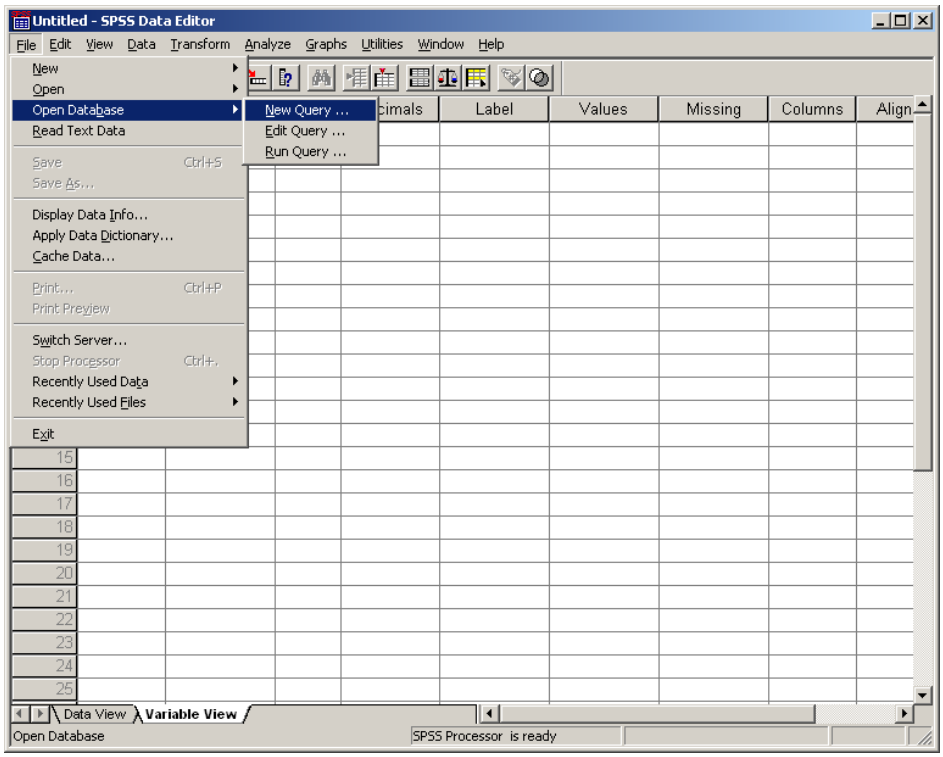
Data Warehouse Users Group Meeting
July 16, 2002
Presenter: Vince Timbers
Phone 865-4253 Email: vlt@psu.edu

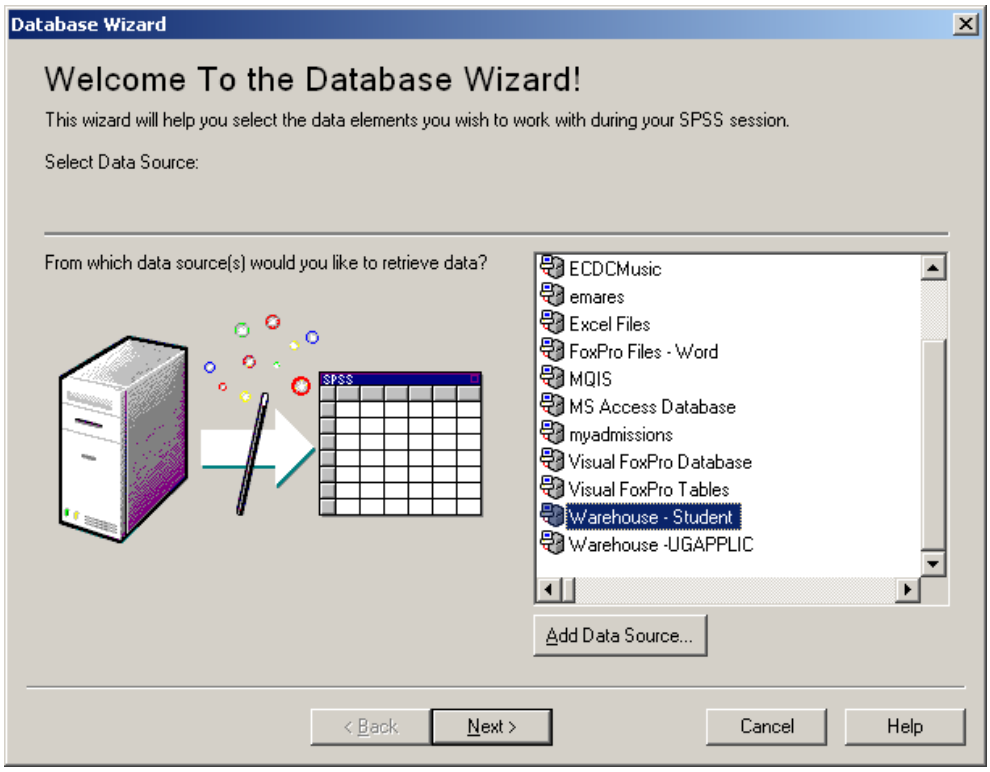
Accessing the Penn State Data Warehouse with SPSS

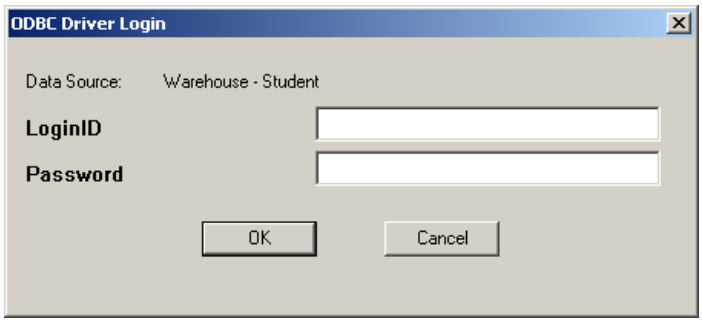
Using SPSS to retrieve data from the data warehouse is quite simple when using the Database Wizard which provides thorough directions. Screen shots from the wizard are provided below with supplemental notes.

To get started select **File**, **Open Database** and **New Query** from the menu bar to open the Database Wizard window.

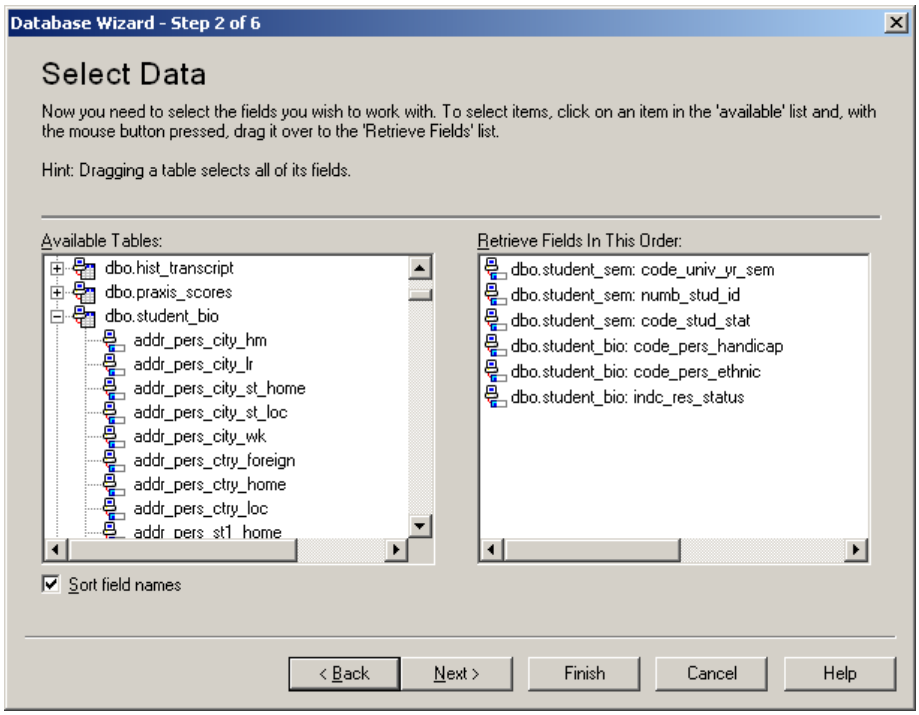


Available data sources are those defined by the ODBC Data Source Administrator. It is important to note that a data source can only access data in one database. In the screen below there are two data sources listed for the data warehouse, one for the STUDENT database and one for the UGAPPLIC database.

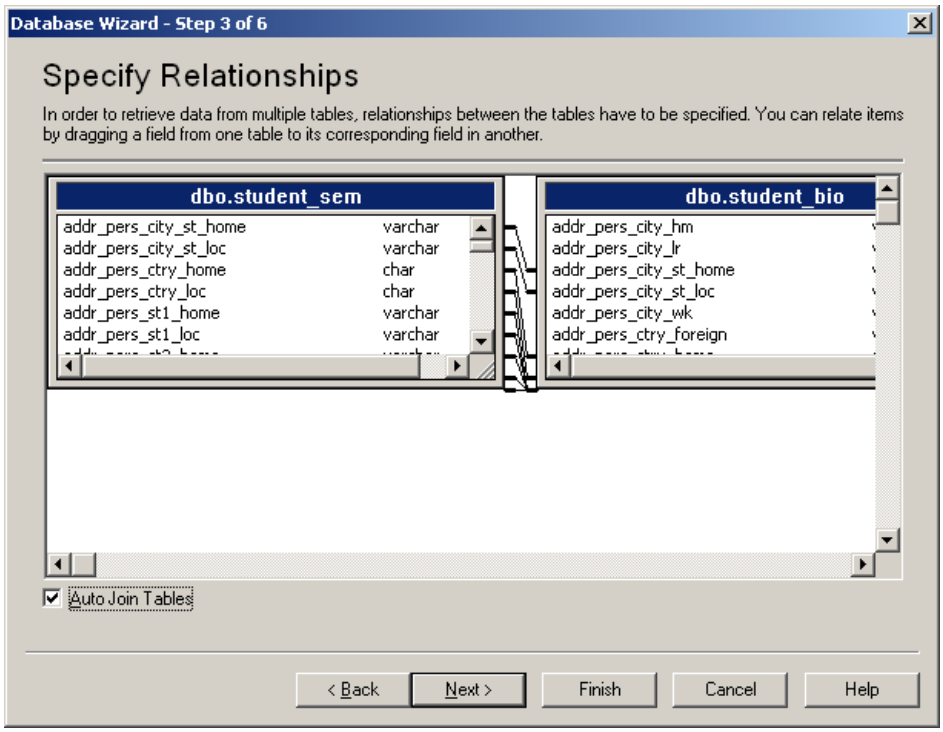




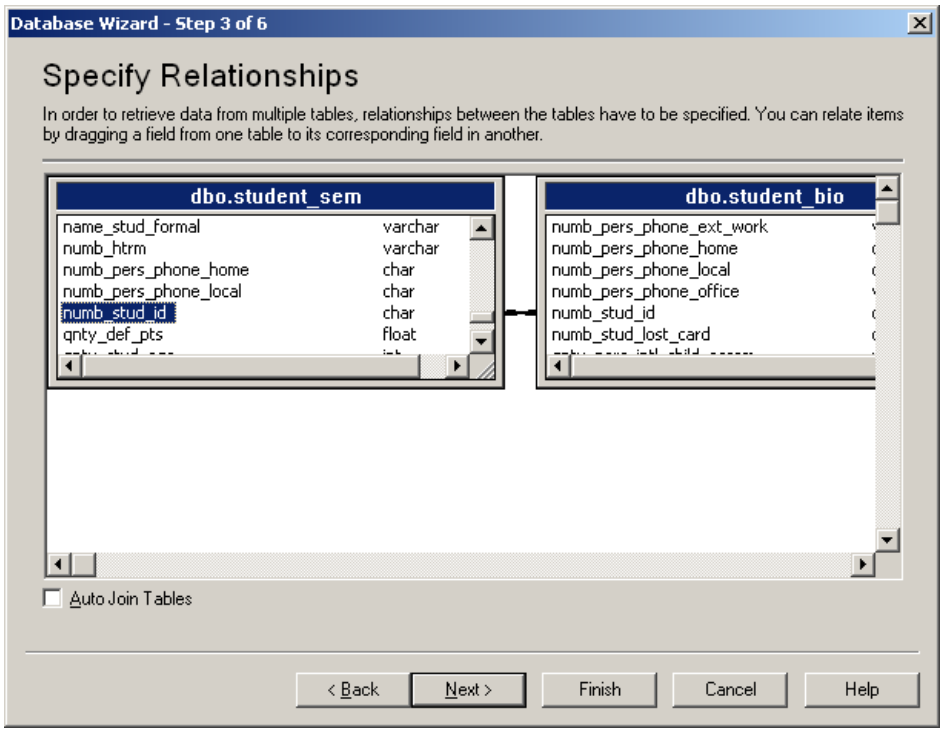
In the window for selecting data, fields may be selected from any table in the database.

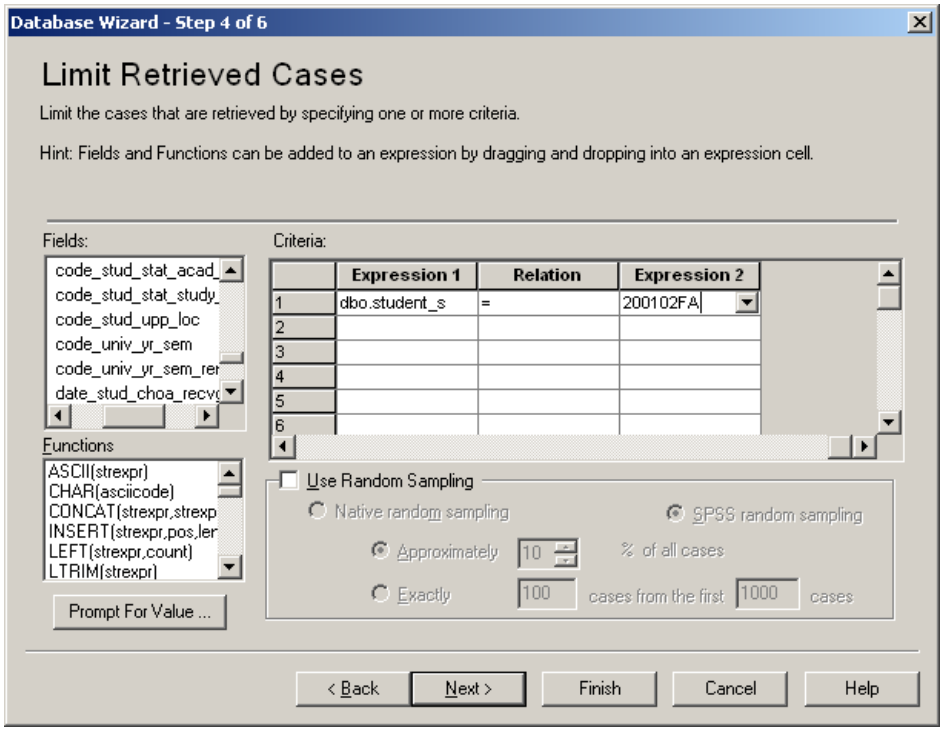


If data elements are selected from multiple tables the window for specifying the relationships between the tables will appear. All common fields between the tables are automatically linked. These relationships may be deleted individually by selecting the link and pressing the delete key. All relationships may be deleted by un-checking the **Auto Join Tables** option in the bottom left corner of the window.

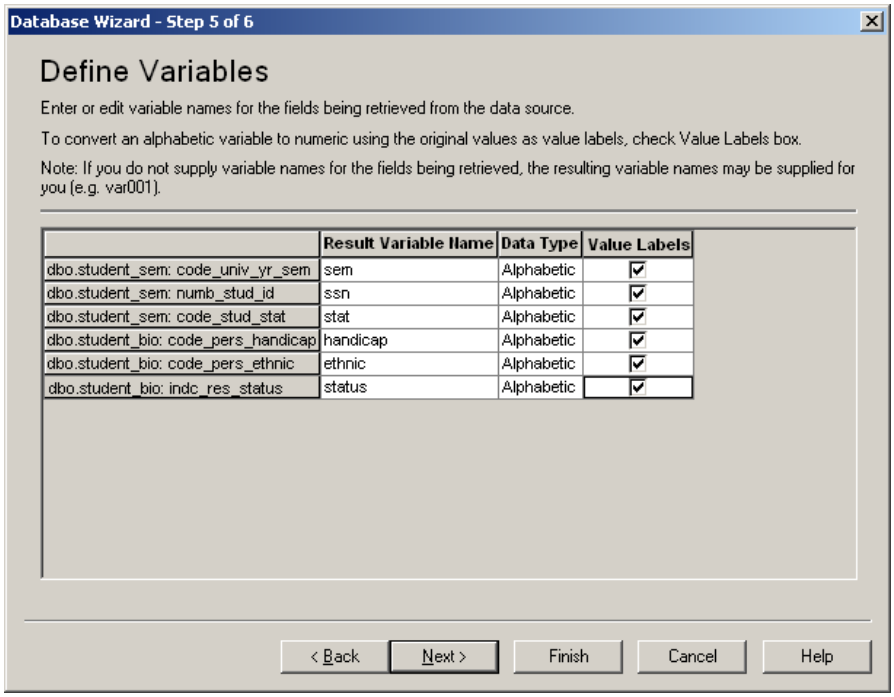


If the Auto Join Tables option is turned off the relationships between tables will need to be specified manually as shown below.

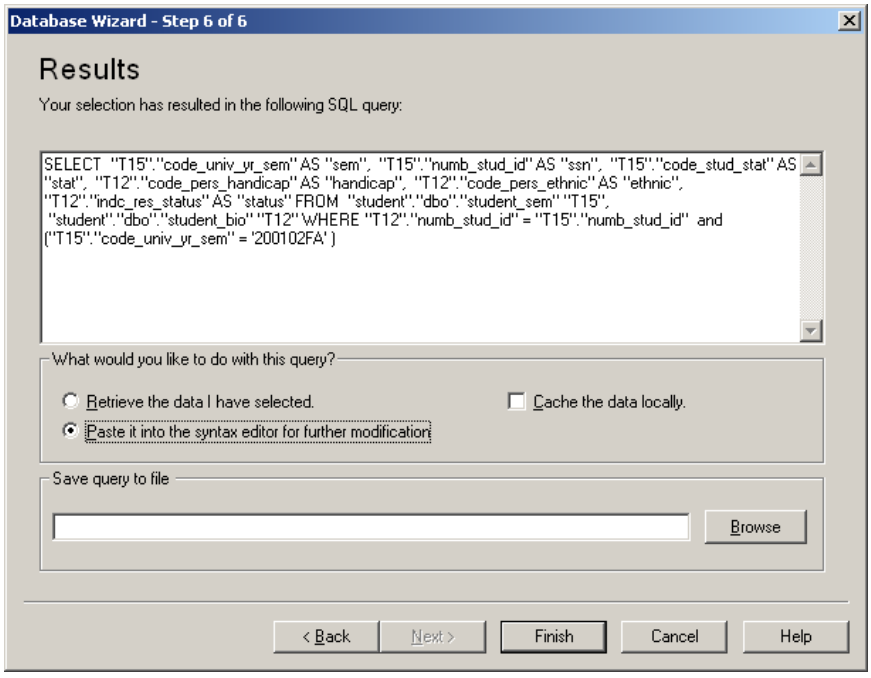




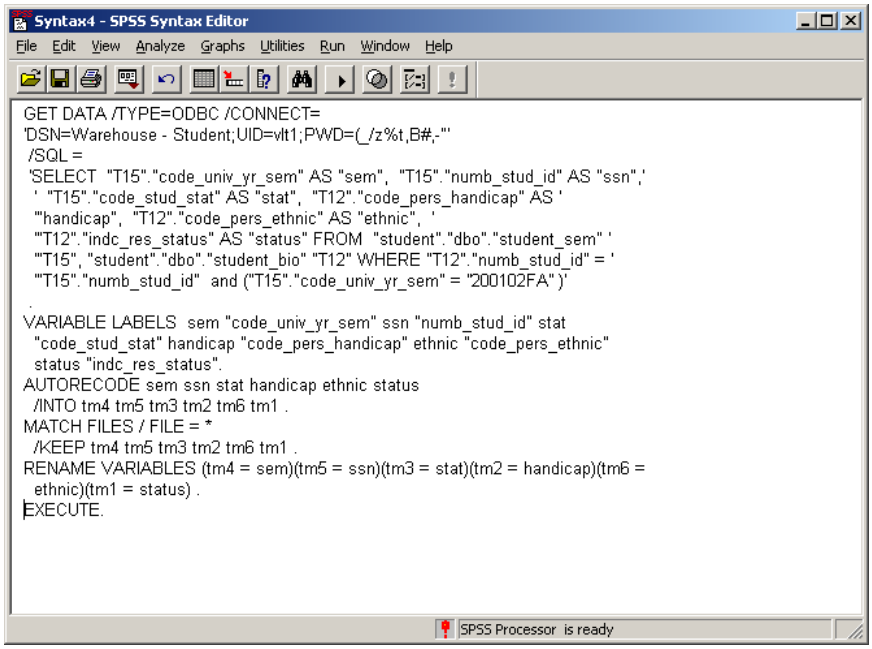
It is important to note that SPSS field names have maximum a length of eight. SPSS will automatically rename the data warehouse fields. By default the first eight characters of the field name from the warehouse are used. When this would give more than one field the same name, the last character is changed to a number. It is often useful to rename the fields as done below. Also note that SPSS field labels are automatically assigned the data warehouse field name.



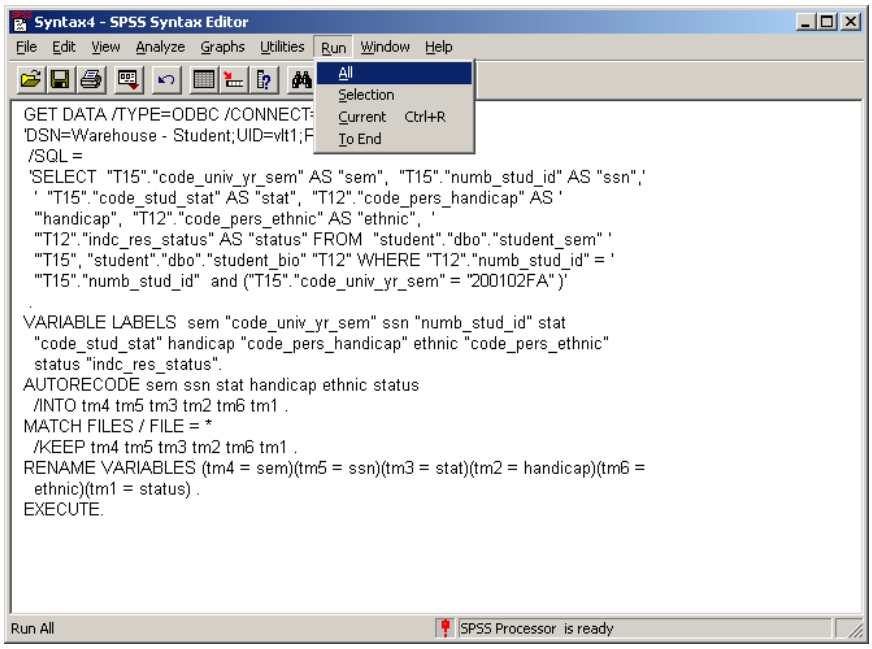
The last step in the wizard gives the option of retrieving the data or pasting the query syntax into the syntax editor. Pasting the query syntax onto the syntax window will allow the query to be edited or saved for future use.



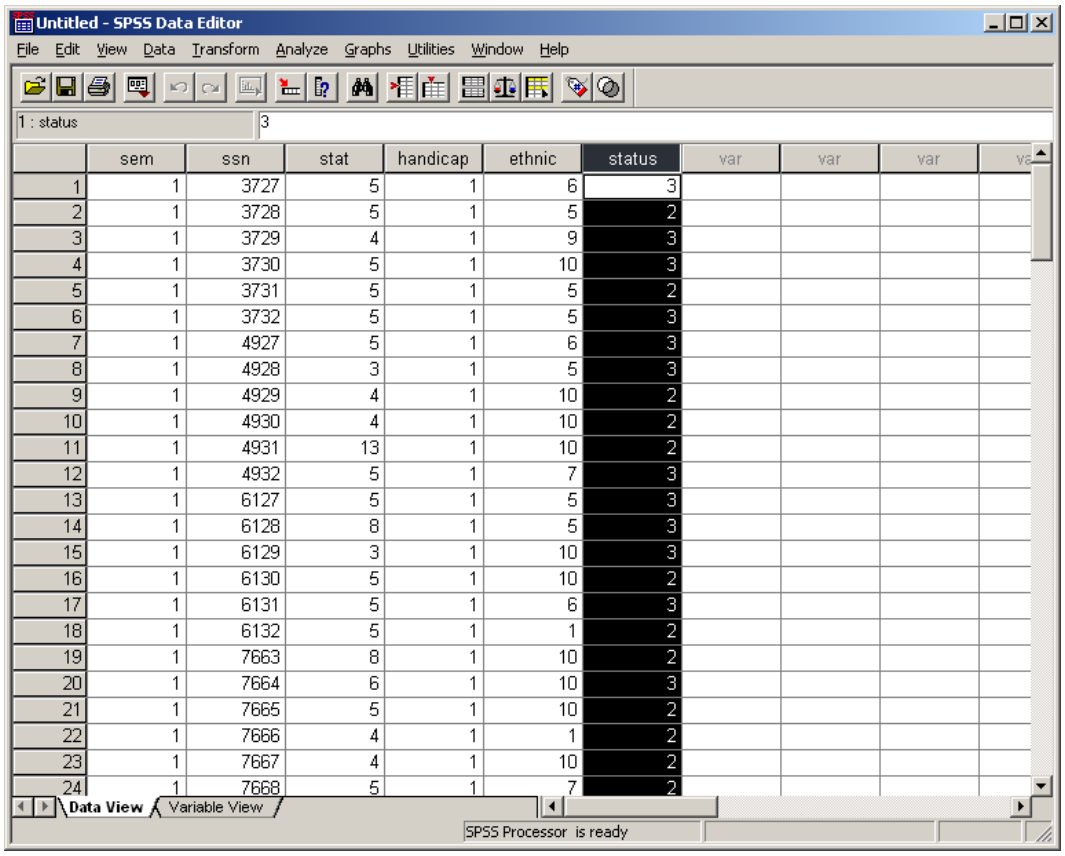
The query syntax in the syntax editor shown below was generated by the Database Wizard. It is possible to write queries in the syntax editor using the syntax below as a guide. Notice that the password has been encrypted. If writing the syntax simply enter the password for the userid being use after **PWD=**. It is important to note that the **SELECT** statement is what gets sent to the data warehouse. All other code before and after the **SELECT** is SPSS code.



To run the query from the Syntax Editor select **Run** and **All** from the menu bar.



When the query is complete the data will be in the Data Editor. From here the data can be edited, saved or used for running reports and analysis.



	sem	ssn	stat	handicap	ethnic	status	var	var	var	va
1	1	3727	5	1	6	3				
2	1	3728	5	1	5	2				
3	1	3729	4	1	9	3				
4	1	3730	5	1	10	3				
5	1	3731	5	1	5	2				
6	1	3732	5	1	5	3				
7	1	4927	5	1	6	3				
8	1	4928	3	1	5	3				
9	1	4929	4	1	10	2				
10	1	4930	4	1	10	2				
11	1	4931	13	1	10	2				
12	1	4932	5	1	7	3				
13	1	6127	5	1	5	3				
14	1	6128	8	1	5	3				
15	1	6129	3	1	10	3				
16	1	6130	5	1	10	2				
17	1	6131	5	1	6	3				
18	1	6132	5	1	1	2				
19	1	7663	8	1	10	2				
20	1	7664	6	1	10	3				
21	1	7665	5	1	10	2				
22	1	7666	4	1	1	2				
23	1	7667	4	1	10	2				
24	1	7668	5	1	7	2				