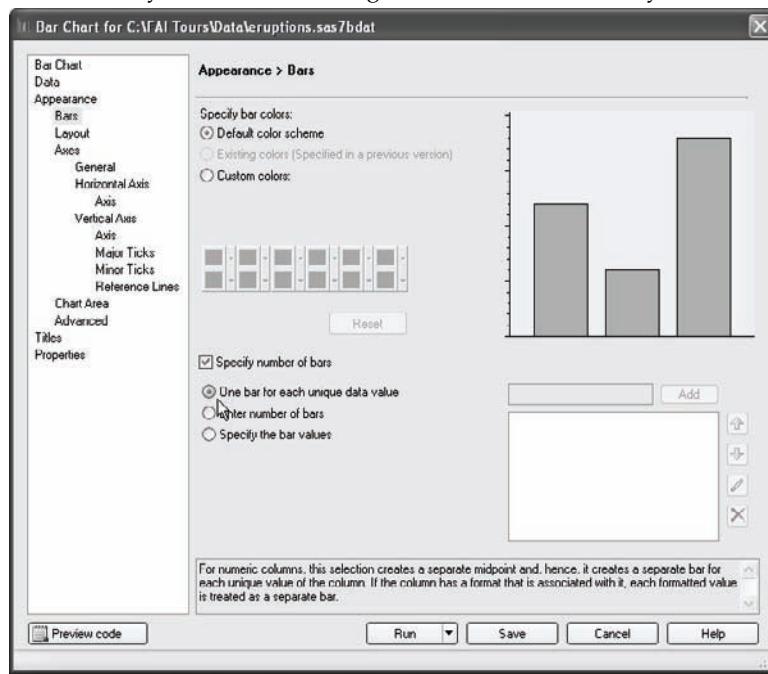


Customizing a bar chart As you can see from the list of options in the selection pane on the left, there are many options for bar charts. You can make changes to the axes and tick marks, add or remove reference lines, specify the size of your chart, change the background color, choose the shape and size of the bars in your chart, and change the order of the bars. If you want to display the mean instead of the sum (or the percentage instead of the frequency), you can change the statistic in the Advanced page.

Options in the Bars page allow you to change the color of the bars, and to specify the number of bars. If the column you are charting is character, then you will get a bar for each value of the column. If your column is numeric, then SAS Enterprise Guide will determine an appropriate number of bars for your chart, and label the bars with the midpoint in the range.

You can override this behavior by specifying the number of bars. In this example, the column charted—VEI—is numeric, but it has only seven discrete values. So, it makes sense to have one bar for each value. To do this, check **Specify number of bars**, and select **One bar for each unique data value**. When you are satisfied with the options, click **Run**.

Results Here is the resulting bar chart of VEI. There is a bar for each unique value of VEI, and the lengths of the bars show a simple count of eruptions for each value of VEI.



10.2 Pie Charts

Pie charts are similar to bar charts. With both types of charts, you can show the frequency, percentage, sum, or mean of values in your data, broken into groups. The type of chart you use depends on your personal preference. Pie charts can be created using the Pie Chart task or wizard. This section describes the task.

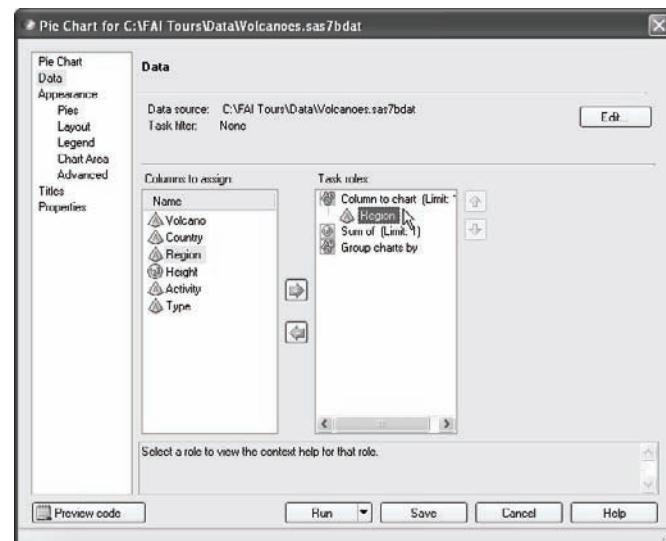
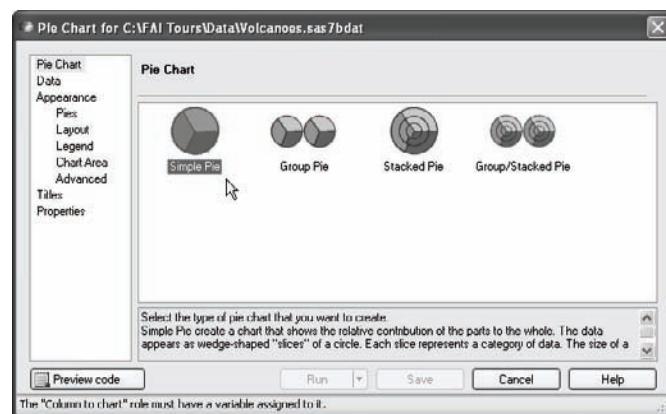
To create a pie chart of the Volcanoes data, click the data icon in the Project Tree or Process Flow to make it active. Then select **Tasks ▶ Graph ▶ Pie Chart**. This opens the Pie Chart window, displaying the Pie Chart page.

Choosing the type of pie chart

Before assigning columns to task roles, you must select the type of pie chart you want to create. If you want one pie chart that shows the frequency or percentage of rows that fall into different categories, then select Simple Pie. If you have a grouping column and you want a separate pie for each level of the group, then you may want to select Group Pie. The Stacked Pie is similar to a Group Pie, except the pies are stacked one on top of the other instead of side by side. Use a Group/Stacked Pie if you have more than one grouping column. In this example, select Simple Pie and then click Data in the selection pane on the left to assign roles to columns.

Assigning task roles For simple pie charts, all you need is a column to chart. Drag the column you want to summarize to the **Column to chart** role. Use the **Sum of** role if you want the size of the pie slice to represent a sum or mean value, instead of simple counts or percentages. The **Group charts by** role is similar to using a Group Pie

Volcano	Country	Region	Height	Activity	Type
Altar	Ecuador	SA	5321	Extinct	Stratovolcano
Arthur's Seat	UK	Eu	251	Extinct	
Baren Island	India	As	354	Active	Stratovolcano
Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
Erebus		An	3794	Active	Stratovolcano
Etna	Italy	Eu	3350	Active	Stratovolcano
Fuji	Japan	As	3776	Active	Stratovolcano
Ganibaldi	Canada	NA	2678		Stratovolcano
Grimsvötn	Iceland	Eu	1725	Active	Caldera
Himawari	Bolivia	SA	6450	Extinct	Stratovolcano



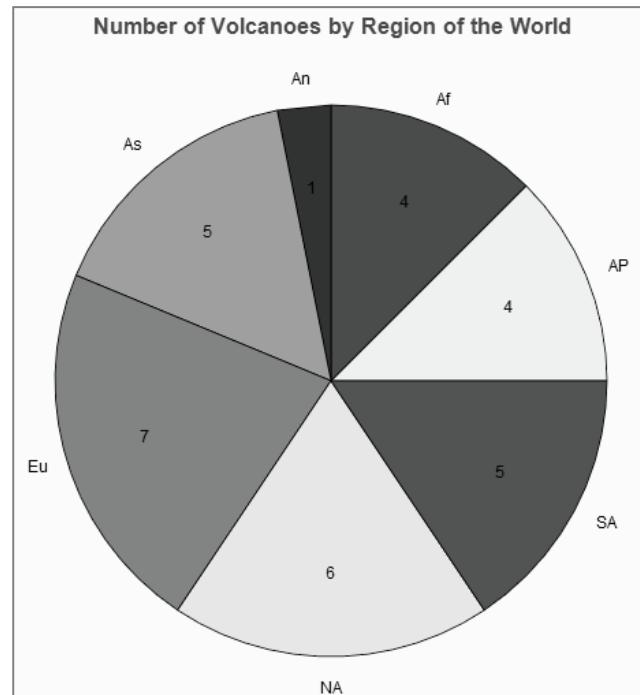
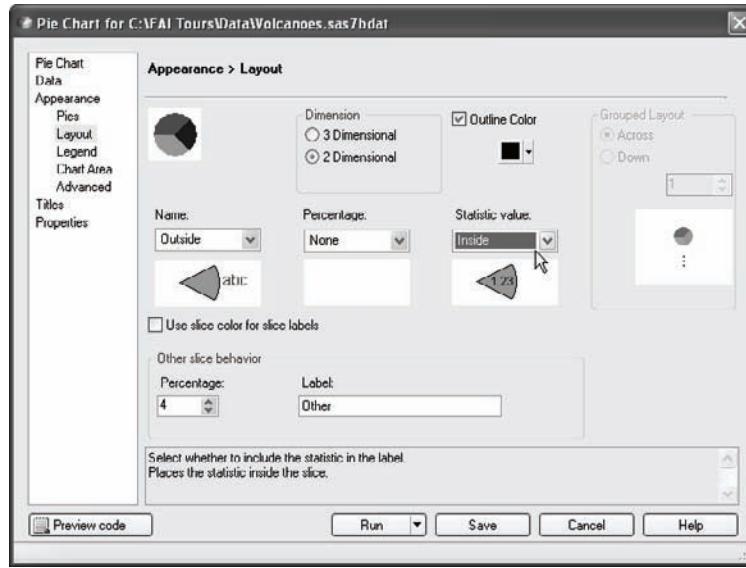
chart, except that selecting a Group Pie will produce pies side by side on the same page, whereas the Group charts by role will produce one pie per page. In this example, Region has been assigned to the Column to chart role.

Other chart options

In the Pies page, you can specify the color scheme to use for the chart, as well as the number of slices for the pie. In the Layout page, you can make the pies two- or three-dimensional, control placement of labels on the pies, specify how pies are placed on the page for grouped pies, and control the “other” slice of pie. Use options in the Legend page to control the legend, the Chart Area page to control the size and background color of the chart, and the

Advanced page to specify if the pie slices should be frequencies or percentages. In this example, in the Layout page, the position of the label for the **Statistic value** has been changed from Outside (the default) to Inside. When you are satisfied with the options, click Run.

Results Here is the result of the Pie Chart task. Because Region is the Column to chart, the size of each pie slice shows the relative number of volcanoes from the region. The label for the statistic value was specified to be inside, so the actual number of volcanoes appears inside each pie slice.



10.3 Simple Line Plots

There are many different types of line plots you can produce using SAS Enterprise Guide. You can plot one or several plots on the same graph, and you can specify different types of interpolation for your plot. If you use a type of interpolation for your plot where the points are connected, make sure that it makes sense to connect the points. Line plots can be created using the Line Plot task or wizard. This section describes the task.

Here is a portion of the NWweather data set which contains data for both Seattle and Portland. For this example, we want to use only the data for Seattle, so a filter will be applied to the data.

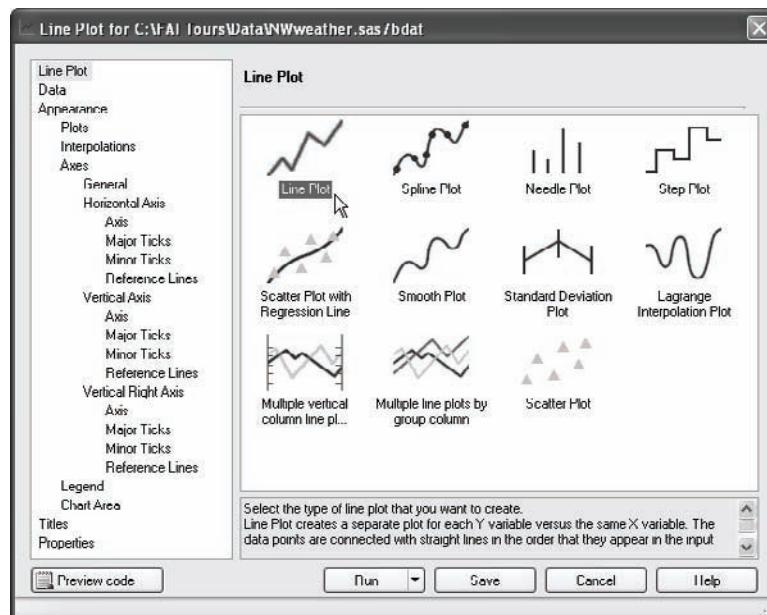
	City	Month	AvgTemp	FromNormal	InchesRain	Bookings
7	Seattle	7	67.9	2.6	0.06	17
8	Seattle	8	66.4	0.8	0.32	17
9	Seattle	9	62.6	1.5	0.89	22
10	Seattle	10	54.3	1.6	8.96	20
11	Seattle	11	42.8	-2.4	6.77	25
12	Seattle	12	41.8	1.1	3.88	31
13	Portland	1	44.8	4.9	7.64	22
14	Portland	2	44.3	1.2	2.37	19
15	Portland	3	49	1.8	5.75	17

To create a line plot, click the data icon in the Project Tree or Process Flow to make it active. Then select Tasks ▶ Graph ▶ Line Plot. This opens the Line Plot window displaying the Line Plot page.

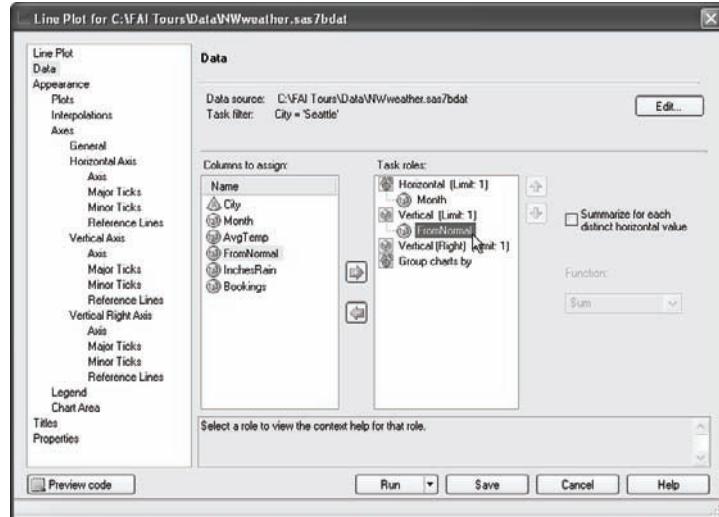
Choosing the type of line plot

Before assigning columns to task roles, you must select the type of line plot you want to create.

Choosing the plot type will automatically change other settings in the Line Plot window to fit that type of plot. For this example, select **Line Plot**, and then click **Data** in the selection pane on the left to assign columns to roles.



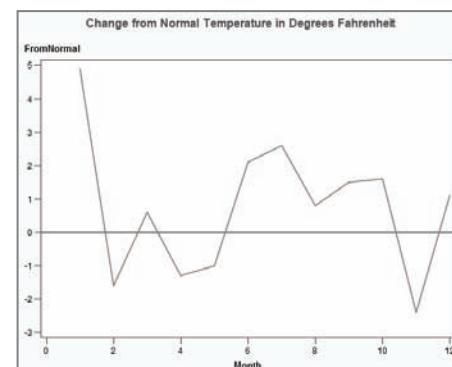
Assigning task roles For a basic line plot, you must assign one column to the **Horizontal** role and one to the **Vertical** role. In this example, Month has been assigned to the **Horizontal** role and FromNormal to the **Vertical** role. To select only rows for Seattle, click **Edit** to open the Edit Data and Filter window and apply the filter **City = 'Seattle'**. For more information about filtering data in a task, see section 5.1.



Other plot options The selection pane on the left lists many categories of options. You can specify the line style, the interpolation method, axes and tick marks, reference lines, the legend, and the size and background color of your plot. For this example, click **Reference Lines** for the Vertical Axis in the selection pane on the left. Then check the boxes labeled **Use reference lines** and **Specify values for lines**. Type the number 0 into the box next to the Add button, and click the **Add** button. When you are satisfied with the options, click **Run**.



Results Here is the result of the Line Plot task. Month is on the horizontal axis and FromNormal on the vertical. The reference line on the vertical axis is set at zero.



10.4 Multiple Line Plots by Group

Sometimes you may want to put more than one line on a plot. If you have a grouping column, then you can create a plot that has separate lines for each value of the grouping column. For example, you may have sales figures by month for different regions and you want one plot of sales by month with a line for each region.

Here is a portion of the NWweather data set. To create a line plot of the data with separate lines for Seattle and Portland, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Graph ▶ Line Plot**. This opens the Line Plot window displaying the Line Plot page.

Choosing the type of line plot Before assigning columns to task roles, you must select the type of line plot you want to create. In this example, select **Multiple line plots by group column**, and then click **Data** in the selection pane on the left to assign columns to roles.

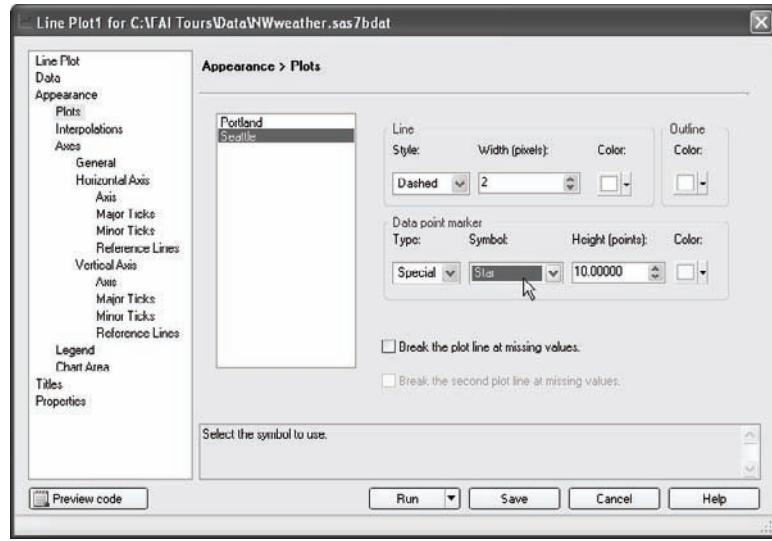
Assigning task roles For this type of plot, you need to assign one column to the **Horizontal** role and one to the **Vertical** role. Then assign the grouping column to the **Group** role. There will be one line for each unique value of the grouping column. In this example, Month has been assigned to the Horizontal role, AvgTemp to the Vertical role, and City to the Group role.

The top screenshot shows the 'Line Plot' selection pane with various plot types like Line Plot, Spline Plot, etc., and the 'Multiple vertical column line pl...' option highlighted. The bottom screenshot shows the 'Data' selection pane where 'Month' is assigned to the Horizontal role, 'AvgTemp' to the Vertical role, and 'City' to the Group role.

City	Month	AvgTemp	FromNormal	InchesRain	Bookings
Seattle	7	67.9	2.6	0.06	17
Seattle	8	66.4	0.8	0.32	17
Seattle	9	62.6	1.5	0.89	22
Seattle	10	54.3	1.6	8.96	20
Seattle	11	42.8	-2.4	6.77	25
Seattle	12	41.8	1.1	3.88	31
Portland	1	44.8	4.9	7.64	22
Portland	2	44.3	1.2	2.37	19
Portland	3	49	1.8	5.75	17

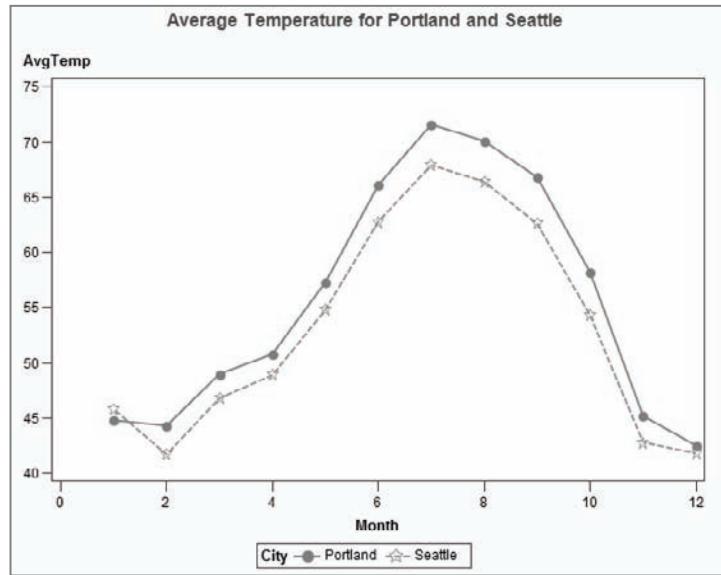
Other plot options

As with simple line plots, there are many options you can specify, such as the type of interpolation, axes and tick marks, reference lines, legends, and the size and background color of your plot. For this example, click **Plots** in the selection pane on the left. By default, all lines are solid, with no data point markers, but different colors. In the **Plots** page, you can assign different symbols and line styles to the



values of the grouping column, making it easier to identify the lines. There are two values for the grouping variable City: Portland and Seattle. Click the city name to choose styles for that city, and select the desired attributes for the line and data point markers. In this example, Seattle will have a dashed line with the star symbol, and Portland a solid line with the dot symbol. When you are satisfied with the options, click **Run**.

Results Here is the result of the Line Plot task. Month is on the horizontal axis and AvgTemp on the vertical. There is a separate line for each city.



10.5 Scatter Plots

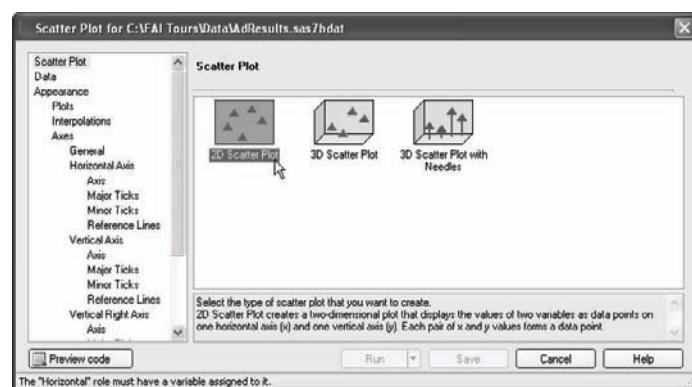
Scatter plots are similar to line plots, but in a scatter plot you do not connect the points. You may simply display the points without any interpolation, or you may add a regression line showing the relationship between two columns. You can produce scatter plots with regression lines using regression analysis, but the Scatter Plot task gives you more control over the appearance of your plot.

This example produces a simple scatter plot using the AdResults data set. This data set contains the amount spent on advertising and the number of tour bookings for each month. To create a scatter plot of the data, click the data icon in the Project Tree or Process Flow to make it active. Then select **Tasks ▶ Graph ▶ Scatter Plot**. This opens the Scatter Plot window, displaying the Scatter Plot page.

	City	Month	AdDollars	Bookings
8	Seattle	8	250	17
9	Seattle	9	250	22
10	Seattle	10	325	20
11	Seattle	11	400	25
12	Seattle	12	500	31
13	Portland	1	325	25
14	Portland	2	290	19
15	Portland	3	250	17
16	Portland	4	300	18

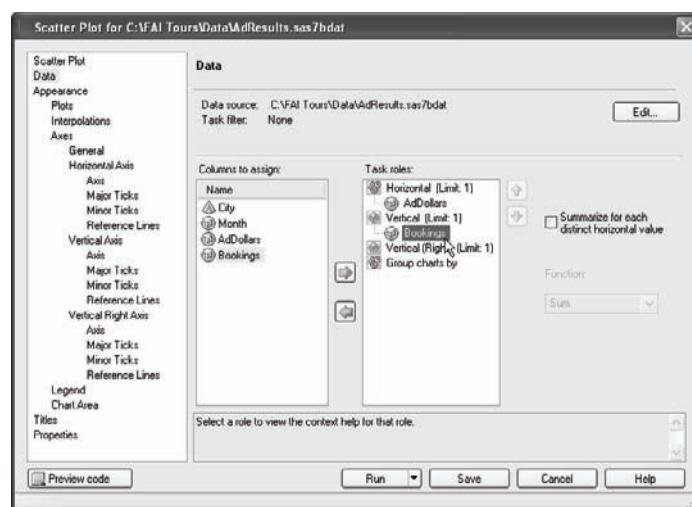
Choosing the type of scatter plot

Before assigning columns to task roles, you must select the type of scatter plot you want to create. You can choose a two- or three-dimensional scatter plot, or a three-dimensional needle plot. For this example, select **2D Scatter Plot**, and then click **Data** in the selection pane on the left to assign columns to roles.

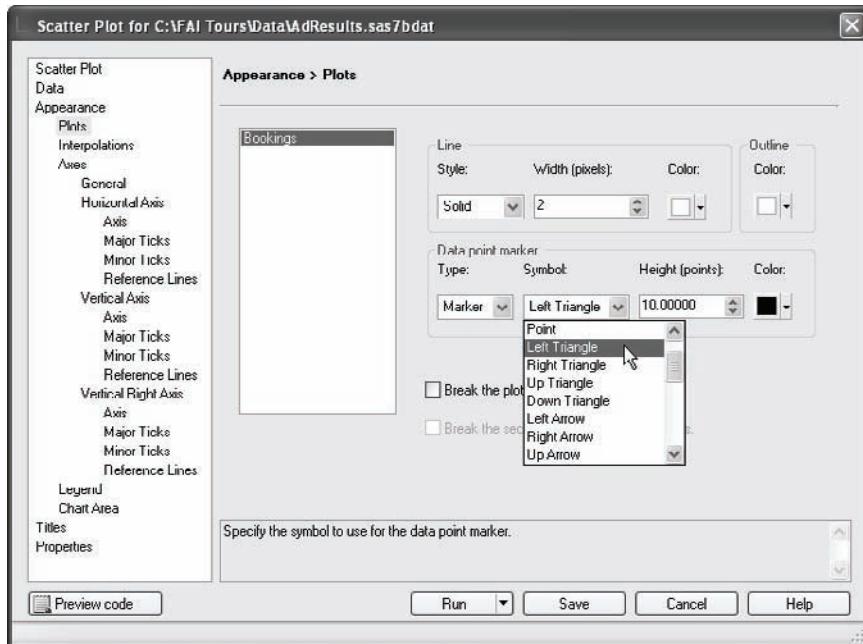


Assigning task roles

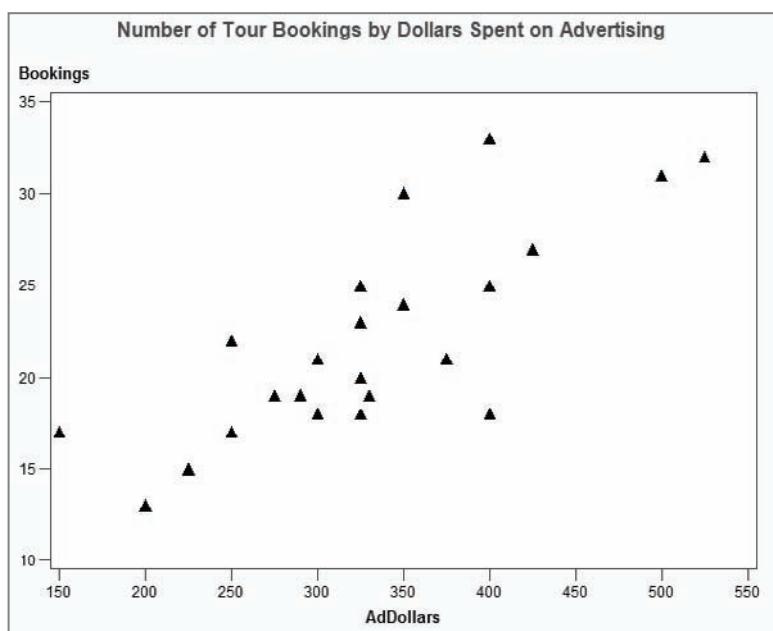
For a two-dimensional scatter plot, you must assign one column to the **Horizontal** role and one to the **Vertical** role. In this example, AdDollars is assigned to the horizontal axis, and Bookings to the vertical axis.



Other plot options As with line plots, there are many options you can specify, such as the type of interpolation (including regression), axes and tick marks, reference lines, and legends. Options in the Chart Area page control the size and background color of your plot. In the Plots page, you can select line styles and point markers. In this example, the data point type is set to Marker, the symbol is Left Triangle and the color is black. When you are satisfied with the options, click Run.



Results Here is the result of the Scatter Plot task. AdDollars is on the horizontal axis and Bookings is on the vertical. A triangle symbol is used for plotting the points instead of the default circle.

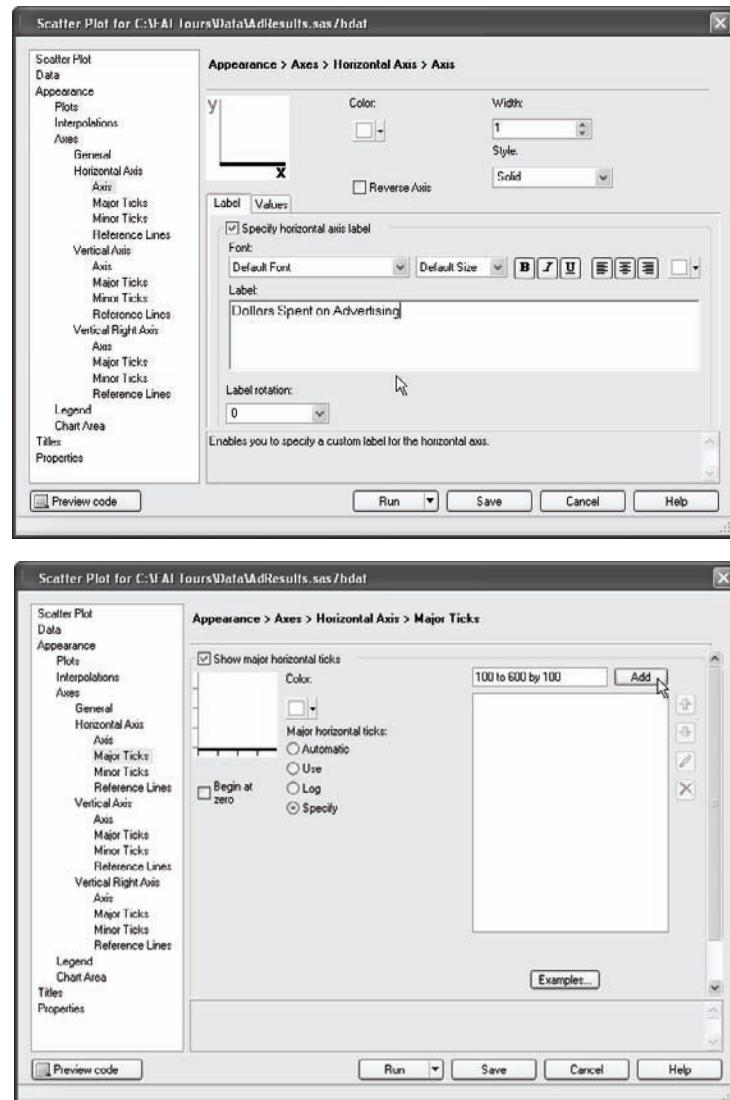


10.6 Controlling the Axes

Most types of plots produced by SAS Enterprise Guide have axes. You can control the values used for tick marks on the axes, and the style and content of the labels. There are several possible types of axes including Horizontal, Vertical, and Vertical Right. In the selection pane on the left of the task window, there will be up to four groups of options for each axis type: Axis, Major Ticks, Minor Ticks, and Reference Lines.

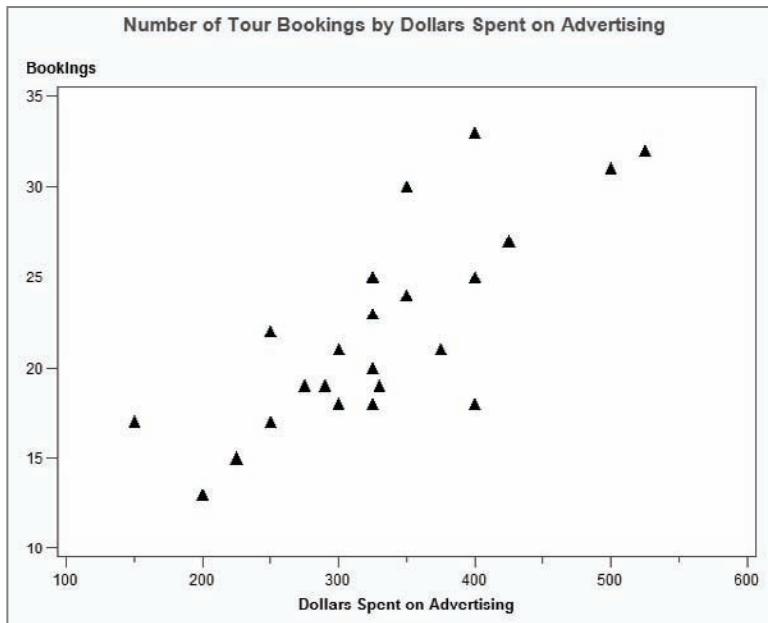
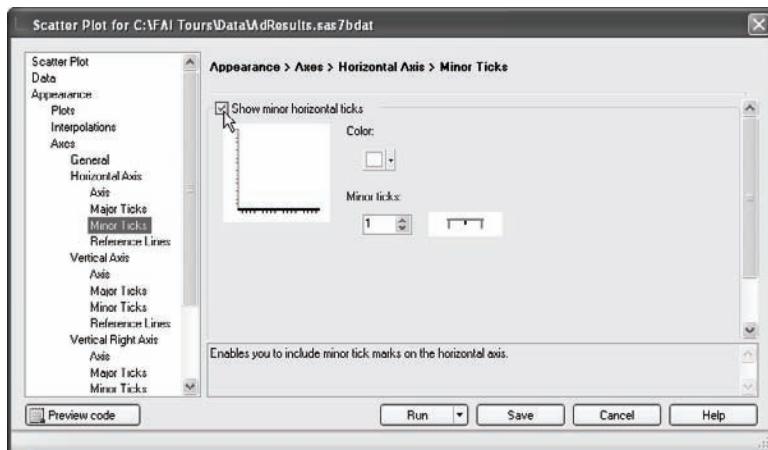
Changing the font, size, and style In the Axis page, you can specify the line color, width, and style for the axis. On the **Label** tab in the Axis page, you can change the label used for your axis including the font, style, and rotation. On the **Values** tab, you can control the font, style, and rotation of the text used to label the tick marks. In this example, the horizontal axis has been given a new label.

Setting the major tick marks Set the major ticks in the Major Ticks page for the axis. By default, the Automatic option is used to determine major ticks. If you select Use, then you can specify the number of ticks to use for the axis. Selecting Log will convert the axis to a log scale and you can choose which base to use for the log: base 2, 10, e, or pi. Selecting Specify allows you to enter your own values for the axis. In the text box to the left of the Add button, enter the desired values and click Add. The values you enter will appear in the text box below.



You can specify single values or ranges. Clicking the **Examples** button will display different methods for specifying values. In this example, the tick marks have been specified by typing **100 to 600 by 100** in the text box and clicking **Add**. This will produce tick marks displayed starting at 100, at each 100 interval, and ending at 600.

Adding minor tick marks By default, no minor tick marks are displayed. To set minor tick marks, open the Minor Ticks page for the axis, check **Show minor ticks**, then choose the number of minor ticks. In this example, one minor tick mark will appear on the horizontal axis.



Results Here are the results of the scatter plot from the previous section with changes to the horizontal axis.

11

“ Style is the dress of thoughts. ”

EARL OF CHESTERFIELD

From “On Education” in *Letters to His Son on the Art of Becoming a Man of the World and Gentleman*. Nov. 24, 1749. As quoted in *The Cyclopedie of Practical Quotations: English, Latin, and Modern Foreign Languages* by Jehiel Keeler Hoyt, 1896.



CHAPTER 11

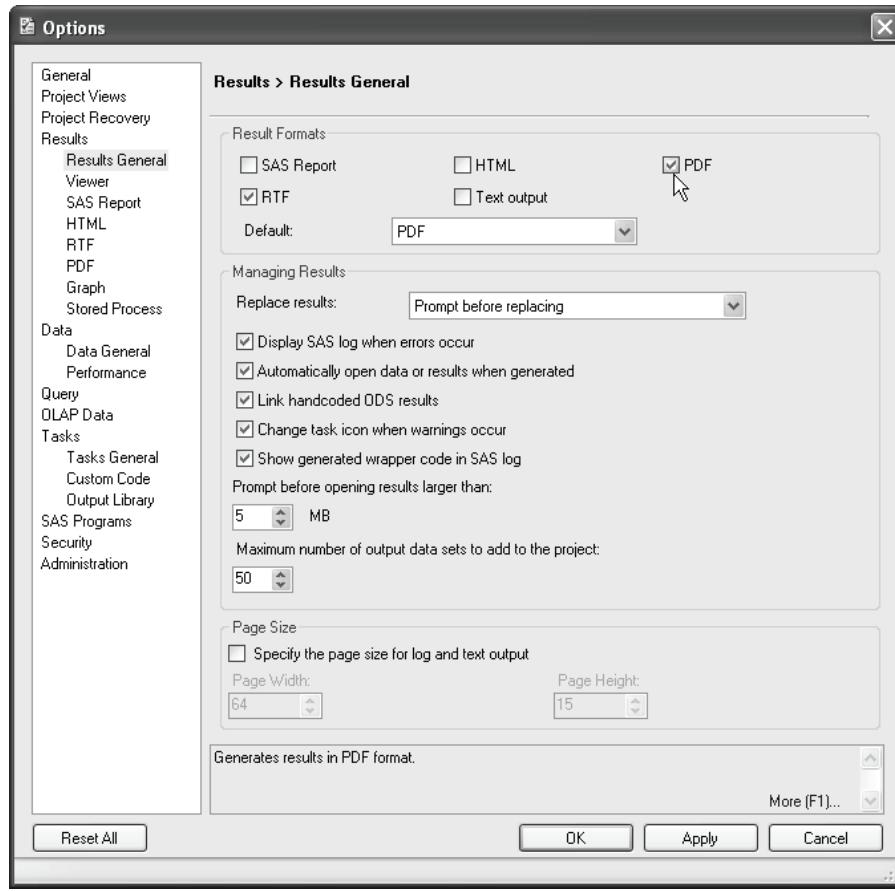
Changing Result Styles and Formats

- 11.1 Changing the Result Format 312
- 11.2 Changing the Result Style 314
- 11.3 Customizing Styles Using the Style Manager 316
- 11.4 Combining Results into a Single Document 318
- 11.5 Adding Text, Images, and Headings to Reports 320
- 11.6 Exporting Results to a File 322

11.1 Changing the Result Format

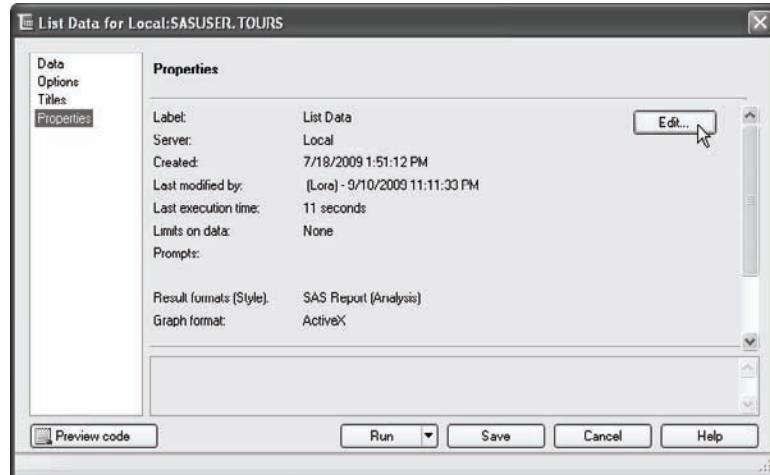
When you run a task that produces output, by default the result is in SAS Report format. SAS Report format can be used to combine several results into one customized report, which can be exported to other file formats if desired. You can also create results directly in other formats. SAS Enterprise Guide can produce results in RTF, which can be opened in Microsoft Word; PDF, which can be opened using Adobe Acrobat or Adobe Reader; HTML, which can be opened in a web browser; and plain text, which can be opened in any text editor.

Setting the default result format To set the default result format for tasks, select **Tools ► Options** from the menu bar. This opens the Options window. Click **Results General** in the selection pane on the left to open the Results General page of options. The available result formats are listed under the heading **Result Formats** in this page. Select one or more formats by clicking the box next to the format. Once you make this change, it will affect all subsequent results. In this example, because both PDF and RTF are checked, any task that you run will produce results in both PDF and RTF format. Because PDF is selected as the default, the PDF Results tab will display on top in the workspace.

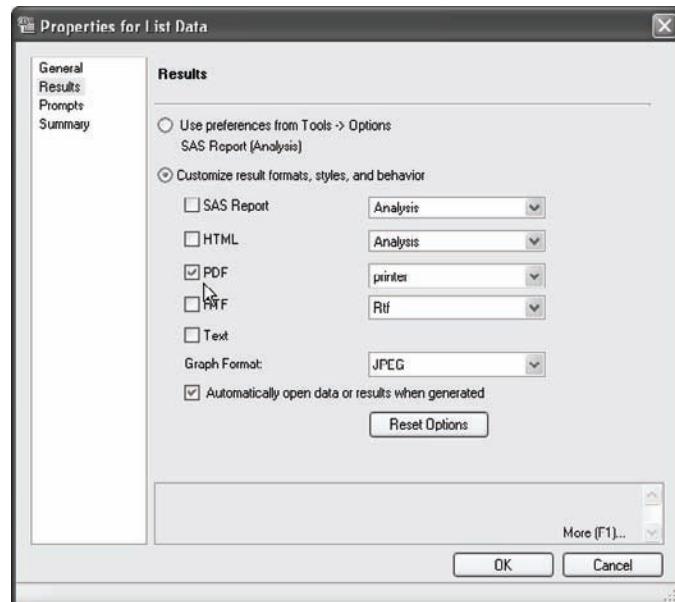


Changing the result format of a task

If you change the result format in the Options window, then the results of all tasks are affected. If you want to change result formats for individual tasks, then click **Properties** in the selection pane for the task, and then click the **Edit** button.



This opens the Properties window for the task. Click **Results** in the selection pane on the left to open the Results page, and then select **Customize result formats, styles, and behavior**. Now you can choose result formats and styles for the task's results. On this same page, you can select the Graph Format for any graphical output the task may produce. In this example, this task will produce results in PDF format only, and any graphs produced will be in JPEG format.



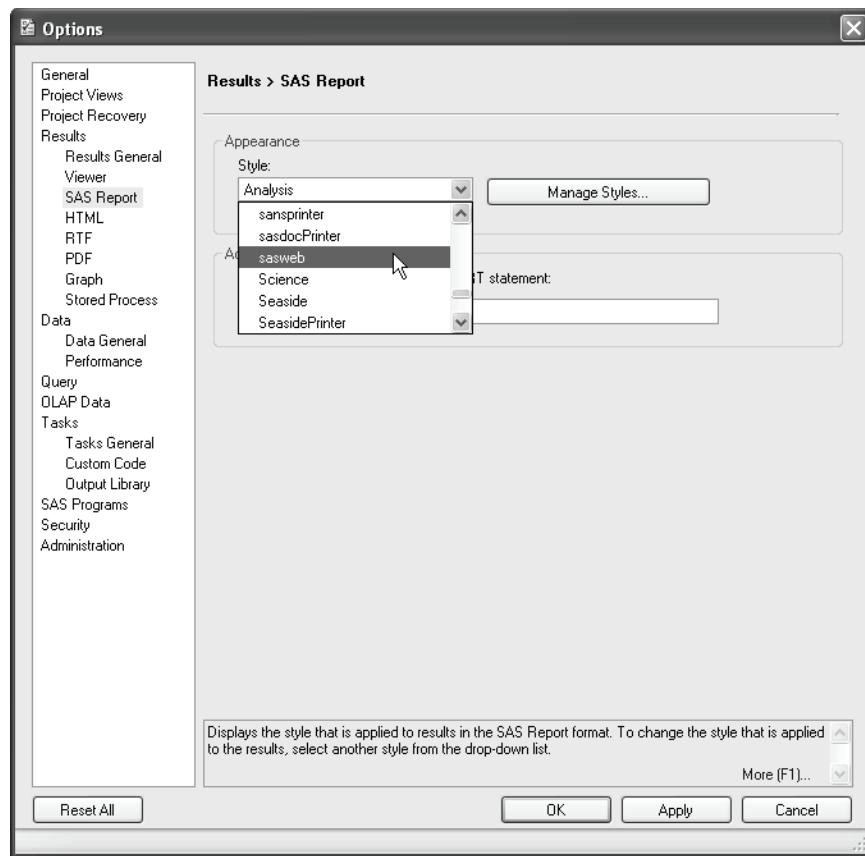
<i>Volcano Tours</i>				
Volcano	Departs	Days	Price	Difficulty
Etna	Catania	7	\$1,075	Moderate
Fuji	Tokyo	2	\$225	Challenging
Kenya	Nairobi	6	\$830	Moderate
Kilauea	Hilo	1	\$55	Easy
Kilimanjaro	Nairobi	9	\$1,310	Challenging
Krakatau	Jakarta	7	\$895	Easy
Poas	San Jose	1	\$65	Easy
Reventador	Quito	4	\$575	Moderate
St. Helens	Portland	2	\$167	Easy
Vesuvius	Rome	6	\$985	Easy

Results When you run a task that produces results in multiple formats, you get an icon in the Process Flow, and a tab in the workspace, for each result format. To open a result, click its result tab, or double-click its icon in the Process Flow. This is what the results of the List Data task look like in PDF format.

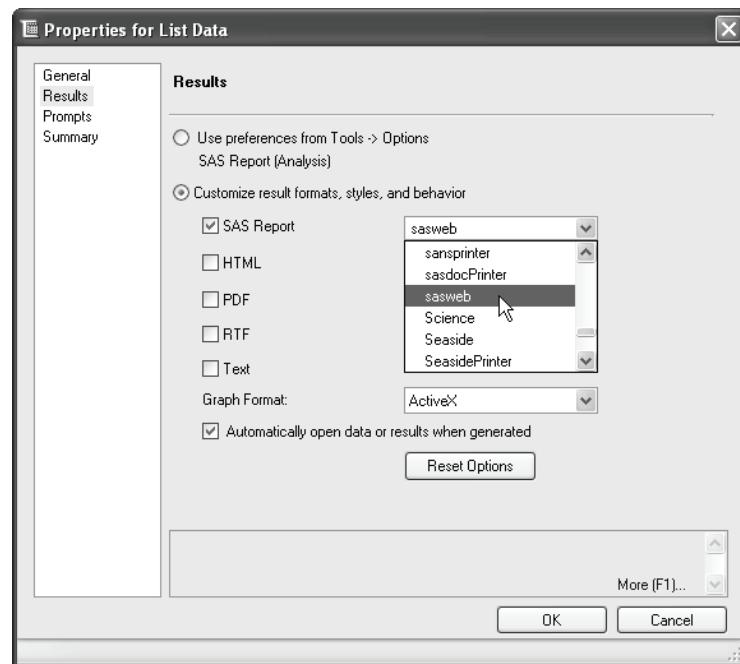
11.2 Changing the Result Style

The style determines the overall look for your results. The colors, fonts, and layout of your results are all defined in the style. Text output does not have a style associated with it, but for the other result formats—HTML, RTF, PDF, and SAS Report—you can choose from a number of styles.

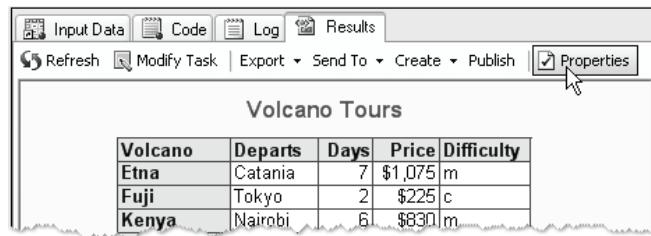
Setting the default result style The default style for results in HTML and SAS Report format is Analysis. For RTF output the default style is RTF, and for PDF it is Printer. You can change the default style for any result format using the Options window. Open the Options window by selecting **Tools ▶ Options** from the menu bar. Click the type of output (SAS Report, HTML, RTF, or PDF) in the selection pane on the left to open the page for that format. Then select the style you want for that result format from the Style drop-down list. Once you make this change, every task you run will use the style you select for that format.



Changing the style for results of a task To change the style for the results of a particular task, do this in the Properties window for the task. To open the Properties window, click **Properties** in the selection pane for the task, then click the **Edit** button. Click **Results** in the selection pane on the left to open the Results page of the Properties window. Check the box next to **Customize result formats, styles, and behavior**, and then select the result format and a style from the drop-down list for each format.



Changing the style after running a task For results in SAS Report and HTML formats, you can change the style after you run the task. Click the **Properties** button on the workspace toolbar for the result to open the Properties window. Then choose a style from the Style drop-down menu. Click **OK** and your result will be displayed with the selected style. You do not need to rerun the task.



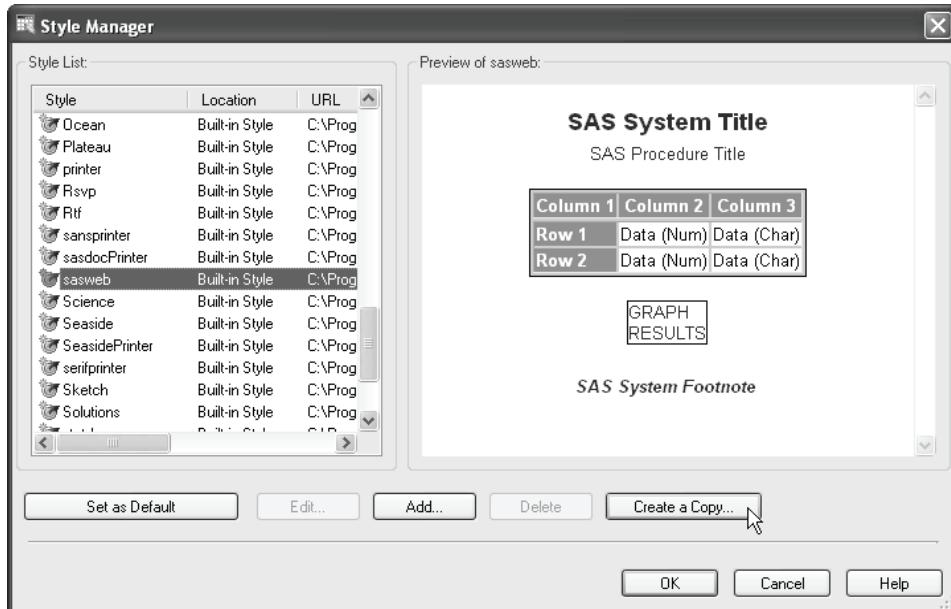
Results Here is what a SAS Report result looks like using the Sasweb style.

Volcano Tours				
Volcano	Departs	Days	Price	Difficulty
Etna	Catania	7	\$1,075	m
Fuji	Tokyo	2	\$225	c
Kenya	Nairobi	6	\$830	m
Kilauea	Hilo	1	\$55	Easy
Kilimanjaro	Nairobi	9	\$1,310	Challenging
Krakatau	Jakarta	7	\$895	Easy
Poas	San Jose	1	\$65	Easy
Reventador	Quito	4	\$575	Moderate
St. Helens	Portland	2	\$167	Easy
Vesuvius	Rome	6	\$985	Easy

11.3 Customizing Styles Using the Style Manager

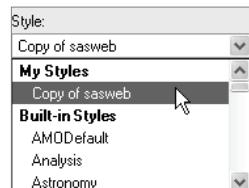
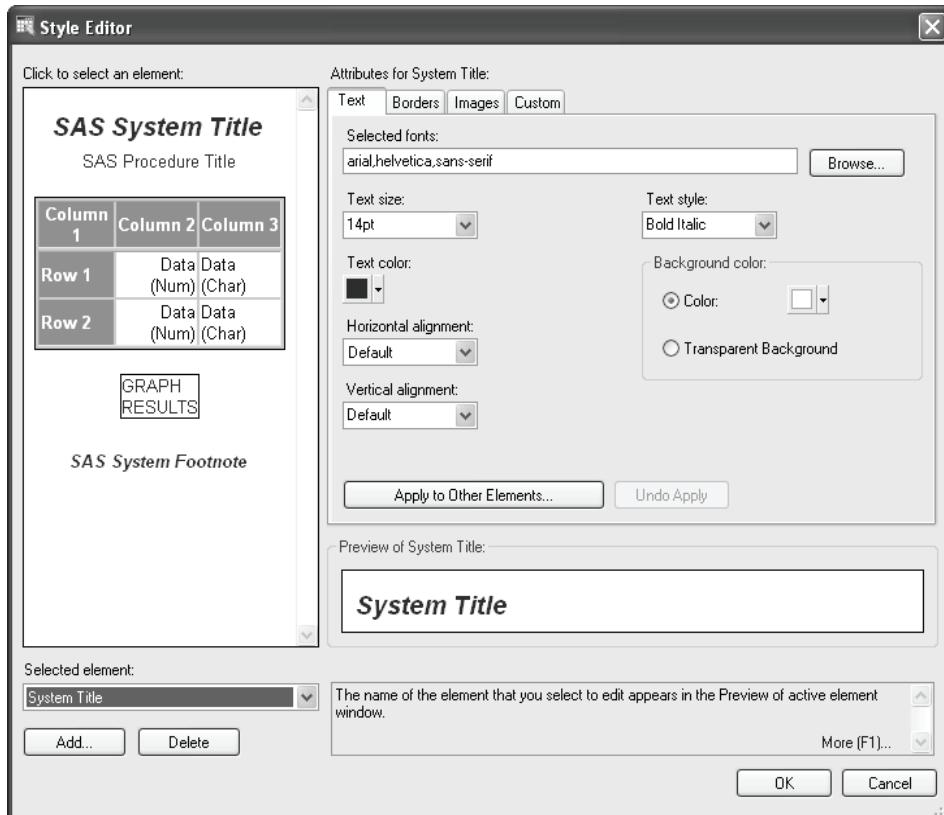
Although SAS Enterprise Guide comes with many styles for results, you still might not find a style that fits your needs. The Style Manager allows you to modify existing styles for SAS Report and HTML results. You cannot modify styles for use with RTF or PDF result formats.

Opening the Style Manager Open the Style Manager by selecting **Tools ▶ Style Manager** from the menu bar. The Style Manager provides you with a list of available styles in the box on the left. When you click a style, you will see a preview of the style in the box on the right.



Editing an existing style To edit an existing style, first create a copy of the style. Click the style in the Style Manager window, and then click **Create a Copy**. This opens the Save Style As window where you give the new style a name and choose a storage location. Now you can edit the copy of the style you just saved. Click the new style name in the Style Manager window, and click **Edit**.

This opens the Style Editor. The preview area on the left shows the current style of various elements such as titles, headers, and data cells. Choose the element you want to edit by clicking it in the preview area on the left, or selecting it from the **Selected element** drop-down list. Then select the attributes to use for that element in the area on the right. In this example, the **SAS System Title** is given a **Bold Italic** text style, and a **14pt** font text size. In addition to changing the style of the text, you can control borders using the Borders tab, and add images to your style using the Images tab. Click **OK** when you are finished making changes to the style. Then click **OK** in the Style Manager window to save your changes.



Using the new style The new style you created will appear in the **My Styles** section of the list of styles. You can select it for any SAS Report or HTML results, or set it to be the default style. See the previous section for details about changing the style for results.

11.4 Combining Results into a Single Document

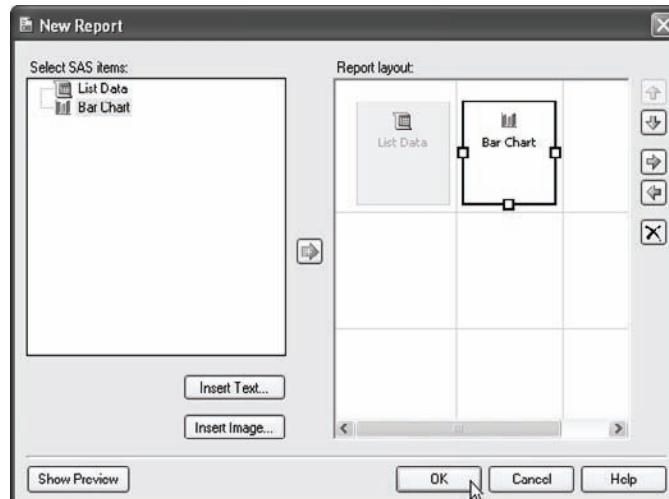
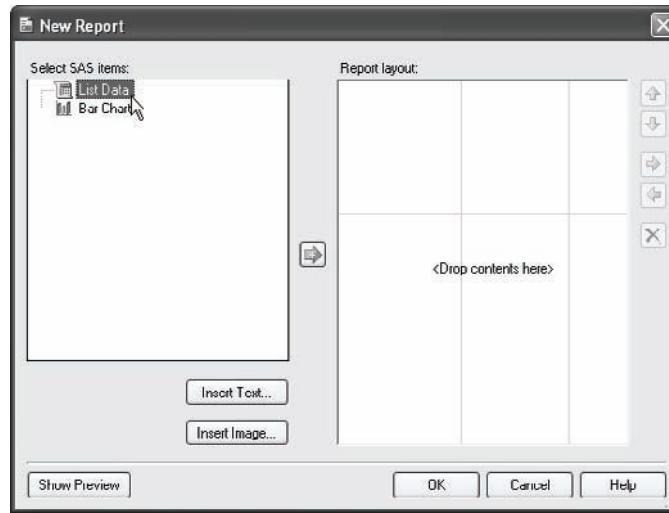
The report editor allows you to combine results from multiple tasks. The task results must be in SAS Report format. You can arrange results one above the other, or side-by-side. You can also add text and images, and add headers or footers. Whenever tasks are run, the results will be automatically updated in the report. The report can then be shared with other SAS applications such as SAS Web Report Studio, or exported to an HTML, XML, or PDF file.

Selecting Items for the report

From the menu bar, select **File ▶ New ▶ Report** to open the New Report window. All the results in your project produced in SAS Report format are listed in the box on the left labeled **Select SAS items**. A grid for the report layout is located on the right side of the window. Drag and drop desired results from the list on the left to one of the grid cells on the right. Results can appear side-by-side or one above the other.

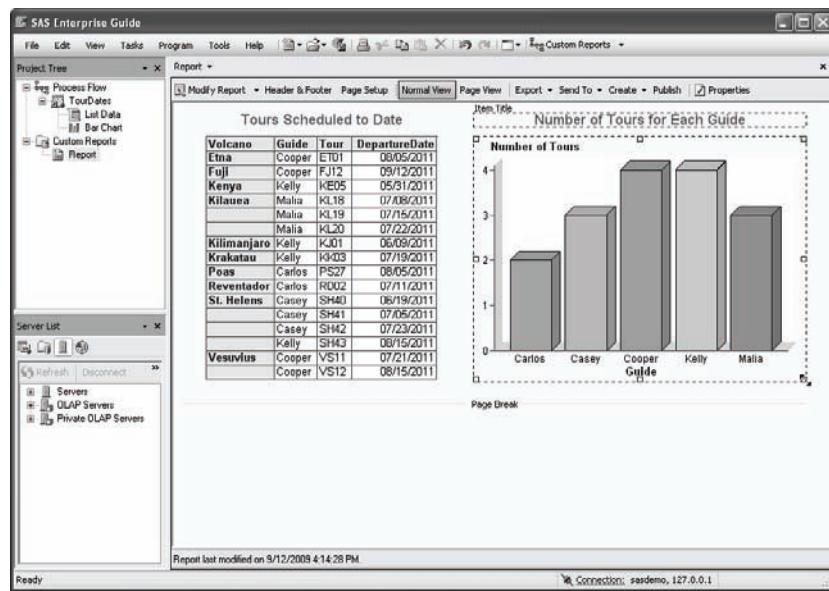
Once you have added the result to the grid, you can use the handles located on the edges of the box to expand the result to occupy multiple grid cells if desired.

In this example, a List Data result is positioned to the left of a Bar Chart result with each result occupying one grid cell. Click **OK** when you have the results positioned as desired.

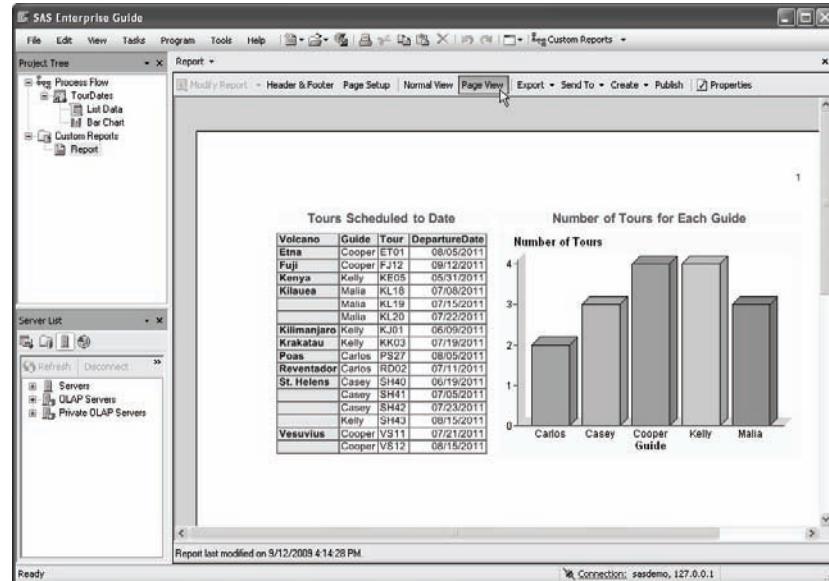


Fitting items on a page

After you close the New Report window, the results will appear in the workspace. You can resize graphs and charts if necessary at this point. You cannot resize tabular results. If you want to make changes to the content of your report, click Modify Report on the workspace toolbar for the report. If you click the down-arrow next to the Modify Report button, you can choose options to insert text or images, remove page breaks, or apply the same style to all report items.



Results To see how your result will look when printed, click the Page View button on the workspace toolbar. You cannot resize results or modify the report while you are in Page View. To make modifications, click Normal View.



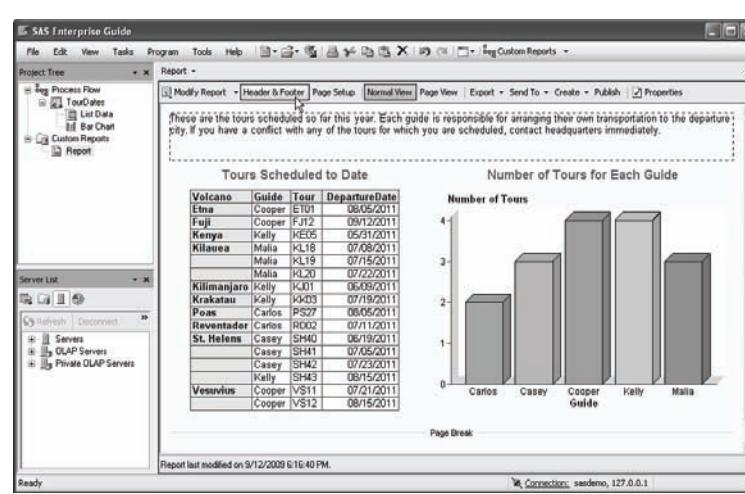
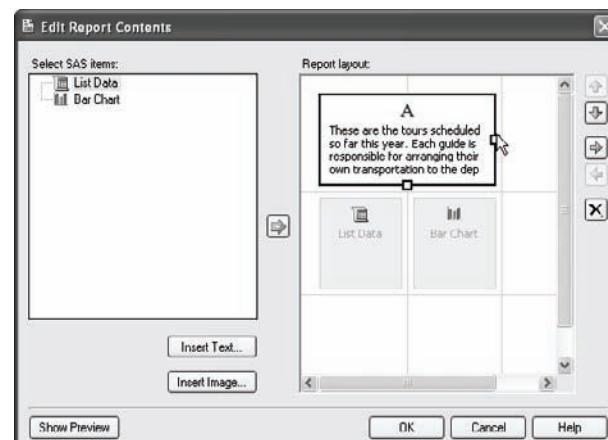
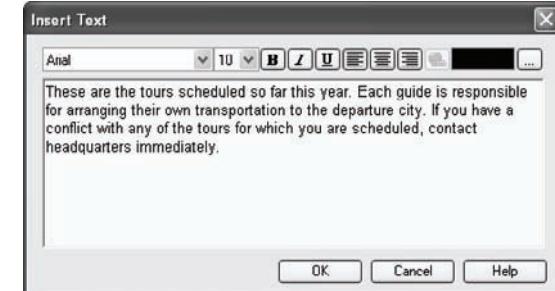
11.5 Adding Text, Images, and Headings to Reports

The report editor allows you to combine results created in the SAS Report format along with text, images, headers, and footers. To create a new report, select **File ▶ New ▶ Report** from the menu bar. To edit an existing report, click **Modify Report** on the workspace toolbar for the report. This example modifies the report created in the previous section.

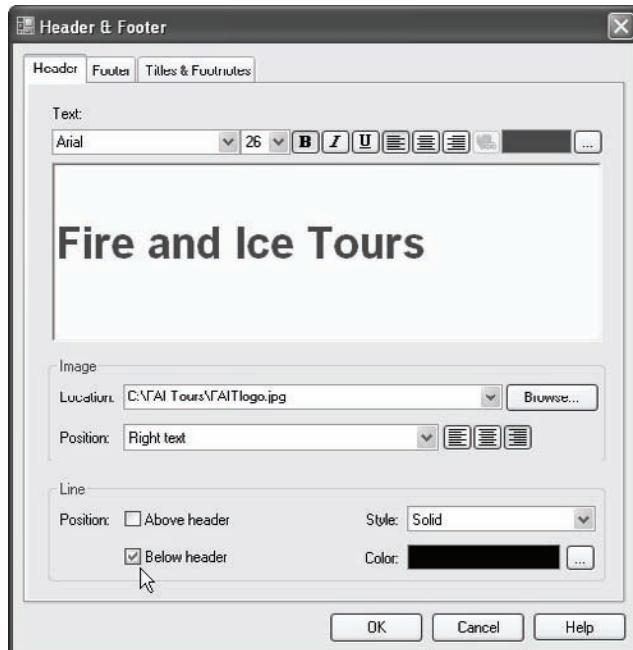
Inserting text or images To insert images into your report, click the **Insert Image** button in the report contents window. This opens a window where you can navigate to your image file. Click **Open** and the image will be inserted into one of the cells in the layout window. To insert text, click **Insert Text** to open an Insert Text window (as in this example) where you can enter the desired text, and choose formatting. Click **OK** and the text will be inserted into one of the cells in the layout window.

Arranging items in the report layout If the inserted text or image is not in the desired location, you can move it to a different cell by simply clicking and dragging. You can also change the relative size of any item by clicking the item and then clicking and dragging the handles on the edges of the box to change the shape. In this example, the text has been placed above the List Data and Bar Chart results and the box expanded to span both results. Click **OK** when you are satisfied with the layout of the report.

Adding headers and footers To add headers or footers to your report, click the **Header & Footer** button on the workspace toolbar for the report.



This opens the Header & Footer window. You can enter and format text in the text box, add and position images, and add lines above or below the header or footer. Any headers or footers you add to the report will appear on each page of the report. On the **Titles & Footnotes** tab, it is also possible to remove (but not edit) the titles and footnotes generated by SAS Enterprise Guide tasks. In this example, a header is added to the report, along with an image to the right of the header text and a line below. When you are finished making changes to the header and footer, click **OK**. The report will display in the workspace.



Results After arranging the items for your report and adding any headers or footers, you can preview your report by clicking **Page View** on the workspace toolbar for the report. Here is what the report will look like when printed. To print the report, choose **File ▶ Print Report** from the menu bar. All report contents generated by SAS Enterprise Guide tasks will be automatically refreshed when you rerun the tasks in the project. If you do not want the results refreshed, then create a report snapshot by right-clicking the report icon in the Project Explorer and selecting **Create Report Snapshot**.

1

Volcano	Guide	Tour	DepartureDate
Etna	Cooper	ET01	08/05/2011
Fuji	Cooper	FJ12	09/12/2011
Kenya	Kelly	KE05	05/31/2011
Kilauea	Malia	KL16	07/08/2011
	Malia	KL19	07/15/2011
	Malia	KL20	07/22/2011
Kilimanjaro	Kelly	KJ01	06/09/2011
Krakatau	Kelly	KK03	07/19/2011
Poas	Carlos	PS27	08/05/2011
Reventador	Carlos	RD02	07/11/2011
St. Helens	Casey	SH40	06/19/2011
	Casey	SH41	07/05/2011
	Casey	SH42	07/23/2011
	Kelly	SH43	08/15/2011
Vesuvius	Cooper	VS11	07/21/2011
	Cooper	VS12	08/15/2011

Number of Tours for Each Guide

Number of Tours

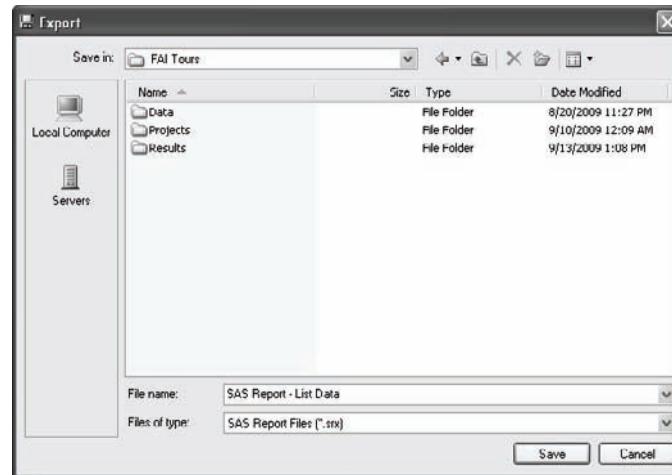
Guide	Number of Tours
Carlos	2
Casey	3
Cooper	4
Kelly	4
Malia	3

11.6 Exporting Results to a File

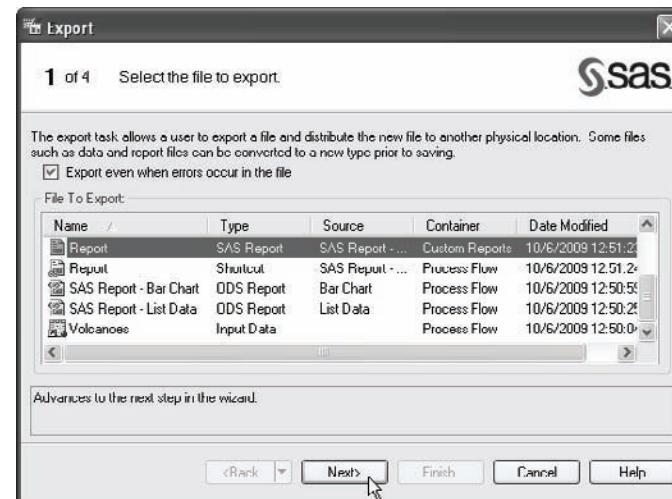
By default, results generated by SAS Enterprise Guide are saved inside the project. If you want, you can export your results to a file. Exporting results can be done manually or automatically each time you run the project.

Exporting results to a file

You can export any result to a file by right-clicking the result icon in the Process Flow and selecting **Export ▶ Export result-format – result-name** or choosing this option from the **Export** menu on the workspace toolbar for the result. This opens an Export window where you can save the file on your local computer or on a SAS server. If your result is in SAS Report format, you can export the file as a SAS Report, HTML, XML, or PDF file. All other result formats can only be exported in the original format.



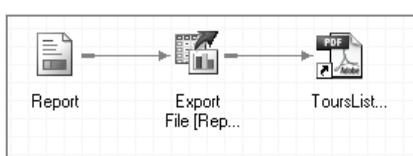
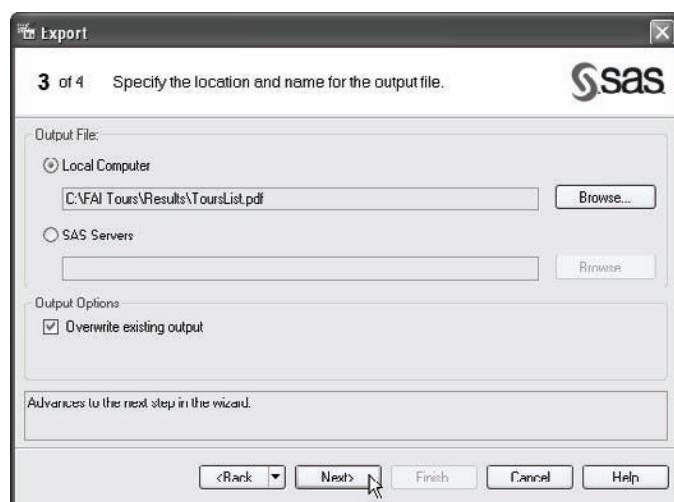
Exporting results as a step in the project If you export results as a step in the project, then each time the project is run, the exported file will be updated. Right-click the result icon in the Process Flow and select **Export ▶ Export result-format – result-name as a step in the project** or choose this option from the **Export** menu of the workspace toolbar for the result. This opens the Export wizard. In the first window, all items in the project are listed with the selected item highlighted. Click **Next**.



If the result you wish to export is in SAS Report format, then the second window will give you a choice for the output file type: SAS Report, HTML, or PDF. If the result format is not SAS Report, then you will not have a choice because you can only export the result in its original format, and you will not see this window. Choose the desired output file type and click **Next**.



In the next window, select the location to save the file. This can be on your local computer, or on a SAS server. By default, if a file with the same name already exists, it will be overwritten. If you uncheck **Overwrite existing output**, then instead of overwriting, a new file will be created with the date and time appended to the file name. Click **Next** to view the final window of the wizard (not shown) that provides a summary of your choices. Click **Finish**.



Results When you export as a step in the project, an icon will appear in the Process Flow showing the export and the result. If your export is not done as a step in the project, then no icons will appear. This example shows the Process Flow for a custom report that was exported as a step in the project to a PDF file.

12

“ The difficulty lies, not in new ideas, but in escaping from old ones. ”

JOHN KEYNES

From *The General Theory of Employment, Interest, and Money*, 1936.



CHAPTER 12

Adding Flexibility with Prompts and Conditions

- 12.1 Creating Prompts for Data Values 326
- 12.2 Using Prompts in Query Filter Conditions 328
- 12.3 Creating Prompts for Variable Names 330
- 12.4 Using Prompts in Tasks 332
- 12.5 Using Prompts in SAS Programs 334
- 12.6 Creating Prompts for Text Values 336
- 12.7 Using Prompts in Project Conditions 338
- 12.8 Running Projects with Conditions 340

12.1 Creating Prompts for Data Values

Prompts allow you to develop projects that are flexible. When you run a project that uses a prompt, a box will appear asking you to specify a value for that prompt. The project will be run using the value you enter. This allows you to create one project that can generate many different results.

There are four ways you can use prompts: in the filter condition of a query, in a task, in a SAS program, or in the condition of a process flow. The example in this section creates a data value prompt that will be used in a query in the next section. The rest of this chapter shows how to create and use other kinds of prompts.

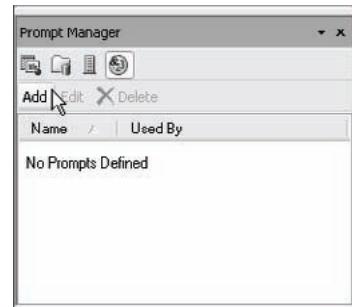
It may be helpful to know that prompts have other names. In SAS Enterprise Guide 4.1, prompts were called parameters. SAS programmers generally refer to prompts as macro variables.

To create prompts, you use the Prompt Manager. The Prompt Manager shares the Resources pane with other windows. To open the Prompt Manager, click the Prompt



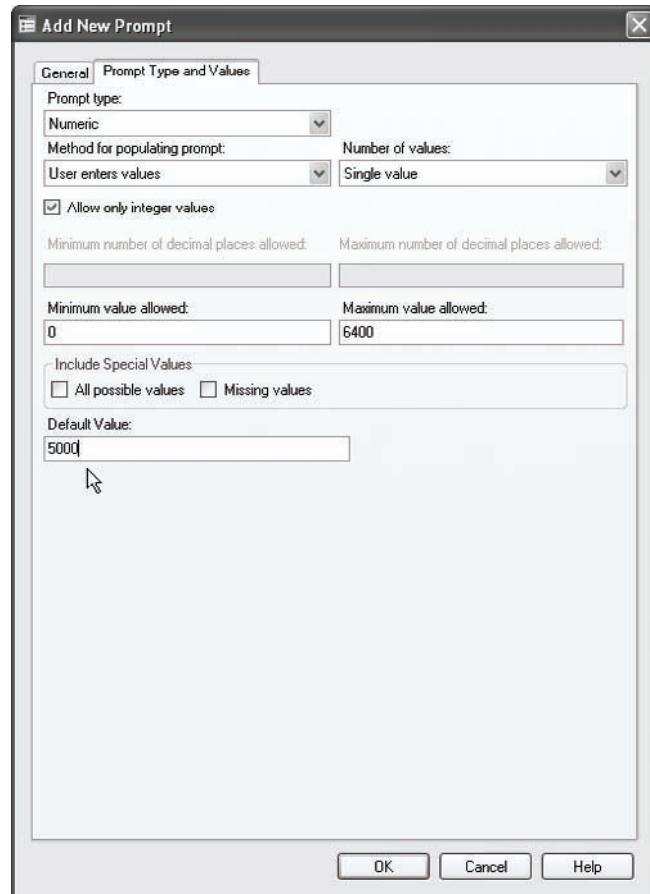
Manager icon in the Resources pane. You can also open the Prompt Manager from the View menu, or by clicking **Prompt Manager** in the Query Builder. In the Prompt Manager, click **Add** to open the Add New Prompt window.

Naming the prompt On the General tab of the Add New Prompt window, you specify a Name, Displayed text, and an optional Description for the new prompt. The name must be 32 characters or fewer in length; start with a letter or underscore; and contain only letters, numerals, and underscores. When you use the prompt, you will see the displayed text and the description, but not the name.

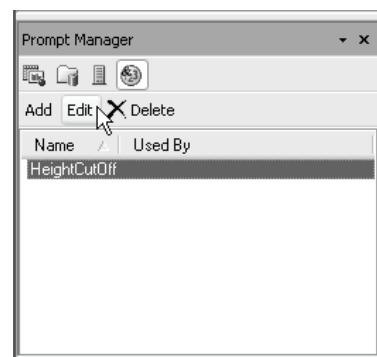


In this example, the prompt has a name of **HeightCutOff**, the displayed text is **Minimum Height**, and there is a description. Next click the **Prompt Type and Values** tab.

Setting the prompt type and values On the **Prompt Type and Values** tab, click the down-arrow on the box labeled **Prompt type**, and select an option from the pull-down list. For this example, select **Numeric**. When you select the prompt type, options for that particular type of prompt will appear in the lower part of the window. For a numeric prompt, you can enter minimum, maximum, and default values. For this example, the **Minimum value allowed** is zero, the **Maximum value allowed** is 6400, the **Default Value** is 5000, and the option **Allow only integer values** is checked. Since the tallest volcano is 6458 meters, setting the maximum value at 6400 ensures that at least one row of the Volcanoes table will always be selected after filtering. When you have set all the properties for the prompt, click **OK**.



The prompt you created will appear in the Prompt Manager. If you want to make changes to the prompt, click its name to highlight it, and click **Edit**. To delete a prompt, click **Delete**.



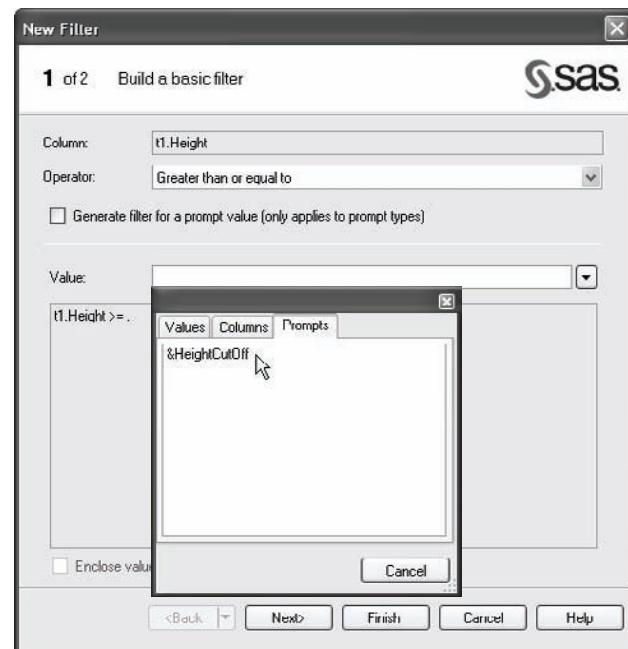
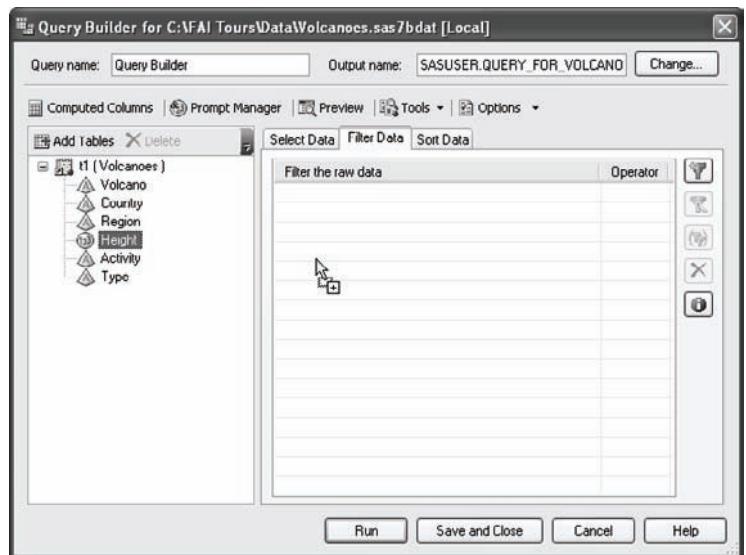
12.2 Using Prompts in Query Filter Conditions

The previous section showed how to create a prompt for a data value. This section shows how to use that prompt in the filter condition of a query.

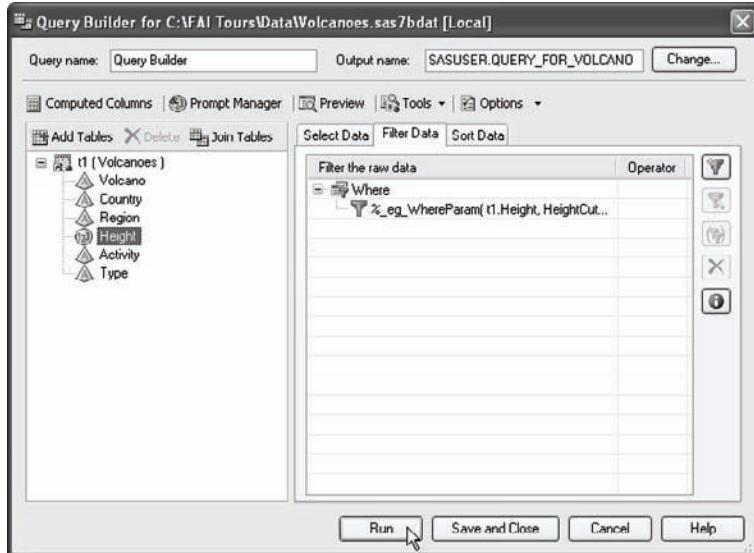
Setting the Filter Condition To open the Query Builder, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Data ▶ Query Builder** from the menu bar. In the Query Builder, drag the columns you want to keep to the **Select Data** tab. For this example, use the Volcanoes data table, and select all the columns.

Then click the **Filter Data** tab and create the filter just like you would any other filter. Drag the desired column to the **Filter Data** tab of the Query Builder. For this example, drag the column Height. The New Filter wizard will open.

In the first window of the New Filter wizard, choose the appropriate operator for the condition, in this case **Greater than or equal to**. Click the down-arrow next to the box labeled Value. Then click the **Prompts** tab, and select the desired prompt. Notice that the prompt name is preceded by an ampersand (&). The ampersand indicates that this is the name of a prompt, not a column. In this example, the filter will select all rows where the value of the Height column is greater than or equal to the value of the **&HeightCutOff** prompt.

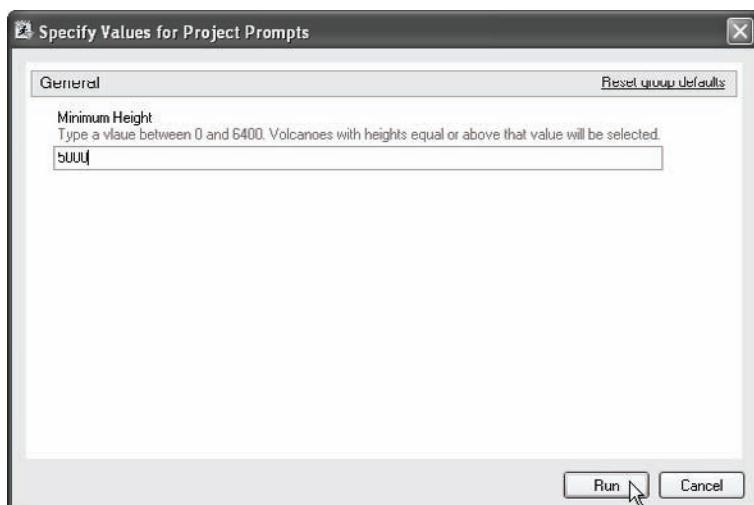


Click **Next** if you want to see a summary of your filter settings. To complete the filter, click **Finish** in the New Filter window. To run the new query, click **Run**.



Running the query

When you run the query, a window will open, asking you to specify a value for the prompt. In this example, you must enter a value between the minimum allowed value (zero) and the maximum (6400). When you are satisfied, click **Run**. In the Process Flow, you will see that the query icon now includes a question mark, , to show that the query uses a prompt.



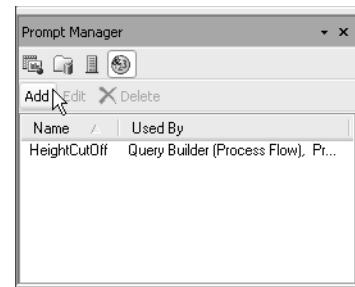
Results Here is the data table with results from selecting 5000 as the cut-off value for height. This table includes all the rows from the Volcanoes table with a height greater than or equal to 5000. Any tasks that use this data table will reflect the selection made in the query. Every time you run the query, you will be asked again to specify a value for the prompt.

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
3	Illimani	Bolivia	SA	6458	Extinct	Stratovolcano
4	Kenya	Kenya	Af	5199	Extinct	
5	Kilimanjaro	Tanzania	Af	5895		Stratovolcano
6	Popocatepetl	Mexico	NA	5426	Active	Stratovolcano
7	Sabancaya	Peru	SA	5976	Active	Stratovolcano

12.3 Creating Prompts for Variable Names

The previous two sections showed how to create a prompt for a data value and then use it in the filter condition of a query. This section shows how you can create a prompt for a variable name, and the next section shows how you can use this prompt in a task. This example creates a prompt that will allow a user to choose a grouping variable when a task is run.

To create a prompt, click the Prompt Manager icon  in the Resources pane, or select **View ▶ Prompts Manager** from the menu bar. The Prompt Manager will open. Click **Add** to open the Add New Prompt window.



Naming the prompt On the General tab of the Add New Prompt window, you specify a **Name**, **Displayed text**, and an optional **Description** for the new prompt. The name must be 32 characters or fewer in length; start with a letter or underscore; and contain only letters, numerals, and underscores. When you use the prompt, you will see the displayed text and the description, but not the name. In this example, the prompt has the name **GroupVar**, the displayed text is **Grouping variable**, and there is a description. Next click the **Prompt Type and Values** tab.

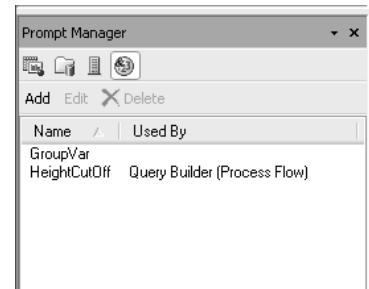
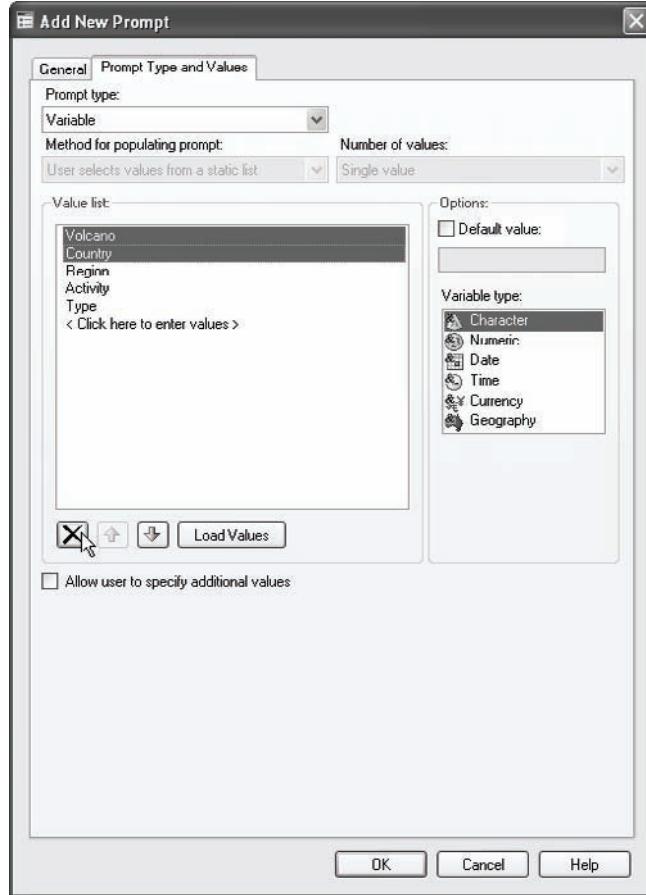


Setting the prompt type and values On the **Prompt Type and Values** tab, click the down-arrow on the box labeled **Prompt type**, and select an option from the pull-down list. For this example, select **Variable**. When you select the prompt type, options for that particular type of prompt will appear in the lower part of the window.

In the section labeled **Options** on the right side of the window, you can choose the variable type. The default variable type is character.

If you click the **Load Values** button near the bottom of the window and navigate to a data set, all the variables of that type for that data set will be listed in the section labeled **Value list**. To delete unwanted variables, highlight their names, and click the delete  button. In this example, all the character variables in the Volcanoes data set were listed, but Volcano and Country are being deleted. Only the variables Region, Activity, and Type make sense as grouping variables. It is possible to choose a different variable type (such as numeric or date), and click Load Values again so that the list will include variables of more than one type. For this example, only three character variables (Region, Activity, and Type) are needed. When you have set all the properties for the prompt, click **OK**.

The prompt you created will appear in the Prompt Manager window.



12.4 Using Prompts in Tasks

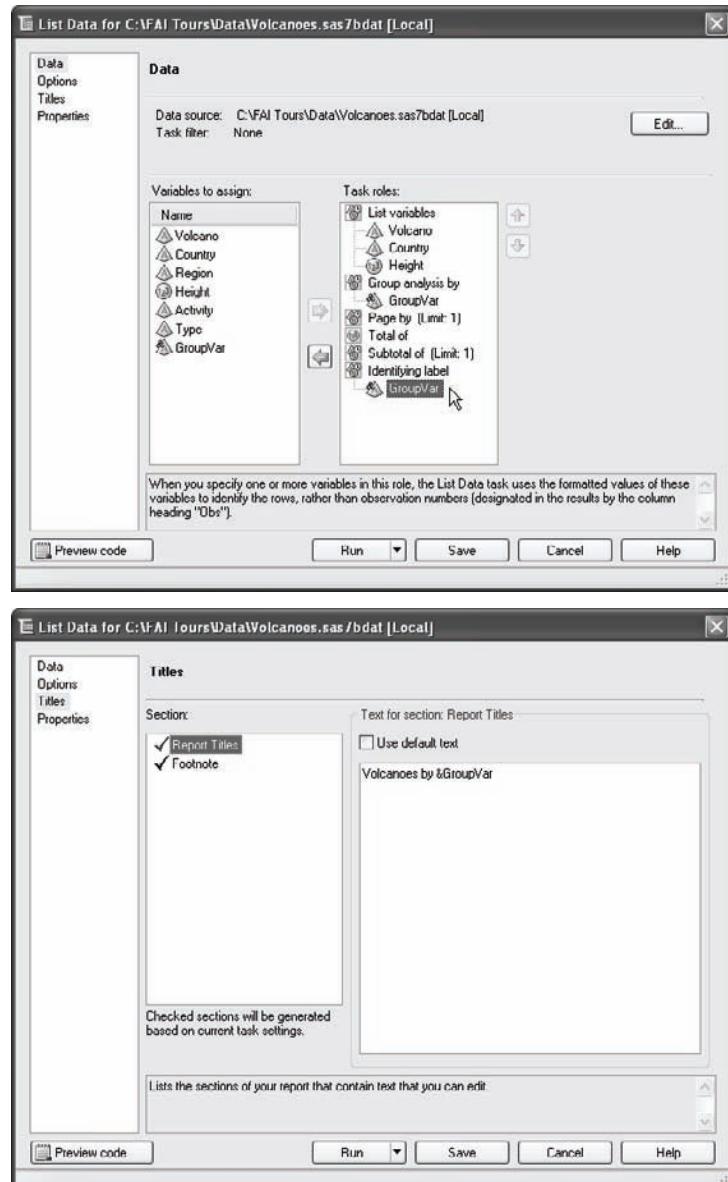
The previous section showed how to create a prompt for the name of a variable. This section shows how to assign that prompt to a task role, and how to run the new task.

This example uses the List Data task and the Volcanoes data table. To open this task, click the data icon in the Project Tree or Process Flow to make it active, and select **Tasks ▶ Describe ▶ List Data** from the menu bar.

Assigning a prompt to a task role Notice that the list of variables in the task roles page includes not only the regular variables, but also any variable prompts you have created for this project. The ampersand (&) in the data icon  indicates that GroupVar is a prompt, not a regular variable.

You assign variables and prompts to roles by dragging them to task roles. In this example, the variables Volcano, Country, and Height have been assigned to the **List variables** role. The prompt GroupVar has been assigned to the **Group analysis by** and **Identifying label** roles.

Using prompts in titles You can use prompts to create custom titles or footnotes. Click **Titles** in the selection pane on the left to open the Titles page.



The Titles page opens, displaying the default title for that particular task. To change the title, uncheck the **Use default text** option and type the new title in the text box. In this example, the title has been changed to “Volcanoes by &GroupVar.” When you are satisfied with the settings, click **Run**.

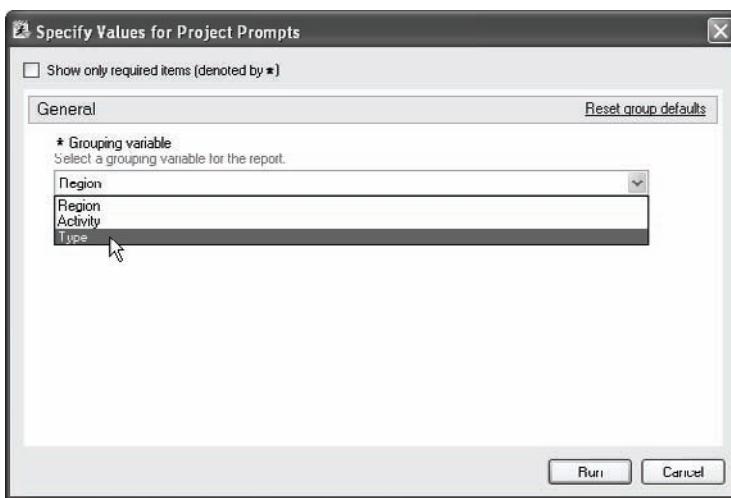
Running the task

When you run the task, a window will open, asking you to specify a value for the prompt. In this window, the variable Type is being selected from a pull-down list of three possible grouping variables. Click **Run** to run the task.

In the Process Flow, you will see that the task icon now includes a question



mark, , to show that the task uses a prompt.



Results Here are the results of the task. This is a List Data report organized by type of volcano. Notice that the value of the prompt (in this case Type) also appears in the title. Every time you run the task, you will be asked again to choose a grouping variable.

Volcanoes by Type			
Type	Volcano	Country	Height
	Arthur's Seat	UK	251
	Kenya	Kenya	5199
Caldera	Grimsvotn	Iceland	1725
	Krakatau	Indonesia	813
Cinder Cone	Puy de Dome	France	1464
Complex	Vesuvius	Italy	1281
Shield	Kilauea	USA	1222
	Mauna Loa	USA	4170
	Nyamuragira	DRCongo	3058
	Santorini	Greece	367
	Warning	Australia	1125
Stratovolcano	Altar	Ecuador	5321
	Barren Island	India	354
	Elbrus	Russia	5633
	Erebus		3794
	Etna	Italy	3350

12.5 Using Prompts in SAS Programs

The previous sections showed how to create prompts and use them in queries or tasks. This section shows how you can create flexible SAS programs by using the same prompts. Then when you run the SAS programs, SAS Enterprise Guide will ask you to specify values for the prompts.

Adding prompts to SAS programs Here is a simple SAS program typed in a Program window in SAS Enterprise Guide.



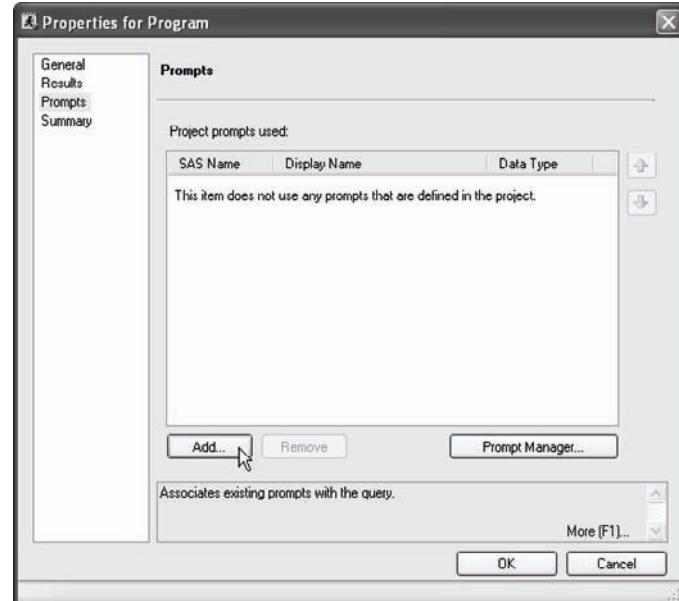
```

Program
Save ▾ Run ▾ Stop Select Server | Export ▾ Send To ▾ Create ▾ Properties
PROC FREQ DATA = 'C:\FAI Tours\Data\Volcanoes';
  WHERE Height >= &HeightCutOff;
  TABLES &GroupVar;
  TITLE1 "Number of Volcanoes &HeightCutOff Meters or Higher";
  TITLE2 "For Each Value of &GroupVar";
RUN;

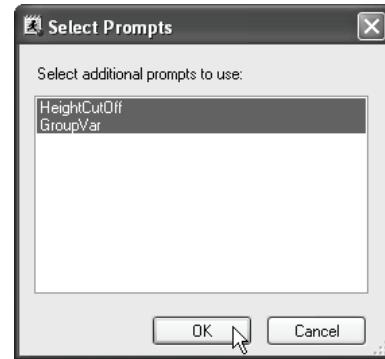
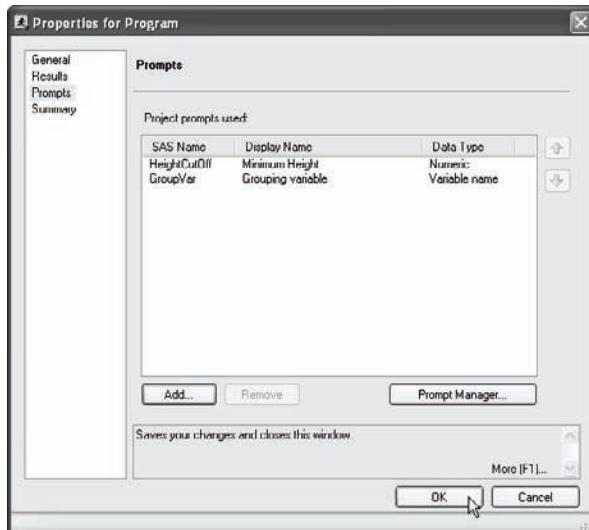
```

This program uses both the data value prompt created in section 12.1 (&HeightCutOff) and the variable name prompt created in section 12.3 (&GroupVar). &HeightCutOff is used in a WHERE statement to select only rows where the values of the Height column are greater than or equal to the prompt value. &GroupVar is used in a TABLES statement so the data will be summarized based on the values of the grouping variable. Then both prompts are used again in the TITLE statements so that the values of the prompts will appear in the titles for the report.

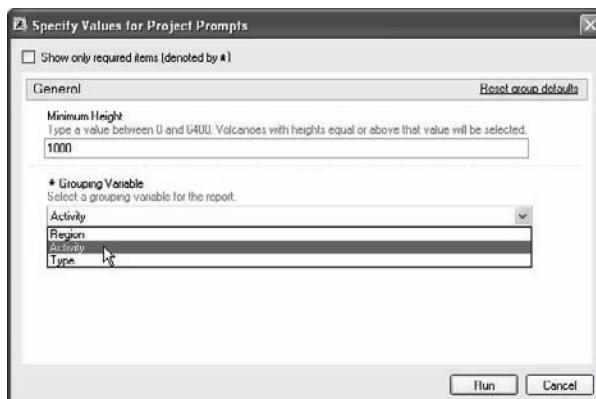
Setting the properties of the program Because the prompts have been defined outside the program, you need to associate the prompts with the program. To do this, right-click the program icon in the Project Tree or Process Flow and select **Properties** from the pop-up menu. Then, in the Properties for Program window, click **Prompts** in the selection pane on the left to display the Prompts page. Click the **Add** button. The Select Prompts window will open.



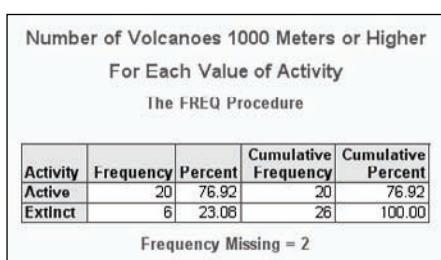
The Select Prompts window lists all the prompts currently defined in the project. Hold down the control key (CTRL), and click the names of all the prompts used in the program. Then click OK to return to the Properties for Program window.



The prompts will appear in the Properties for Program window. This window shows that both the HeightCutOff and GroupVar prompts have been associated with this program. Click OK to save the changes and close the window.



Running the SAS Program To run the program, click Run in the Program window, or click the program icon in the Project Tree or Process Flow, and select **Program ► Run program-name On server-name** from the menu bar. When you do this, a window will open, asking you to specify values for each prompt. In this window, the cut-off height has been set to 1000, and the grouping variable has been set to Activity. When you are satisfied, click Run. In the Process Flow, you will see that the program icon now includes a question mark, , to show that the program uses a prompt.

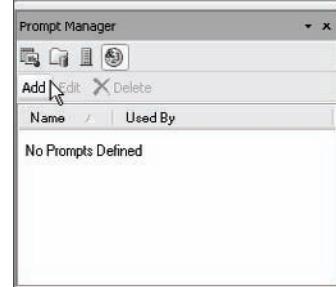


Results This report is summarized by Activity, includes all the volcanoes over 1000 meters, and has custom titles.

12.6 Creating Prompts for Text Values

Earlier sections in this chapter showed how to create a prompt for a data value or a variable name, and how to use those prompts in a query, task, or program. This section shows how to create a prompt for a text value, and the next two sections show how to use this prompt in a project condition. The text value prompt and project condition in this example allow a user to choose which of two tasks to run.

To create a prompt, click the Prompt Manager icon  in the Resources pane, or select **View ▶ Prompts Manager** from the menu bar, or click **Prompt Manager** in the Query Builder. The Prompt Manager will open. Click **Add** to open the Add New Prompt window.



Naming the prompt On the **General** tab of the Add New Prompt window, you specify a **Name**, **Displayed text**, and an optional **Description** for the new prompt. The name must be 32 characters or fewer in length; start with a letter or underscore; and contain only letters, numerals, and underscores. In this example, the prompt has the name **ChooseReport**, the displayed text is **Choose a report to run**, the description is **Select one option**, and the option **Requires a non-blank value** has been checked. Next click the **Prompt Type and Values** tab.



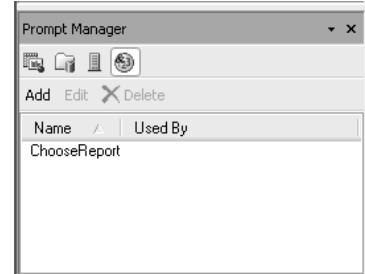
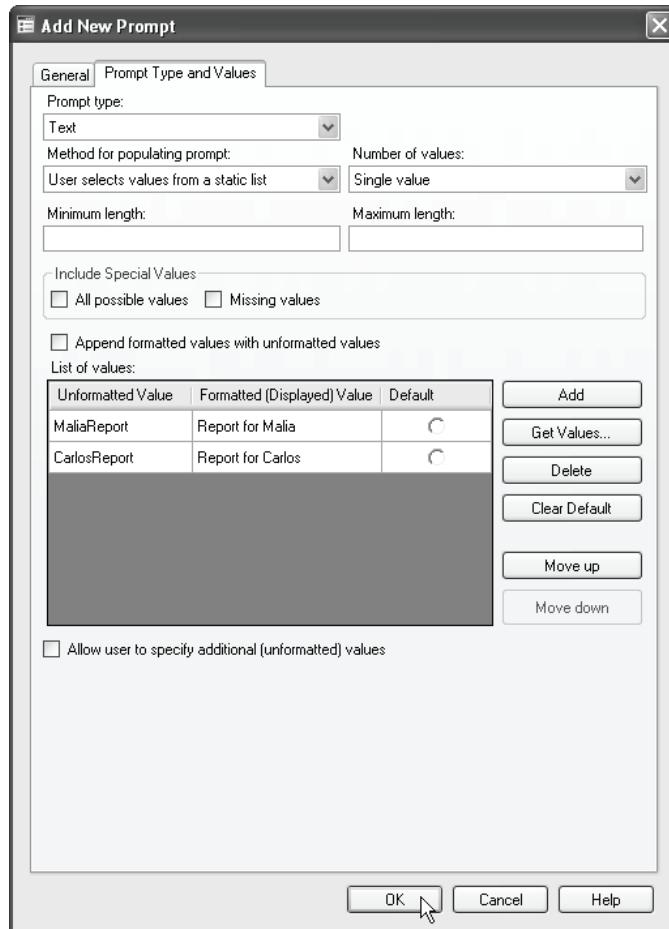
Setting the prompt type and values

On the Prompt Type and Values tab, click the down-arrow on the box labeled **Prompt type**, and select an option from the pull-down list. For this example, select **Text**. When you select the prompt type, appropriate options will appear in the lower part of the window.

Click the down-arrow on the box labeled **Method for populating prompt**, and select the method. Users can type in a value, or choose from a static or dynamic list. For this example, choose **User selects values from a static list**. In the box labeled **Number of values**, select **Single value**.

For a static list, you must specify possible values for the prompt. To do this, click the **Add** button. Empty cells will appear in the area labeled **List of values**. Click a cell and type in the unformatted value and the formatted value for the first choice. Then click **Add** again, and type in the values for the second choice, and so on until you have typed in all the values you want to appear in the static list. For this example, the text prompt has two possible values: **MaliaReport**, which has a formatted value of **Report for Malia**, or **CarlosReport**, which has a formatted value of **Report for Carlos**.

When you have set all the properties for the prompt, click **OK**. The prompt you created will appear in the Prompt Manager window.



12.7 Using Prompts in Project Conditions

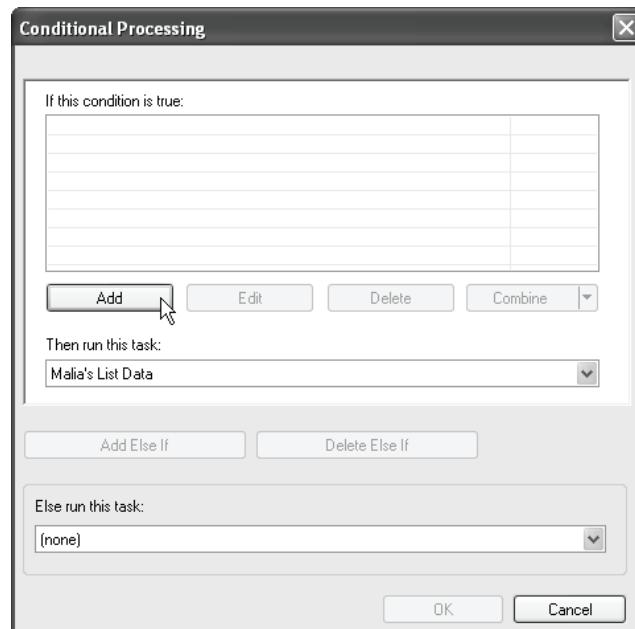
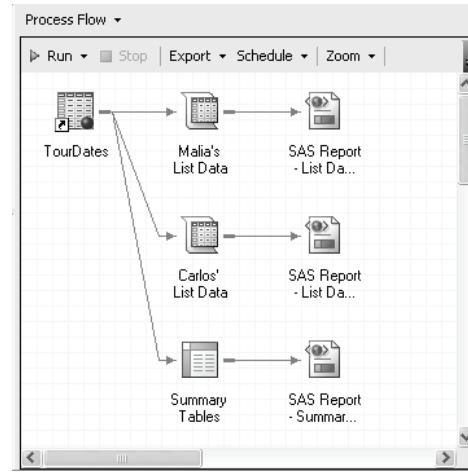
A project condition is an expression that is evaluated by SAS Enterprise Guide at run time. Any item that you can run in SAS Enterprise Guide can have a project condition associated with it. If the condition is true, then that item will run. If it is not true, then that item will not run.

Project conditions can be based on the value of a prompt, on a data value, or on the date or time when the project runs. So, for example, you could have a query that runs only if you choose it from a list of options, or a data set that is exported whenever a particular data value appears, or a report that runs only on Wednesdays. The example in this section creates a project condition based on the text value prompt created in the preceding section.

The project in this example contains three items that can be run: two List Data tasks and a Summary Tables task. The two List Data tasks have been renamed as Malia's List Data and Carlos' List Data. The project condition will prompt the user to choose which of the two List Data tasks to run.

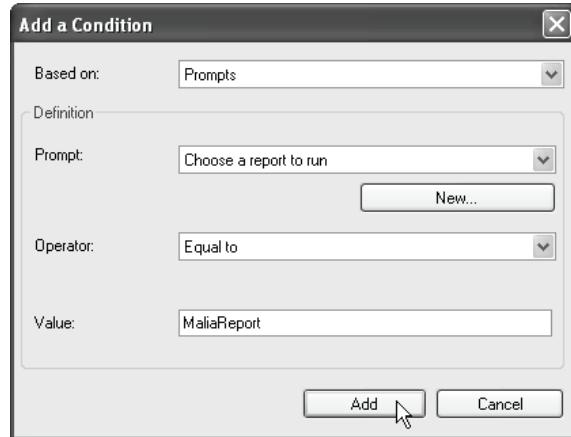
To apply a project condition to an item in a Process Flow, right-click the icon for that item, and select **Condition ▶ Add** from the pop-up menu. A Conditional Processing window will open. For this example, the Conditional Processing window has been opened by right-clicking the icon for Malia's List Data.

In the Conditional Processing window, click **Add** to open the Add a Condition window.



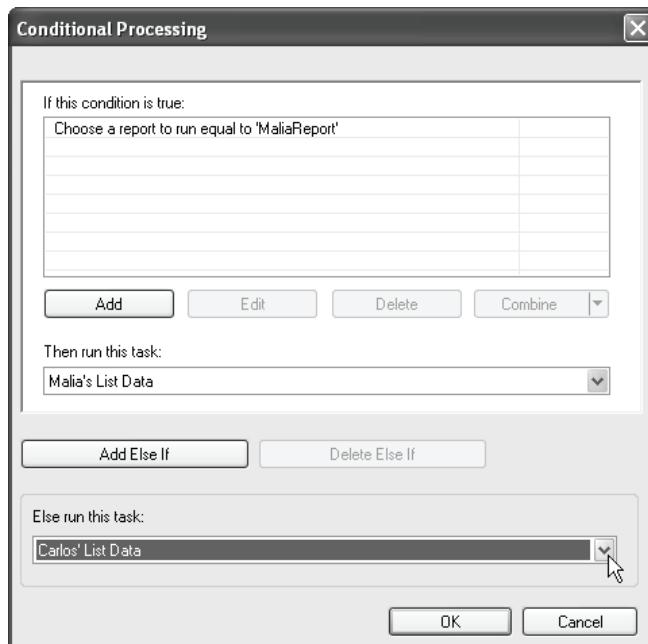
Adding a condition In the Add a Condition window, click the down-arrow in the **Based on** box, and select the type of condition. Your condition can be based on prompts, the current time or date, or a data value in your project. When you select the Based on value, options for that type of condition will appear in the lower part of the window. In this example, the condition will be based on **Prompts**. Click the down-arrow for the box labeled **Prompt** to display a list of text prompts already defined in the project, or click **New...** to create a new one. Choose the operator from the box labeled **Operator**, and type the value of the prompt for this condition in the box labeled **Value**.

For this example, the prompt is **Choose a report to run** (the prompt created in the previous section), the operator is **Equal to**, and the value is **MaliaReport**. When you are satisfied, click **Add** to close the Add a Condition window.



Specifying an action The condition you just created will appear under the words **If this condition is true**. Click the down-arrow on the box labeled **Then run this task**, and select the item in the project that should run when that particular condition is true.

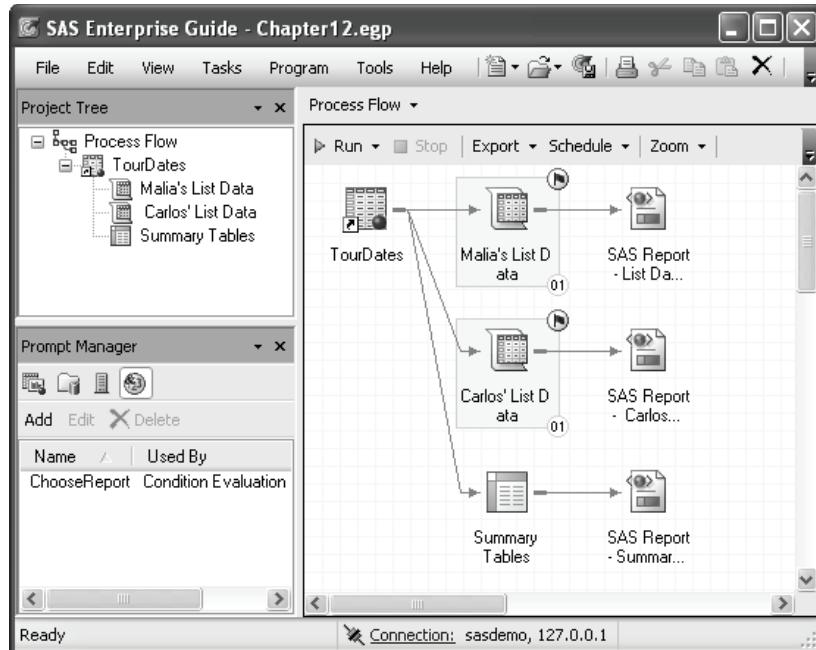
Specifying an Else action At this point, you must specify what to do if the first condition is not true. You have two choices. You can click **Add Else If** to create another condition, or you can click the down-arrow on the box labeled **Else run this task** and choose an alternate item to run whenever the first condition is not true. Since this project condition will include only the two List Data tasks, you do not need to specify another condition. Click the down-arrow on the box labeled **Else run this task**, and select the item that should run when the condition is not true.



In this Conditional Processing window, you can see that when the value of the prompt (named **Choose a report to run**) is **MaliaReport**, then the task named **Malia's List Data** will run. Otherwise, the task named **Carlos' List Data** will run. When you are satisfied with the settings, click **OK** to close the window and apply the condition to the project. The next section shows the Process Flow with the condition applied.

12.8 Running Projects with Conditions

When you add a condition to a project, it is illustrated in the process flow diagram. Here is what the process flow looks like for the project condition that was created in the preceding section.



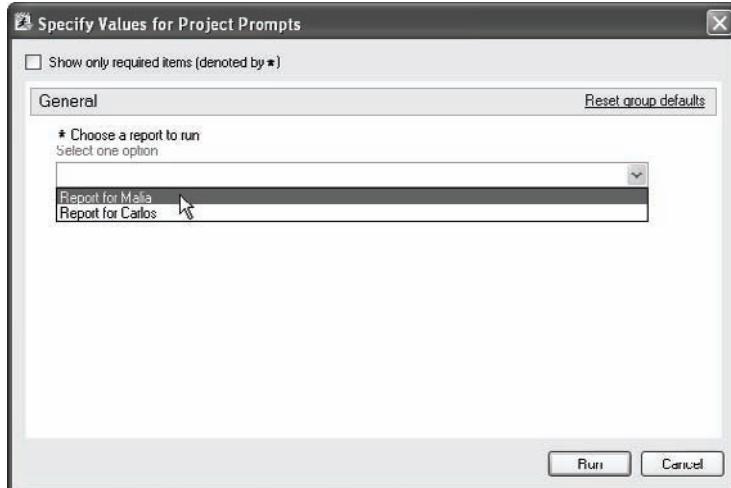
Condition indicators Boxes appear around any items associated with a condition. Numbers in the lower-right corner of each box indicate which items are part of the same condition. In this process flow, the number 01 shows that the two List Data tasks are alternate paths in the same condition, but the Summary Tables task is not. You can specify more than one condition for a process flow. If you have a second condition, then items in that condition will be marked with 02. A third condition would be marked with 03, and so on.

In the upper-right corner of a condition box appears a symbol. At first, this symbol looks like a waving flag. After the process flow has run, the symbol changes to either a check mark indicating that the branch ran, or an X indicating that the branch did not run.

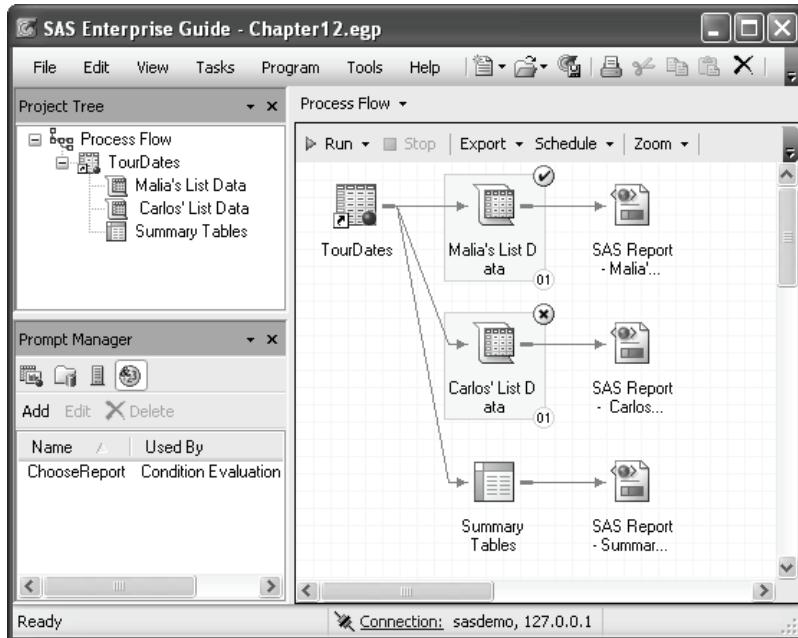
Running the project You run a project with conditions in the same way that you run an ordinary project. Select **Run** from the workspace toolbar for the Process Flow, or select **File ▶ Run project-name** from the menu bar. You can also run one branch of a process flow by right-clicking an icon in the Process Flow window and selecting **Run Branch from item-name**.

Project conditions do not have to use prompts. Instead of depending on prompts, the condition could depend on the current date or time, or on a value in your data. In those cases, the Process Flow would simply run without prompting for any user input.

If the condition uses a prompt, then, when you run it, a window will open allowing you to specify a value for the prompt. In this example, the prompt is **Choose a report to run**, and it has two possible values. This is the text prompt that was created in section 12.6. The value **Report for Malia** is being selected. When you are satisfied with your choice, click **Run**.



Here is the Process Flow after the project has run.

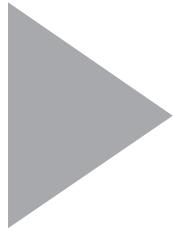


You can see that Malia's List Data task has a check mark in the corner, and that Carlos' List Data task has an X. These condition indicators show that the first task ran, but the second one did not. Since the Summary Tables task is not part of the condition, it will always run. Every time you open a project with conditions, all the condition indicators will be set back to waving flags.



APPENDIX A

A Data Used in This Book	345
Tours Data	345
Tour Dates Data	346
Tour Bookings Data	347
Volcanoes Data	349
Eruptions Data	351
Latitude and Longitude Data	354
Portland Flights Data	356
Seattle Flights Data	356
Northwest Weather Data	357
On-Time Status of Flights Data	358
Advertising Results Data	359
Ages Data	360



“Begin at the beginning,”
the King said, very gravely,
“and go on till you come to the
end: then stop.”

LEWIS CARROLL

The King speaking with the White Rabbit in *Alice's Adventures in Wonderland*, 1865.

A Data Used in This Book

Reading about a topic is good, but many people learn better by doing. However, before you can do the examples in the tutorials or reference sections, you need the data. One way to get the data is to type the data shown in this appendix into a SAS Enterprise Guide Data Grid, a Microsoft Excel spreadsheet, or a text file. Another way is to download the data sets by going to

support.sas.com/authors.

Select the name of one of the authors (Susan Slaughter or Lora D. Delwiche). Then find the picture of this book, click the words **Example Code and Data** underneath the picture, and follow the instructions.

Tours Data

Filename: Tours.sas7bdat

File Type: SAS data set

Columns:

Name	Description	Possible Values
Volcano	Name of the volcano	
Departs	City from which tour departs	
Days	Length of the tour in days	
Price	Price of the tour in U.S. dollars	
Difficulty	Strenuousness of the tour	c (Challenging) m (Moderate) e (Easy)

Tours.sas7bdat

	Volcano	Departs	Days	Price	Difficulty
1	Etna	Catania	7	\$1,075	m
2	Fuji	Tokyo	2	\$225	c
3	Kenya	Nairobi	6	\$830	m
4	Kilauea	Hilo	1	\$55	e
5	Kilimanjaro	Nairobi	9	\$1,310	c
6	Krakatau	Jakarta	7	\$895	e
7	Poas	San Jose	1	\$65	e
8	Reventador	Quito	4	\$575	m
9	St. Helens	Portland	2	\$167	e
10	Vesuvius	Rome	6	\$985	e

Tour Dates Data

Filename: TourDates.sas7bdat
File Type: SAS data set

Columns:

Name	Description
Tour	Code for tour
Volcano	Name of the volcano
DepartureDate	Date of tour departure
Guide	Name of guide for tour

TourDates.sas7bdat

	Tour	Volcano	DepartureDate	Guide
1	PS27	Poas	08/05/2011	Carlos
2	SH40	St. Helens	06/19/2011	Casey
3	SH41	St. Helens	07/05/2011	Casey
4	SH42	St. Helens	07/23/2011	Casey
5	SH43	St. Helens	08/15/2011	Kelly
6	FJ12	Fiji	09/12/2011	Cooper
7	ET01	Etna	08/05/2011	Cooper
8	KE05	Kenya	05/31/2011	Kelly
9	KL18	Kilauea	07/08/2011	Malia
10	KL19	Kilauea	07/15/2011	Malia
11	KL20	Kilauea	07/22/2011	Malia
12	RD02	Reventador	07/11/2011	Carlos
13	VS11	Vesuvius	07/21/2011	Cooper
14	VS12	Vesuvius	08/15/2011	Cooper
15	KJ01	Kilimanjaro	06/09/2011	Kelly
16	KK03	Krakatau	07/19/2011	Kelly

Tour Bookings Data

There are two versions of the bookings data. One is a Microsoft Excel file, and the other is a SAS data set. They both contain the same columns and data values.

Filename: Bookings.xls Bookings.sas7bdat
File Type: Microsoft Excel spreadsheet SAS data set

Columns:

Name	Description
Office	Office where reservation was made
CustomerID	Customer identification number
Tour	Code for tour
Travelers	Number traveling in party
Deposit	Amount of deposit
Deposit_Date	Date of deposit

Bookings.xls

	A	B	C	D	E	F	G
1	Office	CustomerID	Tour	Travelers	Deposit	Deposit_Date	
2	Portland	SL28	SH43	10	425	7/5/2011	
3	Portland	DE27	PS27	6	75	7/11/2011	
4	Portland	SL34	FJ12	4	200	7/19/2011	
5	Portland	DI33	SH43	4	150	7/23/2011	
6	Portland	BU12	SH43	2	75	7/23/2011	
7	Portland	DE31	FJ12	3	175	7/25/2011	
8	Portland	WI48	FJ12	2	100	7/26/2011	
9	Portland	NG17	PS27	5	65	7/26/2011	
10	Portland	RA28	PS27	2	30	7/28/2011	
11	Portland	ME11	PS27	2	30	7/28/2011	
12	Portland	GI08	SH43	8	300	7/31/2011	
13	Portland	HI15	SH43	4	150	7/31/2011	
14	Portland	MA09	SH43	2	75	7/31/2011	
15							
16							
17							

Bookings.sas7bdat

	Office	CustomerID	Tour	Travelers	Deposit	Deposit_Date
1	Portland	SL28	SH43	10	425	05JUL2011
2	Portland	DE27	PS27	6	75	11JUL2011
3	Portland	SL34	FJ12	4	200	13JUL2011
4	Portland	DI33	SH43	4	150	23JUL2011
5	Portland	BU12	SH43	2	75	23JUL2011
6	Portland	DE31	FJ12	3	175	25JUL2011
7	Portland	WI48	FJ12	2	100	26JUL2011
8	Portland	NG17	PS27	5	65	26JUL2011
9	Portland	PA28	PS27	2	30	28JUL2011
10	Portland	ME11	PS27	2	30	28JUL2011
11	Portland	GI08	SH43	8	300	31JUL2011
12	Portland	HI15	SH43	4	150	31JUL2011
13	Portland	MA09	SH43	2	75	31JUL2011

Volcanoes Data

Filename: Volcanoes.sas7bdat
File Type: SAS data set

Columns:

Name	Description	Possible Values
Volcano	Name of the volcano	
Country	Country where the volcano is located	
Region	Region where the volcano is located	Af (Africa) An (Antarctica) AP (Australia/Pacific) As (Asia) Eu (Europe) NA (North America) SA (South America)
Height	Height of the volcano in meters	
Activity	Activity of the volcano	Active Extinct
Type	Kind of volcano	Caldera Complex Shield Stratovolcano

Volcanoes.sas7bdat

	Volcano	Country	Region	Height	Activity	Type
1	Altar	Ecuador	SA	5321	Extinct	Stratovolcano
2	Arthur's Seat	UK	Eu	251	Extinct	
3	Barren Island	India	As	354	Active	Stratovolcano
4	Elbrus	Russia	Eu	5633	Extinct	Stratovolcano
5	Erebus		An	3794	Active	Stratovolcano
6	Etna	Italy	Eu	3350	Active	Stratovolcano
7	Fuji	Japan	As	3776	Active	Stratovolcano
8	Garibaldi	Canada	NA	2678		Stratovolcano
9	Grimsvotn	Iceland	Eu	1725	Active	Caldera
10	Illimani	Bolivia	SA	6458	Extinct	Stratovolcano
11	Kenya	Kenya	Af	5199	Extinct	
12	Kilauea	USA	AP	1222	Active	Shield
13	Kilimanjaro	Tanzania	Af	5895		Stratovolcano
14	Kliuchevskoi	Russia	As	4835	Active	Stratovolcano
15	Krakatau	Indonesia	As	813	Active	Caldera
16	Lassen	USA	NA	3187	Active	Stratovolcano
17	Mauna Loa	USA	AP	4170	Active	Shield
18	Nyamuragira	DRCongo	Af	3058	Active	Shield
19	Nyiragongo	DRCongo	Af	3470	Active	Stratovolcano
20	Pinatubo	Philippines	As	1486	Active	Stratovolcano
21	Poas	Costa Rica	NA	2708	Active	Stratovolcano
22	Popocatepetl	Mexico	NA	5426	Active	Stratovolcano
23	Puy de Dome	France	Eu	1464	Extinct	Cinder Cone
24	Reventador	Ecuador	SA	3562	Active	Stratovolcano
25	Ruapehu	NZ	AP	2797	Active	Stratovolcano
26	Sabancaya	Peru	SA	5976	Active	Stratovolcano
27	Santorini	Greece	Eu	367	Active	Shield
28	Shishaldin	USA	NA	2857	Active	Stratovolcano
29	St. Helens	USA	NA	2549	Active	Stratovolcano
30	Vesuvius	Italy	Eu	1281	Active	Complex
31	Villarrica	Chile	SA	2847	Active	Stratovolcano
32	Warning	Australia	AP	1125	Extinct	Shield

Eruptions Data

There are two versions of the eruptions data. One is a text file, and the other is a SAS data set. They both contain the same columns and data values.

Filename:	Eruptions.csv	Eruptions.sas7bdat
File Type:	Text file with comma-separated values	SAS data set

Columns:

Name	Description	Possible Values
Volcano	Name of the volcano	
StartDate	Date the eruption started in MMDDYY10. format	
EndDate	Date the eruption ended in MMDDYY10. format	
VEI	Volcanic Explosivity Index	0–8

Eruptions.csv

Volcano, StartDate, EndDate, VEI
Barren Island, 12/20/1795, 12/21/1795, 2
Barren Island, 12/20/1994, 06/05/1995, 2
Erebus, 12/12/1912, . , 2
Erebus, 01/03/1972, . , 1
Etna, 02/06/1610, 08/15/1610, 2
Etna, 06/04/1787, 08/11/1787, 4
Etna, 01/30/1865, 06/28/1865, 2
Etna, 12/16/2005, 12/22/2005, 1
Fuji, 12/16/1707, 02/24/1708, 5
Grimsvotn, 10/31/1603, 11/01/1603, 2
Grimsvotn, 01/08/1873, 08/01/1873, 4
Grimsvotn, 12/18/1998, 12/28/1998, 3
Kilauea, 05/30/1840, 06/25/1840, 0
Kilauea, 05/24/1969, 07/22/1974, 0
Kliuchevskoi, 09/25/1737, 11/04/1737, 2
Kliuchevskoi, 03/25/1931, 03/27/1931, 4
Kliuchevskoi, 01/20/2005, 04/07/2005, 2
Krakatau, 05/20/1883, 10/21/1883, 6
Krakatau, 07/04/1938, 07/02/1940, 3
Krakatau, 05/29/2000, 10/30/2000, 1
Lassen, 05/30/1914, 06/29/1917, 3
Mauna Loa, 06/20/1832, 07/15/1832, 0
Mauna Loa, 03/25/1984, 04/15/1984, 0
Nyamuragira, 11/07/1907, 12/05/1907, 3
Nyamuragira, 02/06/2001, 04/05/2001, 2
Nyiragongo, 06/21/1982, 10/17/1982, 1
Nyiragongo, 01/17/2002, 02/03/2002, 1
Pinatubo, 04/02/1991, 09/02/1991, 5
Poas, 12/29/1898, 12/31/1907, 1
Poas, 04/08/1996, 04/08/1996, 1
Popocatepetl, 10/13/1663, 10/19/1665, 3
Popocatepetl, 12/21/1994, 08/05/1995, 2
Reventador, 12/12/1856, 12/13/1856, 3
Reventador, 02/24/1944, 03/01/1944, 3
Reventador, 11/03/2002, 01/10/2003, 4
Ruapehu, 02/13/1861, 05/16/1861, 2
Ruapehu, 06/17/1996, 09/01/1996, 3
Sabancaya, 05/01/1997, 05/02/1997, 3
Santorini, 09/27/1650, 12/06/1650, 4
Santorini, 05/23/1707, 09/14/1711, 3
Santorini, 01/26/1866, 10/15/1870, 2
Santorini, 01/10/1950, 02/02/1950, 2
Shishaldin, 03/13/1999, 05/27/1999, 3
St. Helens, 03/26/1847, 03/30/1847, 2
St. Helens, 03/27/1980, 10/28/1986, 5
St. Helens, 10/01/2004, 01/27/2008 , 2
Vesuvius, 12/15/1631, 01/31/1632, 5
Vesuvius, 12/25/1732, 06/04/1737, 3
Vesuvius, 12/18/1875, 04/22/1906, 4
Vesuvius, 07/05/1913, 04/04/1944, 3
Villarrica, 11/07/1837, 11/21/1837, 2
Villarrica, 10/26/2008, . , 1

Eruptions.sas7bdat

	Volcano	StartDate	EndDate	VEI
1	Barren Island	12/20/1795	12/21/1795	2
2	Barren Island	12/20/1994	06/05/1995	2
3	Erebus	12/12/1912	.	2
4	Erebus	01/03/1972	.	1
5	Etna	02/06/1610	08/15/1610	2
6	Etna	06/04/1787	08/11/1787	4
7	Etna	01/30/1865	06/28/1865	2
8	Etna	12/16/2005	12/22/2005	1
9	Fuji	12/16/1707	02/24/1708	5
10	Grimsvotn	10/31/1603	11/01/1603	2
11	Grimsvotn	01/08/1873	08/01/1873	4
12	Grimsvotn	12/18/1998	12/28/1998	3
13	Kilauea	05/30/1840	06/25/1840	0
14	Kilauea	05/24/1969	07/22/1974	0
15	Kliuchevskoi	09/25/1737	11/04/1737	2
16	Kliuchevskoi	03/25/1931	03/27/1931	4
17	Kliuchevskoi	01/20/2005	04/07/2005	2
18	Krakatau	05/20/1883	10/21/1883	6
19	Krakatau	07/04/1938	07/02/1940	3
20	Krakatau	05/29/2000	10/30/2000	1
21	Lassen	05/30/1914	06/29/1917	3
22	Mauna Loa	06/20/1832	07/15/1832	0
23	Mauna Loa	03/25/1984	04/15/1984	0
24	Nyamuragira	11/07/1907	12/05/1907	3
25	Nyamuragira	02/06/2001	04/05/2001	2
26	Nyiragongo	06/21/1982	10/17/1982	1
27	Nyiragongo	01/17/2002	02/03/2002	1
28	Pinatubo	04/02/1991	09/02/1991	5
29	Poas	12/29/1898	12/31/1907	1
30	Poas	04/08/1996	04/08/1996	1
31	Popocatepetl	10/13/1663	10/19/1665	3
32	Popocatepetl	12/21/1994	08/05/1995	2
33	Reventador	12/12/1856	12/13/1856	3
34	Reventador	02/24/1944	03/01/1944	3
35	Reventador	11/03/2002	01/10/2003	4
36	Ruapehu	02/13/1861	05/16/1861	2
37	Ruapehu	06/17/1996	09/01/1996	3
38	Sabancaya	05/01/1997	05/02/1997	3
39	Santorini	09/27/1650	12/06/1650	4
40	Santorini	05/23/1707	09/14/1711	3
41	Santorini	01/26/1866	10/15/1870	2
42	Santorini	01/10/1950	02/02/1950	2
43	Shishaldin	03/13/1999	05/27/1999	3
44	St. Helens	03/26/1847	03/30/1847	2
45	St. Helens	03/27/1980	10/28/1986	5
46	St. Helens	10/01/2004	01/27/2008	2
47	Vesuvius	12/15/1631	01/31/1632	5
48	Vesuvius	12/25/1732	06/04/1737	3
49	Vesuvius	12/18/1875	04/22/1906	4
50	Vesuvius	07/05/1913	04/04/1944	3
51	Villarica	11/07/1837	11/21/1837	2
52	Villarica	10/26/2008	.	1

Latitude and Longitude Data

There are two versions of the latitude and longitude data. One is a text file, and the other is a SAS data set. They both contain the same columns and data values.

Filename: Latlong.txt Latlong.sas7bdat
File Type: Fixed-width text file SAS data set

Columns:

Name	Description
Volcano	Name of the volcano
Latitude	Latitude
Longitude	Longitude

Latlong.txt

Volcano	Latitude	Longitude
Altar	-1.67	-78.42
Barren Island	12.28	93.52
Elbrus	43.33	42.45
Erebus	-77.53	167.17
Etna	37.73	15.00
Fuji	35.35	138.73
Garibaldi	49.85	-123.00
Grimsvotn	64.42	-17.33
Illimani	-16.39	-67.47
Kenya	-0.09	37.18
Kilauea	19.43	-155.29
Kilimanjaro	-3.07	37.35
Kliuchevskoi	56.06	160.64
Krakatau	-6.10	105.42
Lassen	40.49	-121.51
Mauna Loa	19.48	-155.61
Nyamuragira	-1.41	29.20
Nyiragongo	-1.52	29.25
Pinatubo	15.13	120.35
Poas	10.20	-84.23
Popocatepetl	19.02	-98.62
Puy de Dome	45.50	2.75
Reventador	-0.08	-77.66
Ruapehu	-39.28	175.57
Sabancaya	-15.78	-71.85
Santorini	36.40	25.40
Shishaldin	54.76	-163.97
St. Helens	46.20	-122.18
Vesuvius	40.82	14.43
Villarrica	-39.42	-71.93

Latlong.sas7bdat

	Volcano	Latitude	Longitude
1	Altar	-1.67	-78.42
2	Barren Island	12.28	93.52
3	Elbrus	43.33	42.45
4	Erebus	-77.53	167.17
5	Etna	37.73	15
6	Fuji	35.35	138.73
7	Garibaldi	49.85	-123
8	Grimsvothn	64.42	-17.33
9	Illimani	-16.39	-67.47
10	Kenya	-0.09	37.18
11	Kilauea	19.43	-155.29
12	Kilimanjaro	-3.07	37.35
13	Kliuchevskoi	56.06	160.64
14	Krakatau	-6.1	105.42
15	Lassen	40.49	-121.51
16	Mauna Loa	19.48	-155.61
17	Nyamuragira	-1.41	29.2
18	Nyiragongo	-1.52	29.25
19	Pinatubo	15.13	120.35
20	Poas	10.2	-84.23
21	Popocatepetl	19.02	-98.62
22	Puy de Dome	45.5	2.75
23	Reventador	-0.08	-77.66
24	Ruapehu	-39.28	175.57
25	Sabancaya	-15.78	-71.85
26	Santorini	36.4	25.4
27	Shishaldin	54.76	-163.97
28	St. Helens	46.2	-122.18
29	Vesuvius	40.82	14.43
30	Villarrica	-39.42	-71.93

Portland Flights Data

Filename: Portland.sas7bdat

File Type: SAS data set

Columns:

Name	Description
Origin	City from which flight departs
Destination	City in which flight arrives
FlightNo	Flight number
FlightPrice	Price of flight in U.S. dollars

Portland.sas7bdat

	Origin	Destination	FlightNo	FlightPrice
1	Portland	Catania	L469	\$779.00
2	Portland	Hilo	HA25	\$703.00
3	Portland	Nairobi	KLM6034	\$1,833.00
4	Portland	Rome	D1576	\$644.00
5	Portland	San Jose	CA1210	\$494.00
6	Portland	Tokyo	UA383	\$705.00

Seattle Flights Data

Filename: Seattle.sas7bdat

File Type: SAS data set

Columns:

Name	Description
Origin	City from which flight departs
Destination	City in which flight arrives
FlightNo	Flight number
FlightPrice	Price of flight in U.S. dollars

Seattle.sas7bdat

	Origin	Destination	FlightNo	FlightPrice
1	Seattle	Catania	BA48	\$802.00
2	Seattle	Hilo	HA21	\$677.00
3	Seattle	Jakarta	AA119	\$1,815.00
4	Seattle	Nairobi	KLM6034	\$1,761.00
5	Seattle	Quito	CA1086	\$833.00
6	Seattle	Rome	USA6	\$596.00
7	Seattle	San Jose	CA1100	\$480.00
8	Seattle	Tokyo	UA875	\$721.00

Northwest Weather Data

Filename: NWweather.sas7bdat
 File Type: SAS data set

Columns:

Name	Description	Possible Values
City	City	
Month	Month	1-12
AvgTemp	Average high temperature in degrees Fahrenheit	
FromNormal	Change from normal temperature in degrees Fahrenheit	
InchesRain	Amount of rain in inches	
Bookings	Number of tours booked for that month	

NWweather.sas7bdat

	City	Month	AvgTemp	FromNormal	InchesRain	Bookings
1	Seattle	1	45.8	4.9	8.39	32
2	Seattle	2	41.7	-1.6	1.76	19
3	Seattle	3	46.8	0.6	6.34	23
4	Seattle	4	48.9	-1.3	2.74	18
5	Seattle	5	54.8	-1	1.16	21
6	Seattle	6	62.8	2.1	0.51	18
7	Seattle	7	67.9	2.6	0.06	17
8	Seattle	8	66.4	0.8	0.32	17
9	Seattle	9	62.6	1.5	0.89	22
10	Seattle	10	54.3	1.6	8.96	20
11	Seattle	11	42.8	-2.4	6.77	25
12	Seattle	12	41.8	1.1	3.88	31
13	Portland	1	44.8	4.9	7.64	22
14	Portland	2	44.3	1.2	2.37	19
15	Portland	3	49	1.8	5.75	17
16	Portland	4	50.8	-0.4	4.37	18
17	Portland	5	57.3	0.2	1.49	21
18	Portland	6	66.1	3.4	0.31	19
19	Portland	7	71.6	3.5	0	13
20	Portland	8	70.1	1.6	0.19	15
21	Portland	9	66.8	3.2	0.85	23
22	Portland	10	58.2	3.9	3.01	24
23	Portland	11	45.2	-0.6	4.09	27
24	Portland	12	42.5	2.3	7.45	33

On-Time Status of Flights Data

Filename: OnTimeStatus.sas7bdat

File Type: SAS data set

Columns:

Name	Description	Possible Values
Month	Month	Dec Jan Feb
Departure	Whether delayed 15 minutes or more	OnTime Late
TimeOfDay	Whether before or after 12 p.m.	BeforeNoon AfterNoon
NumOfFlights	Number of flights in that category	

OnTimeStatus.sas7bdat

	Month	Departure	TimeOfDay	NumOfFlights
1	Dec	OnTime	AfterNoon	25
2	Dec	Late	AfterNoon	33
3	Dec	OnTime	BeforeNoon	80
4	Dec	Late	BeforeNoon	13
5	Jan	OnTime	AfterNoon	15
6	Jan	Late	AfterNoon	12
7	Jan	OnTime	BeforeNoon	30
8	Jan	Late	BeforeNoon	13
9	Feb	OnTime	AfterNoon	20
10	Feb	Late	AfterNoon	8
11	Feb	OnTime	BeforeNoon	43
12	Feb	Late	BeforeNoon	1

Advertising Results Data

Filename: AdResults.sas7bdat
File Type: SAS data set

Columns:

Name	Description	Possible Values
City	City	
Month	Month	1–12
AdDollars	Money spent on advertising in U.S. dollars	
Bookings	Number of tours booked for that month	

AdResults.sas7bdat

	City	Month	AdDollars	Bookings
1	Seattle	1	350	30
2	Seattle	2	330	19
3	Seattle	3	525	32
4	Seattle	4	400	18
5	Seattle	5	375	21
6	Seattle	6	325	18
7	Seattle	7	150	17
8	Seattle	8	250	17
9	Seattle	9	250	22
10	Seattle	10	325	20
11	Seattle	11	400	25
12	Seattle	12	500	31
13	Portland	1	325	25
14	Portland	2	290	19
15	Portland	3	250	17
16	Portland	4	300	18
17	Portland	5	300	21
18	Portland	6	275	19
19	Portland	7	200	13
20	Portland	8	225	15
21	Portland	9	325	23
22	Portland	10	350	24
23	Portland	11	425	27
24	Portland	12	400	33

Ages Data

Filename: Ages.sas7bdat
File Type: SAS data set

Columns:

Name	Description	Possible Values
Difficulty	Strenuousness of the tour	c (Challenging) m (Moderate) e (Easy)
Age	Age in years	

Ages.sas7bdat

	Difficulty	Age
1	e	45
2	e	38
3	e	65
4	e	43
5	e	29
6	e	72
7	e	66
8	e	57
9	e	39
10	e	33
11	m	26
12	m	37
13	m	42
14	m	27
15	m	31
16	m	39
17	m	35
18	m	30
19	m	41
20	m	55
21	c	65
22	c	39
23	c	59
24	c	55
25	c	50
26	c	47
27	c	42
28	c	60
29	c	58
30	c	67

Index

A

Access files 169, 186-187
active data table 142
addition in expression editor 178-179, 206-207
advanced expression 81-85, 179, 206-207
advanced filters 236-237
ALL universal class variable 271
analysis of variance 294-295
AND logic in filters 99, 234-235
ANOVA procedure (One-Way ANOVA task) 294-295
ANYDTDTew.informat 152-153
Append Table task 242-243
appending
 data tables 240, 242-243
 rows in Data Grid 177
arguments in functions 208-209
arranging items in a process flow 140
ascending sort order 101, 227-228, 231
ASCII sort order 229
Assign Project Library task 172-173
assignment of values in a query 218-219
associations, measures of 288-289
Auto Arrange option for process flows 140
Auto Hide 134
automatically running SAS code 165
axes 308-309

B

Background Color option for process flows 140
Bar Chart task 298-299
batch programs, running 156-157
BESTw.format 154-155
box area in Summary Tables 275

C

character data
 definition 12, 146-147
 formats 154-155
 functions 210-211
 informats 152-153
 labels for 194-195
 user-defined formats 194-195

charts

 bar 298-299
 pie 300-301
 result format 313

chi-square test 288-289

Cochran-Mantel-Haenszel statistics

 288-289

code files, SAS

 code template 159
 editing 158-159
 embedding in project 161
 inserting code in task 158-159

 item in project 139

 log 162

 options 165

 prompts in 334-335

 properties 160-161, 334-335

 running 158-161

 server for 159

 stopping 161

 viewing 137, 158-159

 writing 160-161

collating sequence 229

colors

 in styles 314-317
 in Summary Tables 274-281

columns

 attributes when importing 181, 183, 185

 computing in Data Grid 178-179

 computing in query 80-86, 206-209

 definition 24, 146

 deleting from Data Grid 16, 177

 deleting in a query 90, 204

 inserting in Data Grid 176-179

 labels 51-56, 150-151

 list of 149

 names 8-9, 147

 order of 90-91, 164-165

 properties 8-17, 87-89, 150-151, 174-177,

 190-193, 198-199, 205, 217, 266-267

 recoding 81, 218-219

 selecting in a query 78-79, 204-205

 sorting data values 226-231

combining data tables 109-117, 123-128, 240-247

 filters 234-235

 results 318-321

comma-delimited files

 exporting 186-187

 reading 168-169, 182-183

362 Index

- COMMAw.d
 - format 154-155
 - informat 152-153
- compound filters 234-235
- computing columns
 - in Data Grid 178-179
 - in query 80-86, 206-209, 212-217
- concatenating data tables 240, 242-243
- conditional selection of rows 92-99, 224-227, 232-237, 328-329
- conditions for projects 338-341
- confidence limits 284-287, 293
- contents of data tables 148-149
- copying
 - data tables 176
 - items in process flows 141
- Correlations task (CORR procedure) 290-291
- COUNT statistic in query 212-217
- counts
 - in Summary Tables 270-271
 - one-way 258-259
 - two-way 260-261
- Create Format task 57-62, 194-197
- cross-tabulations
 - in Summary Tables 270-271
 - in Table Analysis 260-261, 288-289
- CSV files
 - exporting 186-187
 - reading 168-169, 182-183
- currency data
 - definition 13, 146
 - formats 154-155
 - informats 152-153
- D**
 - data engines, SAS 173
 - data files
 - exporting 186-187
 - options 164-165
 - reading 168-171, 180-185
 - types 168-169
 - Data Grid
 - appending rows 177
 - computing columns 178-179
 - creating 5-21, 174-175
 - definition 146
 - deleting columns 16, 177
 - deleting rows 20, 177
 - editing 19, 176-177
- formats, applying 190-191
- inserting columns 176-179
- inserting rows 177
- data sets, definition 6, 146
- data tables
 - active 142
 - appending 240, 242-243
 - compared to data sets 6
 - copying 176
 - creating summary 264-265
 - creating with a query 220-221
 - definition 6, 146
 - editing 19, 150, 176-177
 - exporting 186-187
 - filtering 92-99, 118-123, 224-227, 232-237, 328-329
 - joining 109-117, 123-128, 240-241, 244-247
 - locations 170-171
 - merging 109-117, 123-128, 240-241, 244-247
 - names 6-7, 174
 - new 5-21, 174-175
 - opening 29, 40-43, 75-76, 170
 - opening large 164-165
 - options 164-165
 - properties 148-149
 - sorting 100-102, 226-231
 - subsetting 92-99, 118-123, 224-227, 232-237, 328-329
- data values
 - appearance of 51-56, 190-193
 - groups 13, 146-147, 150-151
 - labels for 57-62, 194-199
 - length of 9, 12
 - missing 18, 147
 - options 164-165
 - recoding in a query 218-219
 - sorting 100-102, 226-231
 - summarizing 212-217, 264-265
 - types 9-14, 146-147, 150-151
- data views 220-221
- date data
 - computations with 206-207
 - definition 146-147
 - formats 154-155, 190-193
 - functions 210-211
 - informats 152-153
- DATEPART function 210-211
- datetime data
 - definition 146-147
 - functions 210-211

DATETIMEw.d
 format 154-155
 informat 152-153

DATEw.
 format 154-155
 informat 152-153

DAY function 208-211

DB2 169

dBASE files 169, 186-187

DDMMYYw. informat 152-153

deleting
 columns from Data Grid 16, 177
 columns in a query 90, 204
 duplicate rows 229
 items from a project 139
 process flows 140
 rows from Data Grid 20, 177
 task templates 144

delimited files
 exporting 186-187
 reading 182-183
 types 168-169

descending sort order 101, 227-228, 231

descriptive statistics
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 262-263, 286-287
 in Summary Tables 272-273

display formats
 applying standard 151
 applying user-defined 63-65, 198-199
 creating user-defined 57-62, 194-197
 grouping with 266-267
 in Data Grid 176-177, 190-191
 in query 204-205, 217
 in tasks 51-56, 192-193
 names 58
 table of 154-155

Distribution Analysis task 284-285

division in expression editor 178-179, 206-207

docked windows 135

document windows 135

documenting
 data tables 147-149
 projects 34-35

DOLLARw.d format 154-155

DTDATEw. format 154-155

duplicate rows, deleting 229

E

EBCDIC sort order 229

Edit button in task window 142, 224-225

editing
 data tables 19, 150, 176-177
 reports 318-321
 SAS programs 158-159
 styles 316-317

EGTASK library 7, 170

embedding program in project 161

engines, SAS data 173

errors 162

EURDFDDw. format 154-155

EUROXw.d format 154-155

Ew. format 154-155

Excel files 169
 exporting 186-187
 importing 180-181

Exchange files 169

Export wizard 186-187

exporting
 data 186-187
 items from a project 139
 results 322-323
 task templates 145

expression editor
 in a filter 236-237
 in a query 82-86, 206-209
 in Data Grid 179

F

filename of data table 149

files, data
 exporting 186-187
 options 164-165
 reading 168-171, 180-185
 types 168-169

filtering data
 compound filters 234-235
 dynamic 328-329
 Filter and Sort task 226-227
 in a query 92-99, 118-123, 232-233
 in a task 224-225
 summarized data 232-233
 using expressions 236-237

Fisher's Exact test 288-289

fixed columns raw data files 184-185

364 Index

- fonts
 - in graphs 308
 - in reports 320-321
 - in styles 316-317
 - in Summary Tables 274-281
- footers, adding to reports 320-321
- footnotes
 - changing 46-50, 252-253
 - changing default 164-165
 - default 48
- FORMAT procedure (Create Format task) 57-62, 194-197
- formats
 - applying standard 151
 - applying user-defined 63-65, 198-199
 - creating user-defined 57-62, 194-197
 - grouping with 266-267
 - in Data Grid 176-177, 190-191
 - in query 204-205, 217
 - in tasks 51-56, 192-193
 - names 58
 - table of 154-155
- FREQ procedure
 - (One-Way Frequencies task) 22-28, 258-259
 - (Table Analysis task) 260-261, 288-289
- frequencies
 - cross-tabulations 260-261
 - in Summary Tables 270-271
 - one-way 22-28, 258-259
 - statistical analysis 288-289
- full outer join 241, 246-247
- functions
 - definition 82
 - in Data Grid 179
 - in filter conditions 236-237
 - in query 208-209
 - table of 210-211
- G**
- GCHART procedure
 - (Bar Chart task) 298-299
 - (Pie Chart task) 300-301
- GPLOT procedure
 - (Line Plot task) 302-305
 - (Scatter Plot task) 28-32, 306-307
- graphs
 - bar charts 298-299
 - controlling axes 308-309
 - line plots 302-305
- pie charts 300-301
- regression plot 293
- result format 313
- scatter plots 28-32, 306-307
- Grid option for process flows 140
- grouping data
 - cross-tabulations 260-261
 - frequency reports 258-259
 - in query 212-217
 - list reports 254-257
 - output data table 264-265
 - recoding values 218-219
 - summary reports 262-263, 270-273
 - with user-defined formats 194-197, 266-267
- H**
- headings
 - adding to reports 320-321
 - changing in a task 51-56
 - in Summary Tables 274-277
- help, SAS Technical Support x
- hiding windows 135
- histograms 298-299
- Hoeffding correlation coefficients 290-291
- HTML data files 169, 186-187
- HTML results 68-69, 312-313
 - styles for 314-317
- I**
- images
 - adding to reports 320-321
 - adding to style 316-317
- Import Data wizard 180-185
- importing data 171
 - delimited raw data 182-183
 - Excel files 180-181
 - fixed columns raw data 184-185
 - JMP files 169
 - SPSS files 169
 - Stata files 169
 - types of data 168-169
- importing, task templates 145
- informats
 - in Data Grid 150-151, 176-177
 - table of 152-153
- inner join 241, 246-247
- inserting
 - columns in Data Grid 176-179
 - rows in Data Grid 177
 - SAS code in a task 158-159

J

JMP files, importing 169
 joining data tables 109-117, 240-241, 244-245
 join types 240-241
 modifying a join 123-128, 246-247
 JULIANw.
 format 154-155
 informat 152-153

K

Kendall correlation coefficients 290-291
 kurtosis 284-285

L

labels
 axes 308-309
 columns 51-56, 150-151
 data values 194-199
 headings in Summary Tables 274-275
 Layout option for process flows 140
 left join 241, 246-247
 LENGTH function 210-211
 length of data values 9, 12, 150-151
 LIBNAME statement 165, 171
 libraries
 for data 7, 41-42, 170-174
 for formats 57, 194
 temporary data 7, 170-171
 libref 170, 172
 Line Plot task 302-305
 Linear Regression task 292-293
 linking items in process flows 141
 List Data task 43-56, 63-65, 250-251, 254-257
 listing files (text output) 312-313
 log files, SAS 162-163
 LOG function 210-211
 LOG10 function 210-211
 Lotus 1-2-3 files 169, 186-187

M

macro variables
 creating for data values 326-327
 creating for text values 336-337
 creating for variable names 330-331
 using in filters 328-329
 using in programs 334-335
 using in project conditions 338-341
 using in tasks 332-333
 many-to-many merge 240

mathematical

functions 210-211
 operators 178-179, 206-207
 MAX
 function 210-211
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 272-273

maximizing workspace 137

MDY function 210-211

MEAN

function 210-211
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 272-273

MEANS procedure (Summary Statistics task)
 262-265, 286-287

means, comparison of 294-295

median

in Distribution Analysis 284-285
 in Summary Statistics 286-287
 in Summary Tables 272-273

menus

compared to Task List window 22

merging data tables 109-117, 123-128, 240-241,
 244-247

MIN

function 210-211
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 272-273

missing data values 18, 147

number of 212-217, 263, 272-273, 286-287

MMDDYYw.

format 154-155
 informat 152-153

mode statistic 284-285

modifying

joins 123-128, 246-247
 tasks 25-26, 46, 143

moments 284-285

MONTH function 210-211

moving items in process flows 140

multiple comparison 294-295

multiplication in expression editor 178-179,
 206-207

My Styles, creating 316-317

366 Index

N

N statistic
 in Distribution Analysis 284-285
 in query (COUNT) 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 270-273
 names
 column 8-9, 147, 150-151, 181, 183, 185
 data tables 6-7, 174
 format 58
 library 172
 prompts 326, 329, 330
 New Computed Column wizard 81-85, 206-209, 218-219
 New Data wizard 5-17, 174-175
 New Filter wizard 93-98, 119-121, 232-237
 NMISS statistic
 in query 212-217
 in Summary Tables 272-273
 normality, tests for 284-285
 notes 34-35, 139
 numeric data
 definition 12, 146-147
 formats 154-155
 groups 13, 146-147
 informats 152-153
 labels for 196-197
 user-defined formats 196-197

O

observations, definition 146
 ODBC 169
 one-to-many merge 240
 one-to-one merge 240
 One-Way ANOVA task 294-295
 One-Way Frequencies task 22-28, 258-259
 opening files 168-171
 delimited files 182-183
 Excel files 180-181
 fixed columns raw data 184-185
 SAS data tables 40-43, 75-76, 170
 opening items in a project 139
 opening tasks 22, 25-26, 142
 operators
 computing new columns using 178-179, 206-207
 in filter conditions 224-227, 232-237, 328-329
 options
 changing default footnote 48

data 164-165
 graph format 313
 query 220-221
 resetting 165
 restore window layout 4, 135, 165
 result format 68-70, 165, 312-313
 result style 69, 165, 314-315
 SAS programs 165
 titles 164-165
 VALIDVARNAME= 147
 OR logic in filters 99, 234-235
 Oracle 169
 ordering
 columns in a query 90-91
 columns in task windows 164-165
 rows with a query 100-102, 230-231
 rows with Filter and Sort task 226-227
 rows with Sort Data task 228-229
 outer join 241, 246-247
 output
 combining 318-321
 data tables 264-265
 graph format 313
 options 164-165
 report format 68-70, 165, 312-313
 report style 165
 saving 322-323
 style for 66-69, 314-317

P

Paradox files 169, 186-187
 parameters 326
 PDF
 results 68-70, 312-313
 styles for 314-315
 Pearson correlation coefficients 290-291
 Pearson tests of association 288-289
 percentages
 in One-Way Frequencies 258-259
 in Summary Tables 272-273
 in Table Analysis 260-261
 percentiles
 in Distribution Analysis 284-285
 in Summary Statistics 286-287
 in Summary Tables 272-273
 PERCENTw.d
 format 154-155
 informat 152-153
 performance, improving 164-165
 Pie Chart task 300-301
 pivot table 270-273

- plots
 controlling axes 308-309
 line 302-305
 regression 293
 result format 313
 scatter 28-32, 306-307
- previewing
 query results 78
 SAS code 158
- PRINT procedure (List Data task) 43-56, 63-65, 250-251, 254-257
- printing
 process flows 141
 reports 321
- procedure titles in results 164-165, 321
- procedures
 ANOVA (One-Way ANOVA task) 294-295
 CORR (Correlations task) 290-291
 FORMAT (Create Format task) 57-62, 194-197
 FREQ (One-Way Frequencies task) 22-28, 258-259
 FREQ (Table Analysis task) 260-261, 288-289
 GCHART (Bar Chart task) 298-299
 GCHART (Pie Chart task) 300-301
 GPLOT (Line Plot task) 302-305
 GPLOT (Scatter Plot task) 28-32, 306-307
 MEANS (Summary Statistics task) 262-265, 286-287
 PRINT (List Data task) 43-56, 63-65, 250-251, 254-257
 REG (Linear Regression task) 292-293
 SORT 228-229
 SQL (Append Table task) 242-243
 SQL (Query Builder) 202-221, 230-237, 240-241, 244-247, 328-329
 SUMMARY (Summary Statistics task) 262-265, 286-287
 TABULATE (Summary Tables task) 270-281
 UNIVARIATE (Distribution Analysis task) 284-285
- process flows 4, 32-33, 134-135, 138-141
- programs, SAS
 code template 159
 editing 158-159
 embedding in project 161
 inserting code in task 158-159
 item in project 139
 log 162
 options 165
- prompts in 334-335
 properties 160-161, 334-335
 running 158-161
 server for 159
 stopping 161
 viewing 137, 158-159
 writing 160-161
- Project Log 162-163
- Project Tree 4, 33, 134, 138-139
- projects
 conditions 338-341
 definition 3, 138-139
 documenting 34-35
 saving 36-37, 139
 scheduling 156-157
- Prompt Manager window 135-136, 326-327, 330-331, 336-337
- prompts
 creating for data values 326-327
 creating for text values 336-337
 creating for variable names 330-331
 using in filters 328-329
 using in programs 334-335
 using in project conditions 338-341
 using in tasks 332-333
- properties
 of columns 8-17, 87-89, 150-151, 174-175, 190-193, 198-199, 205, 217, 266-267
 of data tables 148-149
 of items in project 139
 of joins 124-126, 246-247
 of projects 157
 of SAS programs 160-161, 334-335
 of tasks 66-69, 313
 protecting data 150
- Q**
- QTR function 210-211
- quantiles
 in Distribution Analysis 284-285
 in Summary Statistics 286-287
 in Summary Tables 272-273
- queries
 computing new columns 80-86, 206-209
 filtering data 92-99, 118-123, 232-237, 328-329
 joining data tables 109-117, 123-128, 240-241, 244-247
 naming 203
 options 220-221

368 Index

queries (*continued*)

- order of columns 90-91, 204
- prompts in 328-329
- recoding values 218-219
- result name 203
- result type 102, 220-221
- selecting columns 78-79, 204-205
- sorting rows 100-102, 230-231
- summary groups 214-217

R

range statistic

- in Distribution Analysis 284-285
- in query 212-217
- in Summary Statistics 263, 286-287
- in Summary Tables 272-273

raw data files

- delimited 182-183
- fixed columns 184-185
- types 168

read-in formats

- in Data Grid 150-151, 176-177
- table of 152-153

read-only mode

- 150, 176

reading

- delimited raw data 182-183
- Excel files 180-181
- fixed columns raw data 184-185
- SAS data tables 40-43, 75-76, 170
- types of data 168-169

recoding values in a query

- 81, 218-219

refreshing

- query preview 86
- Server List window 173
- tasks 143

REG procedure (Linear Regression task)

- 292-293

regression, linear

- 292-293
- renaming
 - columns 150
 - columns from Excel 181
 - columns from raw data 183, 185
 - items in a project 139
 - process flows 140
 - queries 202

reopening tasks

- 25-26, 46, 143

replacing values in a query

- 218-219

reports

- creating custom 318-321
- creating with a query 220-221
- cross-tabulations 260-261

footnotes

- 164-165, 252-253
- frequency 22-28, 258-259
- grouping with formats 266-267
- list 43-56, 250-251
- list by group 254-255
- list with totals 256-257
- SAS report result format 318-321
- summary statistics 262-263
- summary tables 270-281
- titles 164-165, 252-253
- titles of procedures in 164-165, 321

resetting options

- 165
- Resources pane 4-5, 134-136, 170-171, 326-327, 330-331, 336-337

restoring windows

- 4, 135, 165
- results 138
 - combining 318-321
 - footnotes 164-165, 252-253
 - format 68-70, 165, 312-313
 - graph format 313
 - queries 102, 202-203, 220-221
 - saving 322-323
 - style of 66-69, 165, 314-317
 - titles 164-165, 252-253
 - titles of procedures in 164-165, 321
 - viewing 137

right join

- 241, 246-247
- row number in list reports 51, 251, 255

rows

- appending in Data Grid 177
- definition 146
- deleting duplicates 229
- deleting from Data Grid 20, 177
- grouping in a report 254-255
- inserting in Data Grid 177
- selecting 92-99, 118-123, 224-227, 232-233, 236-237

RTF

- results 68-69, 312-313
- styles for 314-315

running

- projects 156-157
- projects with conditions 340-341
- SAS programs 158-161, 165
- tasks 143

S

SAS data engines

- 173
- SAS Folders window 135-136
- SAS listing (text) output 312-313

SAS log, viewing 162-163
 SAS names 8, 147, 172, 174, 181, 183, 185
 SAS Report result format 45, 68-69, 312-313, 318-321
 styles for 314-317
SAS/ACCESS 169
 Excel files 181
SASUSER library 7, 41-42, 170
 saving
 data 186-187
 items in a project 139
 projects 36-37, 139
 results 220-221, 322-323
 SAS programs 160-161
 Scatter Plot task 28-32, 306-307
 scheduling projects 156-157
 Scheffe's multiple comparison procedure 294-295
 scientific notation format 154-155
 selecting
 columns in a query 78-79, 204-205
 rows in a query 92-99, 118-123, 232-235
 Server List window 135-136, 170-171, 173
 server, SAS
 for data 41-42, 170
 for program 159, 161
 for task 143
 Show Multiple 136
 Show One 136
 skewness 284-285
 Sort Data task (SORT procedure) 228-229
 sorting data values
 in a query 100-102, 230-231
 in Filter and Sort task 226-227
 in Sort Data task 228-229
 space-delimited files
 exporting 186-187
 reading 168-169, 182-183
 Spearman correlation coefficients 290-291
 splitting
 Project Log 163
 Resources pane 136
 workspace 137
 spreadsheets 180-181, 186-187
 SPSS files, importing 169
 SQL procedure
 (Append Table task) 242-243
 (Query Builder) 202-221, 230-237, 240-241, 244-247, 328-329
 stacking workspace 137

standard deviation
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 272-273
 standard error
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287
 in Summary Tables 272-273
 starting SAS Enterprise Guide 3
 Stata files, importing 169
 statistics
 analysis of variance 294-295
 association 288-289
 correlation 290-291
 descriptive 212-217, 262-263, 272-273, 284-287
 distribution 284-285
 functions 210-211
 multiple comparisons 294-295
 regression 292-293
 saving to a data table 264-265
 stopping
 SAS programs 159, 161
 tasks 143
 Style Manager 316-317
 styles
 changing 66-69, 165, 314-315
 creating new 316-317
 subsetting data tables 92-99, 118-123, 224-227, 232-237, 328-329
 SUBSTR function 210-211
 subtraction in expression editor 178-179, 206-207
 summarized columns
 filtering 232-233
 groups 212-217
 in query 81-85, 212-215
 Summary Statistics task (SUMMARY procedure) 262-265, 286-287
 Summary Tables task 270-281
 box area properties 275
 class level properties 276-277
 data value properties 280-281
 heading properties 274-275
 table properties 278-279
 sums
 function 210-211
 in List Data 256-257

370 Index

- sums (*continued*)
in query 212-217
in Summary Statistics 263, 286-287
in Summary Tables 272-273
- T**
- tab-delimited files
exporting 186-187
reading 168-169, 182-183
- Table Analysis task 260-261, 288-289
- TABULATE procedure (Summary Tables task)
270-281
- Task List window 4-5, 135-136, 142
compared to menu bar 22
- task roles, assigning 24, 250
- Task Status window 5, 135
- tasks 138
Append Table 242-243
Assign Project Library 172-173
assigning formats 51-56, 192-193
Bar Chart 298-299
code generated by 158-159
compared to wizards 22
Correlations 290-291
Create Format 57-62, 194-197
data for 224-225
Distribution Analysis 284-285
dynamic 330-333
Filter and Sort 226-227
filtering in 224-225
footnotes 46-50, 164-165, 252-253
Line Plot 302-305
Linear Regression 292-293
List Data 43-56, 63-65, 250-251, 254-257
log 162
modifying 25-26, 46, 143
One-Way ANOVA 294-295
One-Way Frequencies 22-28, 258-259
opening 22, 25-26, 46, 142-143
options for 164-165
Pie Chart 300-301
prompts in 330-333
properties 66-69, 313
Query Builder 244-247
refreshing 143
result format 312-313
running 143
Scatter Plot 28-32, 306-307
server for 143
Sort Data 228-229
- stopping 143
Summary Statistics 262-265, 286-287
Summary Tables 270-281
Table Analysis 260-261, 288-289
templates 144-145
titles 46-50, 164-165, 252-253
- Technical Support x
- templates
code 159
task 144-145
- temporary library 7, 170
- text
adding to reports 320-321
results files 68-69, 312-313
- text data files
exporting 186-187
reading delimited 182-183
reading fixed columns 184-185
- tick marks 308-309
- time data
definition 147
formats 154-155
informats 152-153
- TIMEw.d
format 154-155
informat 152-153
- titles
changing 46-50, 252-253
changing default 164-165
of procedures in results 164-165, 321
prompts in 332-333
- TODAY function 210-211
- toolbar, workspace 25, 135
- totals
in List Data 256-257
in query 212-215
in Summary Statistics 262-265
in Summary Tables 272-273
- TRANSLATE function 210-211
- U**
- UNIVARIATE procedure (Distribution Analysis task) 284-285
- UPCASE function 210-211
- update mode 150, 176
- uppercase
format 154-155
informat 152-153
- user-defined formats
applying 63-65, 198-199

creating character 57-62, 194-195
 creating numeric 196-197
 grouping with 266-267

V

VALIDVARNAME= option 147
 variables, definition 24, 146
 variance
 in Distribution Analysis 284-285
 in query 212-217
 in Summary Statistics 263, 286-287

W

w.d
 format 154-155
 informat 152-153
 warnings 162-163
 WEEKDATEw. format 154-155
 windows 4-5, 134-135
 document 135
 hiding 135
 maximizing workspace 137
 Process Flow 4, 32-33, 134-135, 138-139
 Project Tree 4, 33, 134, 138-139
 Resources pane 134-135, 170-171, 326-327,
 330-331, 336-337
 restore window layout 4, 135, 165
 splitting workspace 137
 wizards
 Assign Project Library 172-173
 compared to tasks 22
 Export 186-187, 322-323
 Import Data 180-185
 New Computed Column 81-85, 206-209,
 218-219
 New Data 5-17, 174-175
 New Filter 93-98, 119-121, 232-237
 WORDDATEw. format 154-155
 WORK library 7, 170
 workspace 4, 134-135, 137
 toolbar 25
 Workspace Layout icon 137

Z

Zoom option for process flows 140

Special Characters

\$UPCASEw.
 format 154-155
 informat 152-153
 \$w.
 format 154-155

informat 152-153
&
 in data icon 332
 in name of prompt 328
? in icon 329, 333, 335
 FREQ variable 265
 TYPE variable 265
 WAY variable 265

