

Paper 117-29

Introduction to Using SAS® Enterprise Guide® for Statistical Analysis

Charles Hallahan, Economic Research Service, USDA

Linda Atkinson, Economic Research Service, USDA

ABSTRACT

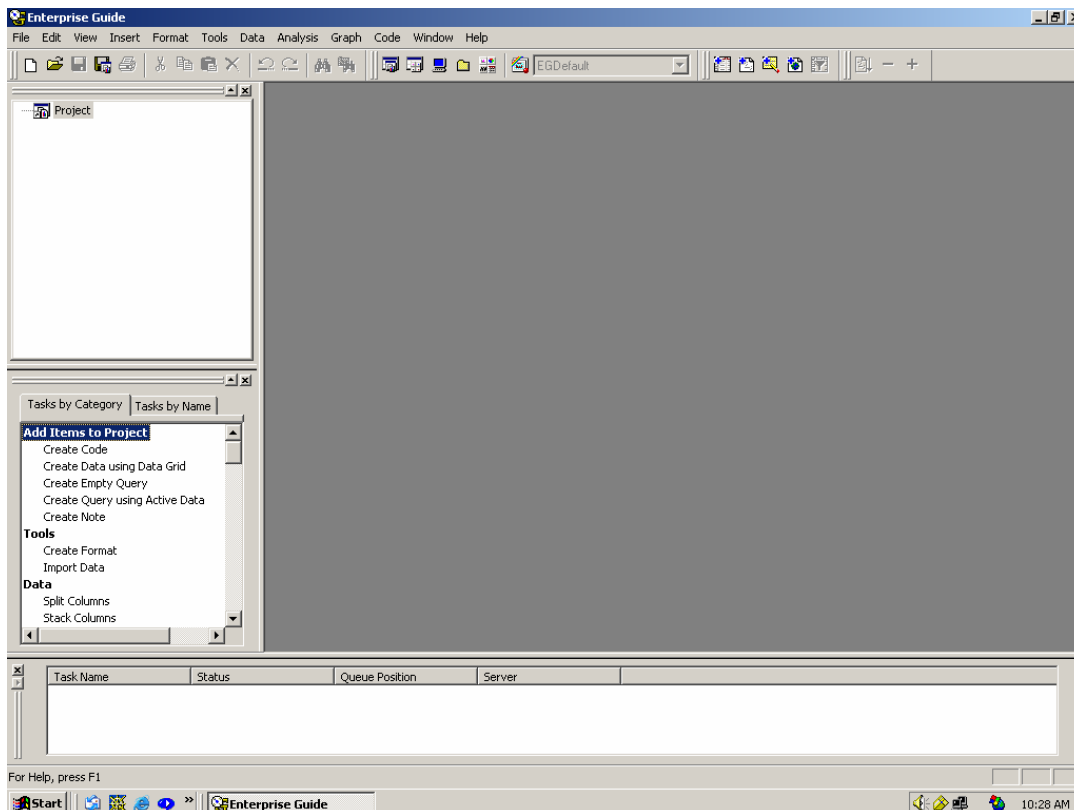
The workshop begins with a quick tour through the layout and purpose of Enterprise Guide (EG) as a menu-based interface to SAS® procedures. Emphasis in this workshop will be on the statistical capabilities of EG. In particular, a set of data will serve as a case study for performing exploratory data analysis, estimating a multiple regression model, and examining graphical diagnostics for the model fit. The workshop will show how code generated by EG can be customized, stored, and rerun, and custom reports saved with the Document Builder.

INTRODUCTION

The SAS System provides a powerful framework for statistical analysis. It has extensive data manipulation capabilities to prepare for analytic and modeling work. It has reporting tools for presenting results. However, for a new user, learning how to write code and run the appropriate procedures can be daunting. Enterprise Guide enables you to get answers without having to write programs, through a point-and-click interface making selections from a series of menus. As a benefit even for experienced SAS programmers, EG provides a framework within which to organize the data, tasks, and results involved in performing a statistical analysis, through the creation and maintenance of “projects”. In this workshop, a set of data will serve as an example for performing exploratory data analysis, estimating a multiple regression model, and examining graphical diagnostics for the model fit. Along the way, we will review the code generated automatically by EG, and demonstrate how it can be customized, stored, and rerun. We’ll also “put it all together” by collecting results and generating custom reports through a tool called the Document Builder.

CREATING A PROJECT

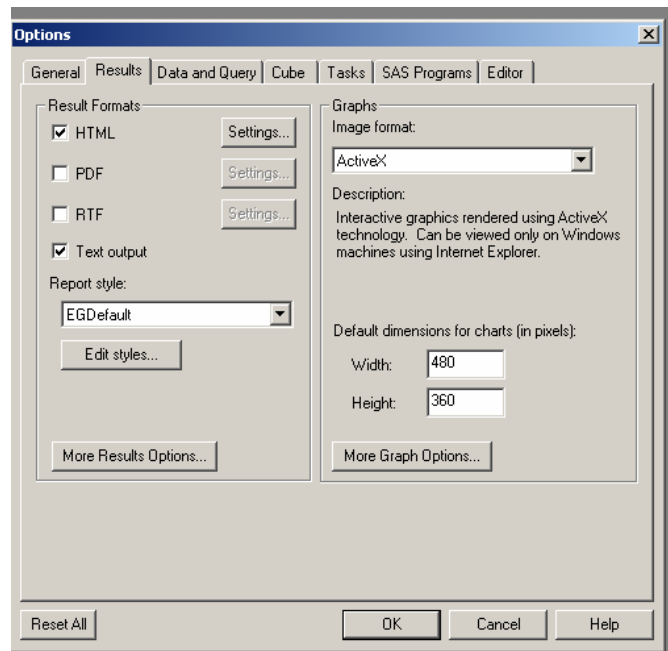
When you first bring up Enterprise Guide, you’ll be asked whether you’d like to create a new project or open an existing one. Click on the “**Create New Project**” button. Alternatively, at any time while in Enterprise Guide, you can click on the **File** → **Open** pulldown menus to indicate you’d like to open a new project. There’s also a **New** button on the toolbar to accomplish the same function.



You'll see a Project entry in the "Project Window" at the upper left of the EG workspace. Other windows open are Task List (also available through pull-down menus), Task Status, and a general workspace area (grey at first but this will fill up quickly). Any of these windows can be closed to give you more space. Note that to close the Status window, you need to click on a box at the upper left rather than upper right as is usual in Windows applications.

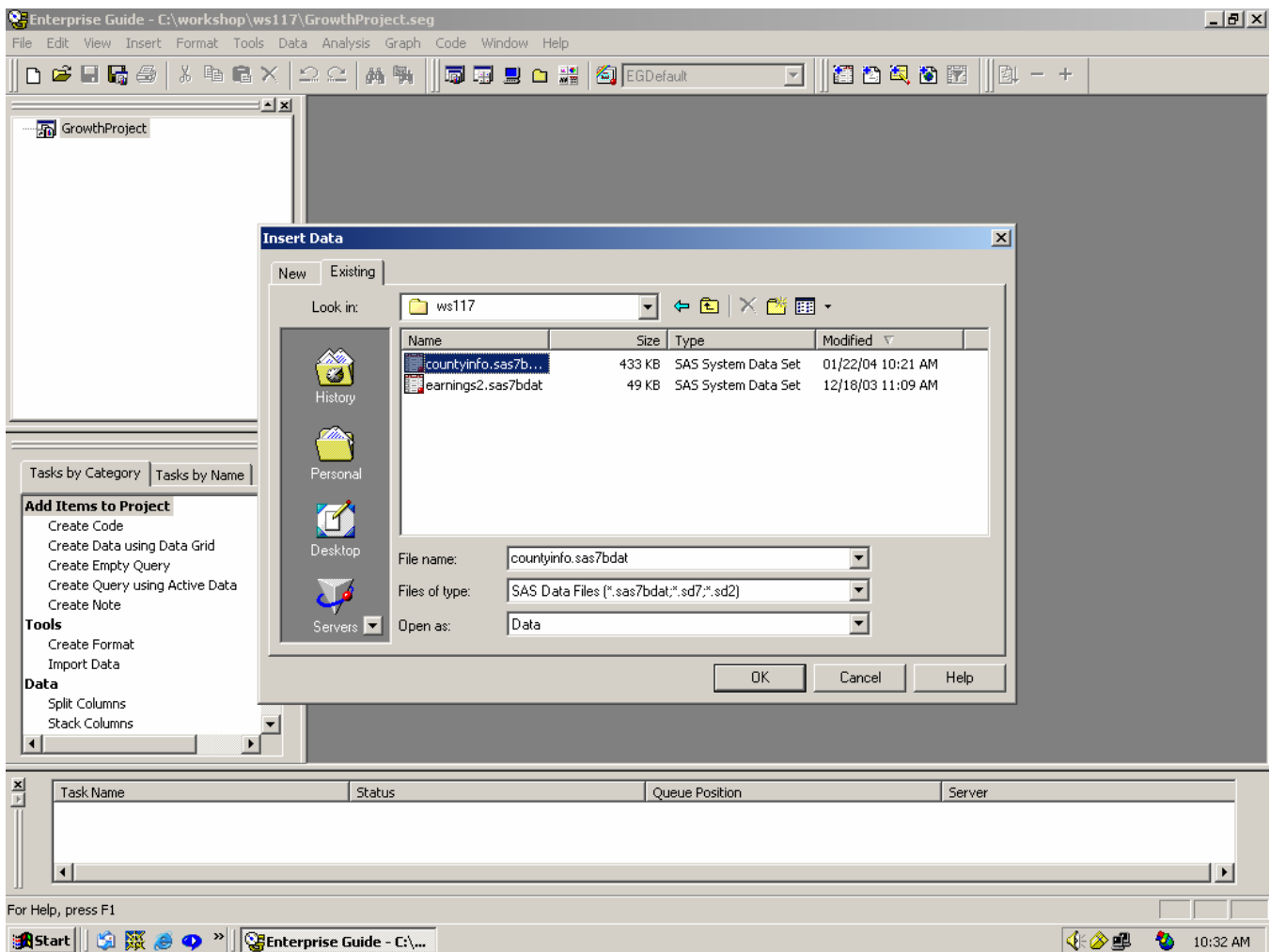
Let's give the Project a better name. This analysis will be about factors associated with local and regional economic growth, so right-click on the word **Project** in the Project window, select "**Save Project As...**" and type "GrowthProject" in the **File name:** box. Enterprise Guide will append the extension .seg.

Besides organizing your work handily, Enterprise Guide also provides lots of opportunity for customizing your work environment. Let's take a look at specifying how we would like our results presented. Click on **Tools → Options** and select the **Results** tab. Be sure that both **HTML** and **Text output** are checked, then click on the **OK** button (you can click on the other tabs to see what other types of things EG gives you control over).



ACCESSING DATA

The first step in an analysis is to get some data to work with. Click on **Insert → Data...** You have the option of selecting **New** to type in data interactively through a spreadsheet-type interface, or **Existing** to access data that are already stored in a file. Use the **Existing** tab and select **countyinfo.sas7bdat** (initially, EG is looking for SAS Data Files, but this can be changed using the dropdown menu by **Files of type:**). Click on the OK button.



Enterprise Guide reads the data and brings it up in a Data Viewer. You can scroll to the right to see more variables or down to see more observations. If you need to modify a data value, EG will ask you if you'd like to switch to Update mode. Note the entry for **countyinfo** in the Project window. An icon indicates that this object is a data table. We can close the data viewer by clicking on the X in the upper right (below the similar box for EG itself). At any time when you'd like to view the data again, double-click on the data table icon.

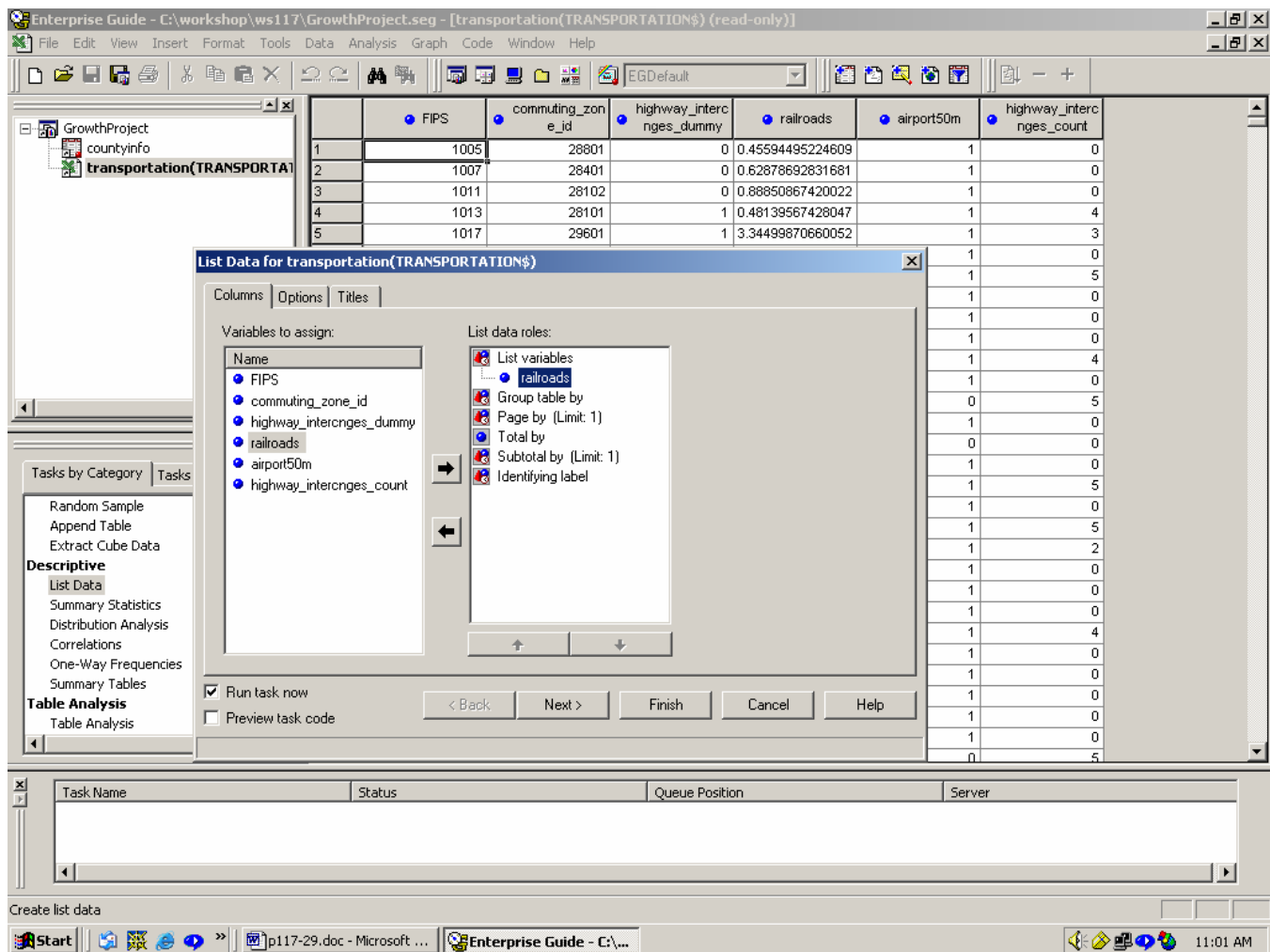
The screenshot displays the SAS Enterprise Guide software interface. The main window shows a project named 'GrowthProject' with a data table named 'countyinfo'. The table has the following columns: FIPSCD, local_tax_level, climate_quality, census_division_5, census_division_1, census_division_2, and census_division_3. The data is organized into rows, with some rows highlighted. On the left side, there is a 'Tasks by Category' pane with options like 'Add Items to Project', 'Create Code', 'Create Data using Data Grid', 'Create Empty Query', 'Create Query using Active Data', 'Create Note', 'Tools', 'Create Format', 'Import Data', 'Data', 'Split Columns', and 'Stack Columns'. At the bottom, there is a 'Task Name' pane with columns for 'Task Name', 'Status', 'Queue Position', and 'Server'. The status bar at the bottom indicates 'For Help, press F1' and shows the system clock as 11:11 AM.

	FIPSCD	local_tax_level	climate_quality	census_division_5	census_division_1	census_division_2	census_division_3
1	1005	4.381111111	0.4315369723	0	0	0	0
2	1007	3.697777778	-0.647139903	0	0	0	0
3	1011	3.39	0.5243712118	0	0	0	0
4	1013	2.728888889	0.5528554209	0	0	0	0
5	1017	2.391111111	0.1853676179	0	0	0	0
6	1019	5.037777778	-0.554645591	0	0	0	0
7	1021	3.355555556	-0.465807628	0	0	0	0
8	1023	7.206666667	0.3295777352	0	0	0	0
9	1025	3.402222222	0.3295777352	0	0	0	0
10	1027	4.131111111	0.1853676179	0	0	0	0
11	1029	5.042222222	0.1853676179	0	0	0	0
12	1031	3.26	0.8530325735	0	0	0	0
13	1035	1.876666667	0.4984341346	0	0	0	0
14	1037	2.046666667	0.1853676179	0	0	0	0
15	1039	3.314444444	0.5528554209	0	0	0	0
16	1041	2.275555556	0.5690956634	0	0	0	0
17	1043	4.503333333	-0.633481251	0	0	0	0
18	1047	3.863333333	0.3405653648	0	0	0	0
19	1049	3.833333333	-0.449675404	0	0	0	0
20	1053	4.733333333	0.9068084624	0	0	0	0
21	1057	4.521111111	-0.51022661	0	0	0	0
22	1059	2.435555556	-0.981464624	0	0	0	0
23	1061	3.855555556	0.5735054273	0	0	0	0
24	1063	6.178888889	0.2148371522	0	0	0	0
25	1065	3.611111111	0.2148371522	0	0	0	0
26	1067	3.375555556	0.6548146581	0	0	0	0
27	1071	7.023333333	-0.680373804	0	0	0	0
28	1075	3.956666667	-0.51022661	0	0	0	0
29	1079	3.494444444	-0.981464624	0	0	0	0

The data for the workshop comes from a USDA/ERS study by Kusmin et al (1996) which models growth in total earnings for nonmetro U.S. counties from 1979 to 1989. Explanatory variables include demographic, labor market, education, transportation, etc. characteristics. The workshop will only use a few of the variables from the original study for illustrative purposes.

Some of the information we need for this analysis (specifically, the transportation variables) is stored in a spreadsheet file. Click on **Insert**→**Data** again, use the **Existing** tab, and for **Files of type**: select **Microsoft Excel Files (*.xls)**, then click on **transportation.xls** and OK. Select the transportation\$ spreadsheet from within the workbook (which represents the entire sheet, whereas the file without the \$ at the end represents a range of cells within the sheet). Click on **OK**. The spreadsheet is added to the project (with a different icon indicating that it is an Excel file), and the data are brought up in a viewer again as the SAS data set was.

As you scroll through the data, you'll notice that some cells are blank (for example, lines 349-351 are blank for several variables). To check that these variables were interpreted as numeric rather than character, right click on the icon for the transportation\$ spreadsheet, select **Properties** and go to the **Columns** tab. All variables were brought in as numeric. In the Tasks window, double-click on **List data** (under Descriptive). In the window that comes up, click on the variable railroads, drag and drop it onto **List variables** (holding down the left mouse button), and click on **Finish**.

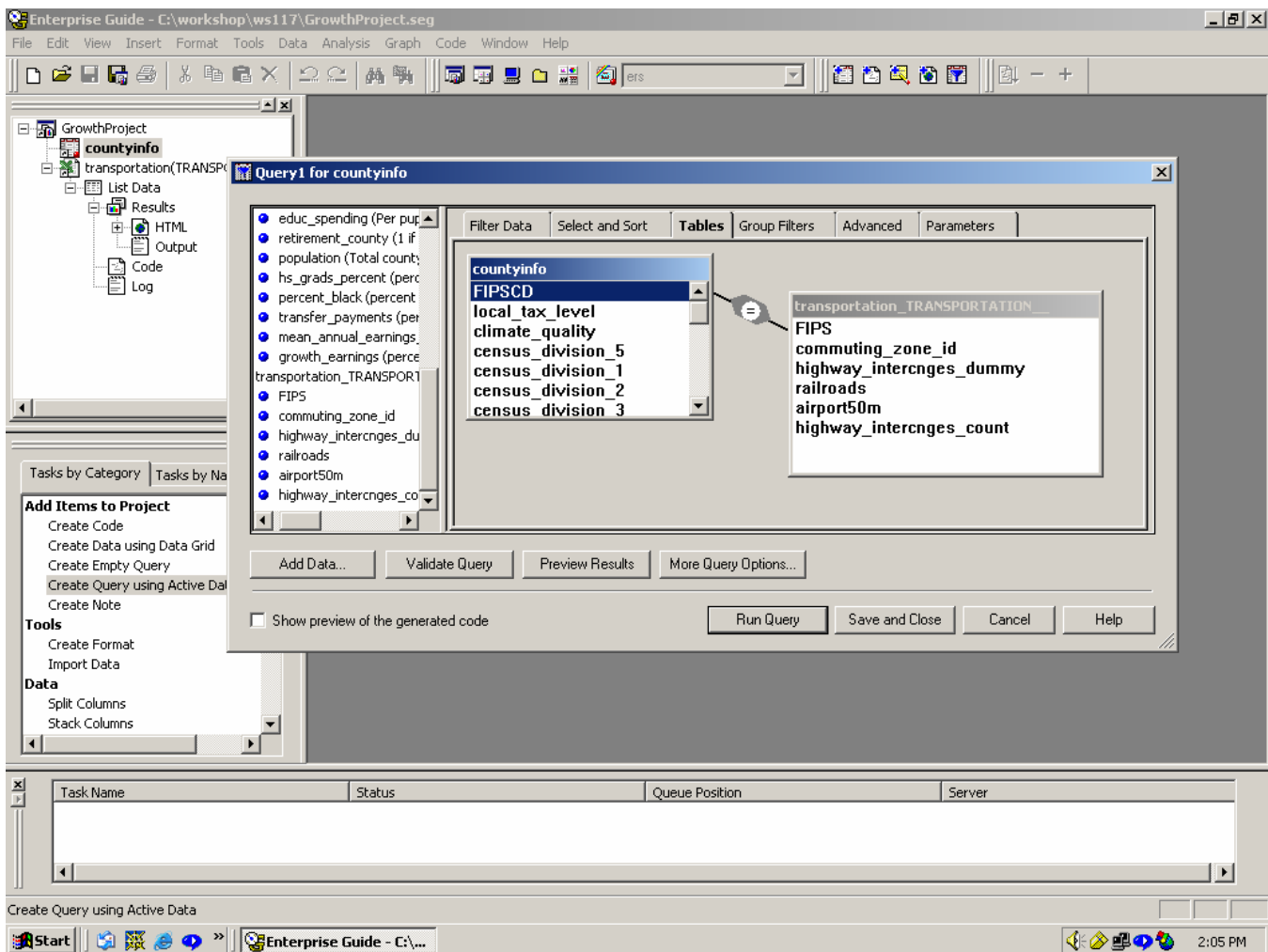


An HTML report lists the observations for the requested variable (other formats such as pdf and rtf can be produced). Scroll down to Obs 349 to see that the observation has the correct missing value indicator.

You've now run a simple task in Enterprise Guide. Note that the **List Data** task has been added to the project window, under the transportation data to which it applies. Different icons indicate the **Results**, **Code**, and **Log** that were generated. Besides the HTML output being displayed, there is also the regular SAS **Output** window for the listing. Double-click on it to view it, then close it. Double-click on the Code to view the Proc Print program that Enterprise Guide generated. This could be directly edited and re-run if you wanted to add options or make changes not provided in the menu systems. Similarly, double-click on the Log to review SAS messages and notes. If errors had been generated, the log icon in the project window would have a red X over it alerting you to the problem. Close any open windows in the workspace area to clean it up.

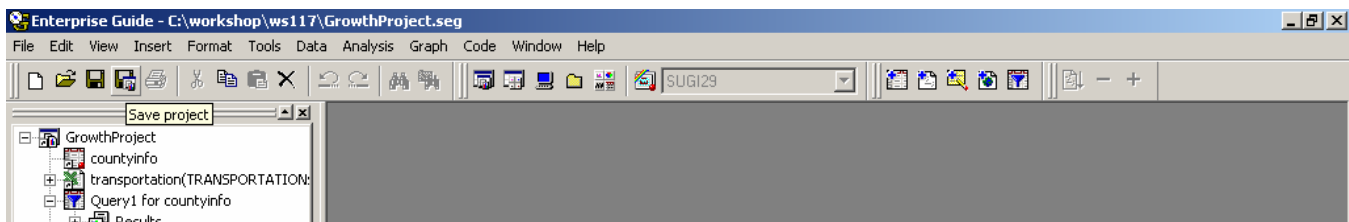
We currently have two sources of data in the project: a SAS data set with some variables, and an Excel spreadsheet with transportation variables. For subsequent analysis and modeling, we'll want to put them together. Click once on the icon for the **countyinfo** data set in the project window to highlight it and make it the active data source. In the **Task** window, double-click on **Create Query using Active Data** (under Add Items to Project). Click on the **Add Data...** button. Select the **transportation** spreadsheet and click on OK. We get a message that we'll need to join the tables manually. This is because we don't have a variable name in common to the data sources to match on. We want to join counties, but in the earnings data set the identifying variable is called FIPSCD, whereas in the spreadsheet it's called FIPS. We'll have to tell Enterprise Guide how we want this done.

Click on **OK** in the message box. You'll see boxes for the two data sources. These can be expanded to show more of their variable lists. Click on FIPSCD in the earnings data set, hold down the left mouse button and drag the cursor until FIPS in the transportation spreadsheet is also highlighted and a dotted line connects the two. Release the mouse button. A symbol indicates how the two sets of data will be joined. Click on the **Run Query** button.



The results of the query, a combined data table, are displayed in the data viewer. Scroll to the right to see that both sets of variables are included. An icon in the Project window shows **Results of Query1 for countyinfo**. We'll be using this as the source for our analyses, so let's give it a better name. Right-click on the Results data table icon and select **Rename**. Type **earnings** and hit **Enter**.

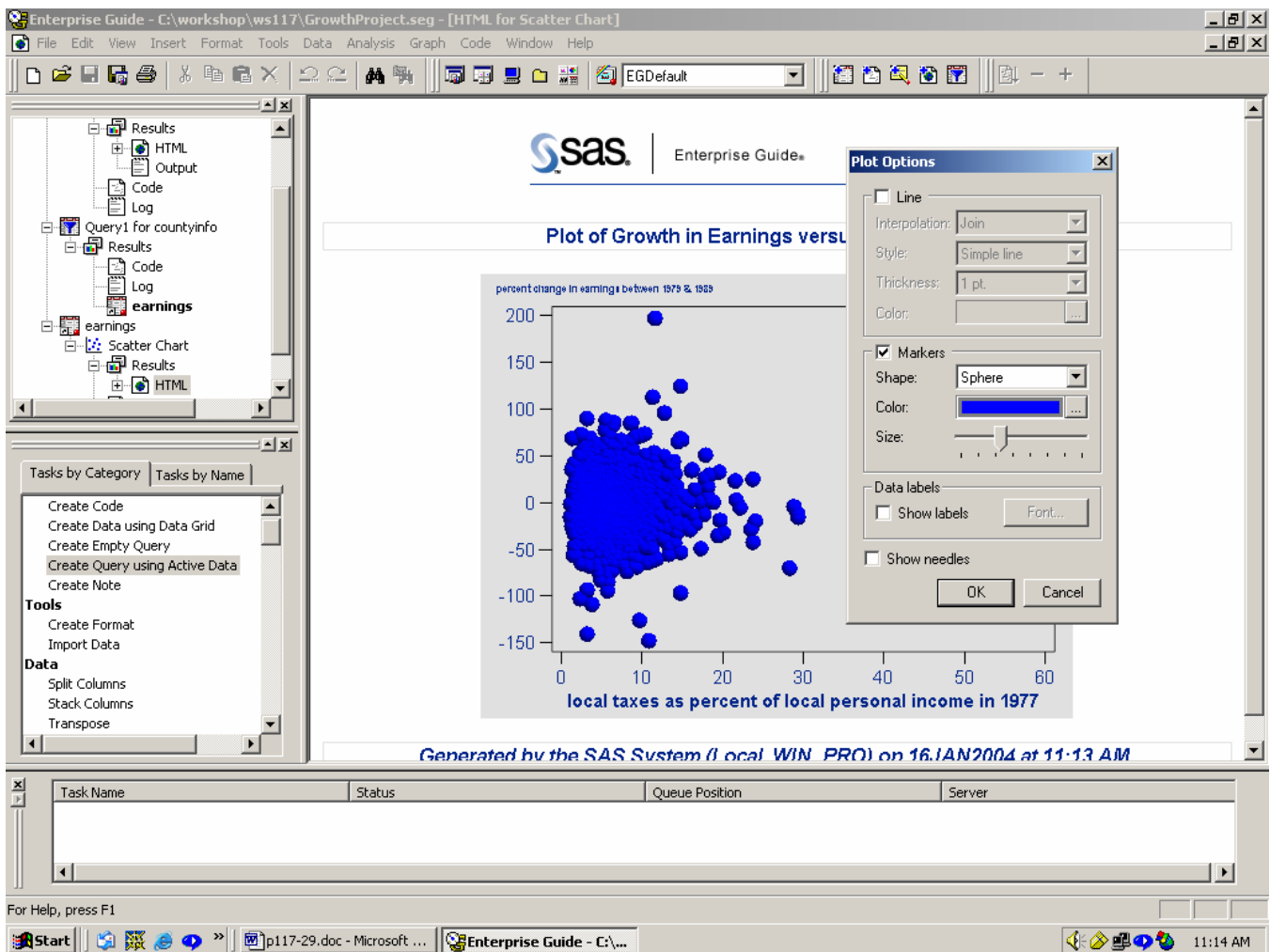
You'll want to periodically click on the **Save project** button on the toolbar. Enterprise Guide will save all of the links listed in your project window, so that the next time you bring it up and access the project, everything is ready for you to pick up from there.



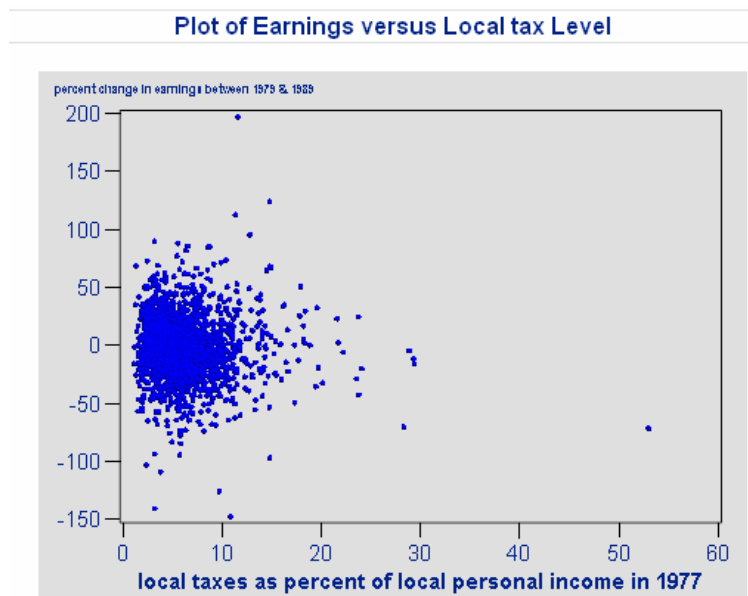
We're now ready to proceed to data analysis.

EXPLORING THE DATA

Any analysis should begin with exploring the data. For example, we could look at some plots, say, the dependent variable, **growth_earnings**, versus a candidate explanatory variable, say, **local_tax_level**. Click on **Graph** → **Scatter** and then on the **Columns** tab. Drag the variable **local_tax_level** to the **Horizontal location** and **growth_earnings** to the **Vertical location**. Click **Next >** or the **Appearance** tab, then on the **Titles** button and enter "Plot of Growth in Earnings versus Local tax Level". Click **Finish**. The following graph appears.



Since the plot symbols seem a little large, right-click within the axes of the graph, select **Plot Options...** and move the slider for **Size** to the left. Click on **OK**.



Two points stand out in the plot - an observation with a growth in earnings just under 200% and an observation with a local tax level over 50%. Since the Chart Tips option is turned on (as can be seen by right-clicking, selecting **Options** and then **Graph...**), moving the mouse over any plot point shows the exact values.

We can define a Query to learn more about these observations. Click on the icon for earnings to make it the active table and then click on **Tools → Query → Create from Active Data ...**. Check that the first tab, **Filter Data**, is selected and drag the column `local_tax_level` from the left pane into the Query workspace. Change the **Operator** to ">, greater than" and enter "50" in the **Value** area. Check **OK** and then **Run Query**. The results of the query show a single observation with a FIPCD value of 48301 and a `local_tax_level` of 53%. However, most of the other variables are missing for this observation, so the observation will not end up being used in the regression analysis. The icon for this latest Query can now be deleted.

The next step would be to look at some simple correlations between the dependent variable and several possible explanatory variables. Checking that earnings is still the Active Table, click on **Analysis → Descriptive → Correlations** and select the five variables `local_tax_level`, `climate_quality`, `coal_mines`, `hs_grads_percent`, and `mean_annual_earnings_log` and move to the **Correlations variables** slot. Then select `growth_earnings` as the **Correlate with** variable. Click **Next >** to move to the **Correlations** tab and keep the default of Pearson correlations. Click **Next >** to move to the **Results** tab and check the box for **Scatter plots** and the box to **Rank correlations from highest to lowest**. Click **Next >** to move to the **Title** tab and leave as is. Click on **Finish**.

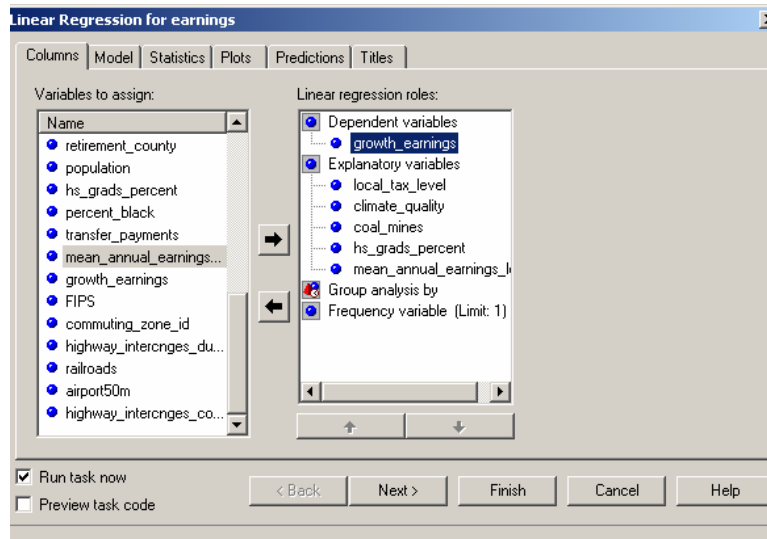
Along with summary statistics, the correlations of the 5 variables with `growth_earnings` are given in descending absolute value.

<code>climate_quality</code>	2229	0.04156	1.90006	92.63412	-3.76902	7.96497	reflecting temperature & humidity in January & July
<code>coal_mines</code>	2220	0.80362	4.24193	1784	0	63.76431	percent employed in coal industry in 1979
<code>hs_grads_percent</code>	2220	0.47131	0.09858	1046	0.18427	0.69917	percent of HS grads in 1980
<code>mean_annual_earnings_log</code>	2220	9.10719	0.19190	20218	8.51551	9.86764	log of mean of earnings/job for 1976-78
Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations							
<code>growth_earnings</code> percent change in earnings between 1979 & 1980	<code>mean_annual_earnings_log</code>	<code>coal_mines</code>	<code>climate_quality</code>	<code>hs_grads_percent</code>	<code>local_tax_level</code>		
	-0.21463	-0.16773	0.14138	-0.10219	-		
	<.0001	<.0001	<.0001	<.0001			
	2220	2220	2229	2220			

The dataset has 2232 observations. Note that some of the variables have missing observations. None of the correlations are dramatically high, c'est la vie. The next step is to fit a regression model.

FITTING A MODEL

Again with earnings as the highlighted data table in the Project window, click on **Analysis → Regression → Linear** and select the dependent and explanatory variables as below.



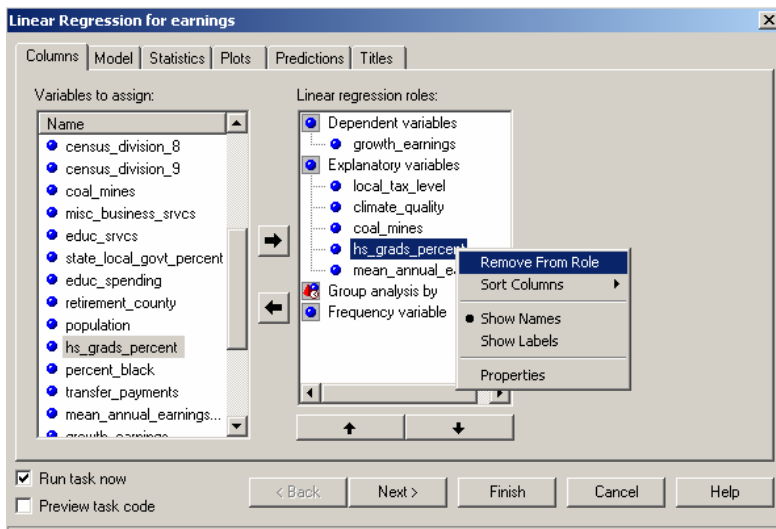
Click on **Finish** to obtain the regression results.

Linear Regression					
Results					
The REG Procedure					
Model: Linear_Regression_Model					
Dependent Variable: growth_earnings percent change in earnings between 1979 & 1989					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	123857	24771	40.93	<.0001
Error	2214	1339970	605.22585		
Corrected Total	2219	1463827			
Root MSE					
Root MSE	24.60134	R-Square	0.0846		
Dependent Mean	-1.47730	Adj R-Sq	0.0825		
Coeff Var	-1665.28589				

While the F-value shows that the overall regression is significant, the R-square of 0.08 is rather low. The next step is to check the significance of individual variables.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	214.33262	28.08188	7.63	<.0001
local_tax_level	local taxes as percent of local personal income in 1977	1	-0.69322	0.18572	-3.73	0.0002
climate_quality	index reflecting temperature & humidity in January & July	1	2.08760	0.28482	7.33	<.0001
coal_mines	percent employed in coal industry in 1979	1	-0.69465	0.13494	-5.15	<.0001
hs_grads_percent	percent of HS grads in 1980	1	-1.85154	6.54638	-0.28	0.7773
mean_annual_earnings_log	log of mean of earnings/job for 1976-78	1	-23.09953	3.21737	-7.18	<.0001

All variables are highly significant, except for `hs_grads_percent`. Suppose we want to drop this variable and replace it in the model with the variable `educ_srvcs`. Double-click on the icon Linear Regression in the Project window. Make sure the **Columns** tab is selected, then right-click on the variable `hs_grads_percent` and select **Remove From Role**.



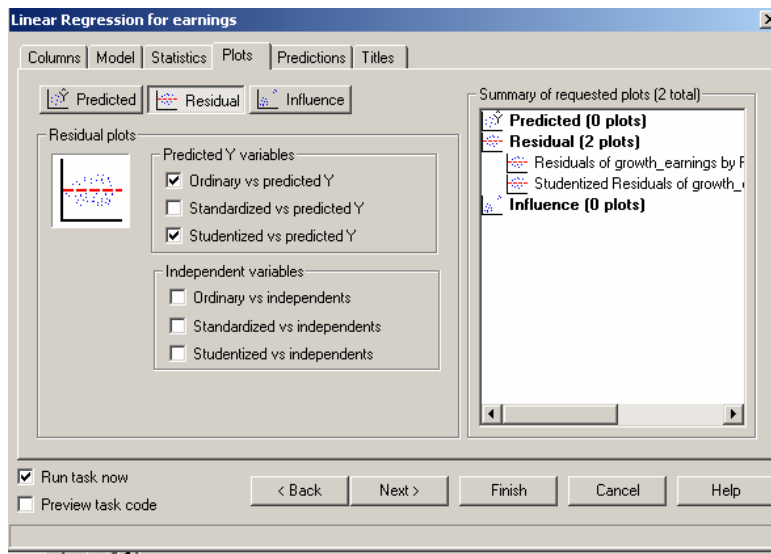
Now right-click on the variable `educ_srvcs` and select **Assign To Role → Explanatory Variable**. Click **Finish** and **Yes** from now on when prompted with "Would you like to replace the results from the previous run?". The R-square shows a slight improvement to 0.09, but now all included variables in the model are significant.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	219.23884	25.94722	8.45	<.0001
local_tax_level	local taxes as percent of local personal income in 1977	1	-0.68628	0.16968	-4.04	<.0001
climate_quality	index reflecting temperature & humidity in January & July	1	2.24345	0.28113	7.98	<.0001
coal_mines	percent employed in coal industry in 1979	1	-0.67790	0.13008	-5.21	<.0001
mean_annual_earnings_log	log of mean of earnings/job for 1976-78	1	-23.85072	2.86405	-8.33	<.0001
educ_srvcs	percent employed in education in 1979	1	1.76715	0.37488	4.71	<.0001

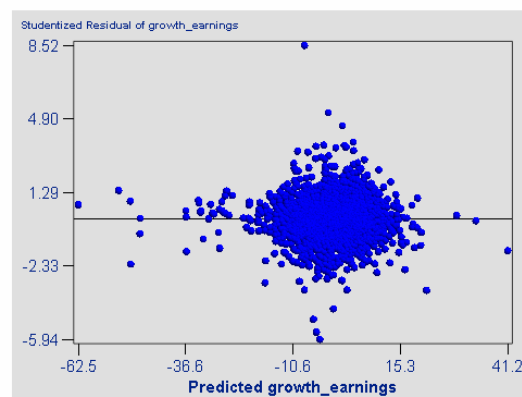
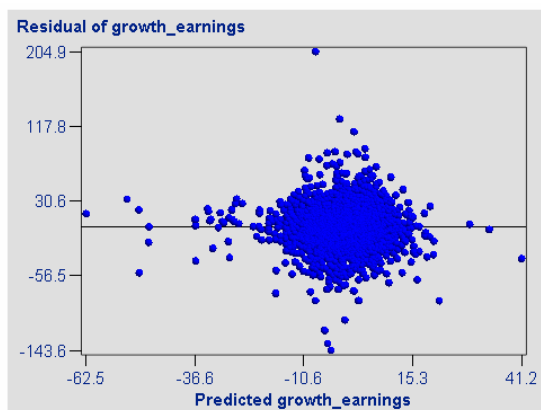
Note also that the parameter estimates for the other variables are essentially the same as the first model. Also, all variables have the correct expected sign. We'll accept this model as our "final" model for the time being and look at some regression diagnostics.

MODEL DIAGNOSTICS

Several kinds of regression diagnostics are available: collinearity, influential observations, and residual plots. We'll start with looking at residual plots. Double-click on the icon Linear Regression and select the tab **Plots** and highlight the middle selection of **Residual**. Select plots of **Ordinary vs predicted Y** and **Studentized vs predicted Y**. Studentized residuals are the ordinary residuals divided by their standard errors where all calculations are done after dropping each observation one at a time. I.e., for the first Studentized residual, the first observation is dropped and parameter estimates obtained. A residual is then obtained for the first observation using these parameter estimates. An observation with an absolute Studentized value greater than 2 or 3 is considered an outlier.

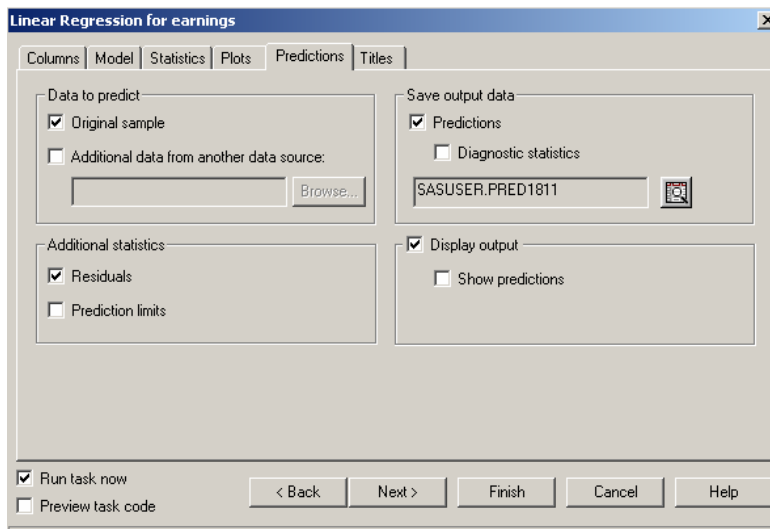


Clicking **Finish** results in the following graphs with several possible outliers.



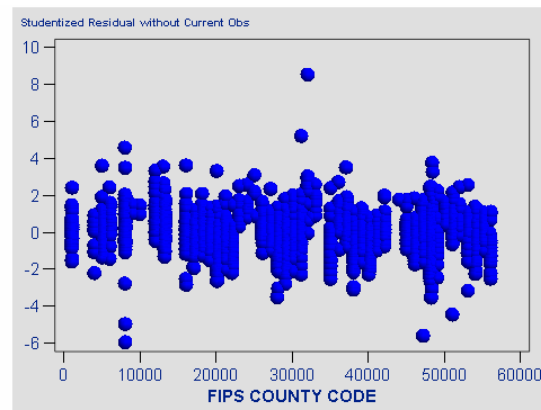
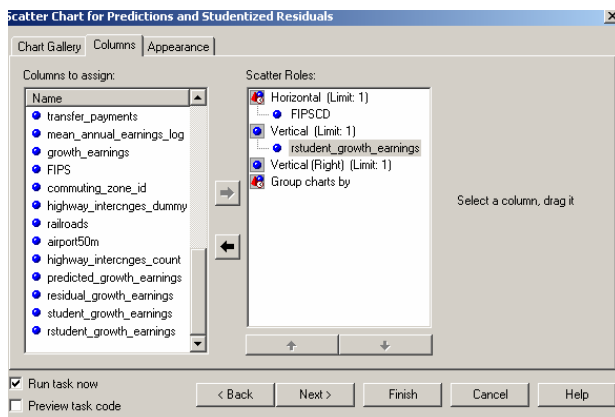
These plots show no obvious violation of the regression model assumptions. Outlying observations can be examined individually for possible explanations. A robust regression technique could be applied (see the new PROC ROBUSTREG) to downweight outlying observations. Suppose we wanted to identify the observation with a Studentized residual of 8.52. This requires a plot not available directly through the menu selections. The idea is to create a scatterplot with the Studentized residuals on the vertical axis and an ID variable, FIPSCD in this case, on the horizontal axis. Using the **Chart Tips** feature will then reveal the value of FIPSCD for that observation.

The first step is to create a dataset containing the Studentized residuals. This is done by re-estimating the model (double-click the icon for Linear Regression) and clicking on the **Predictions** tab. Click on **Original sample** and select **Residuals** from **Additional statistics**. Also click on the **Prediction** box under **Save output data** (see below).



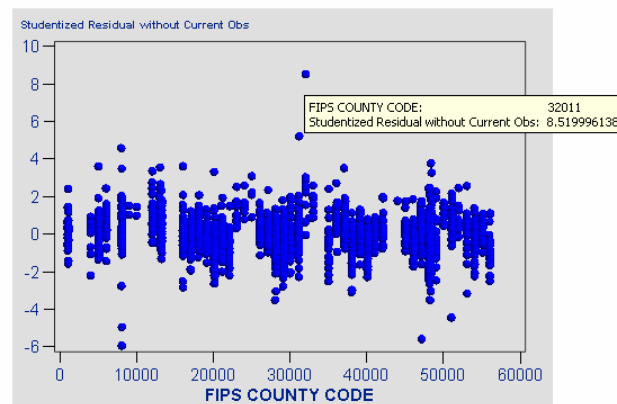
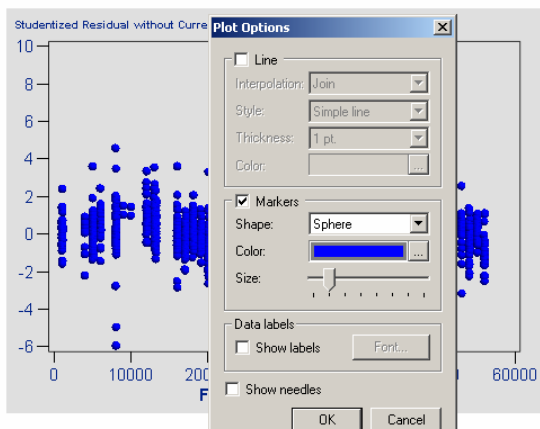
By default, a permanent SAS dataset is saved in the library SASUSER with a name determined by SAS. Click **Finish** and note the new icon labeled "Predictions for growth_earnings = local_tax_level ..." for the new output dataset. This long label can be simplified by right clicking on the icon, select **Rename**, and edit label to just "Predictions and Studentized Residuals".

To create the scatterplot, click on **Graph → Bar...** and scroll to find **2D Scatterplot**. Double-click on the 2D Scatterplot icon and drag FIPSCD to the Horizontal position and rstudent_growth_earnings to the Vertical position. Click **Finish**.



The default graph has rather large markers. To change the appearance of the default graph, right click on the graph and select **Plot Options...** and move the slider to the left to decrease the size of the markers.

Then move the mouse pointer over the highest point to identify the outlier as the observation with a FIPSCD value of 32011.



Defining a query to examine this observation reveals that the unusually large residual is due to the unusually large value of the dependent variable, namely, a growth_earnings value of 197 percent.

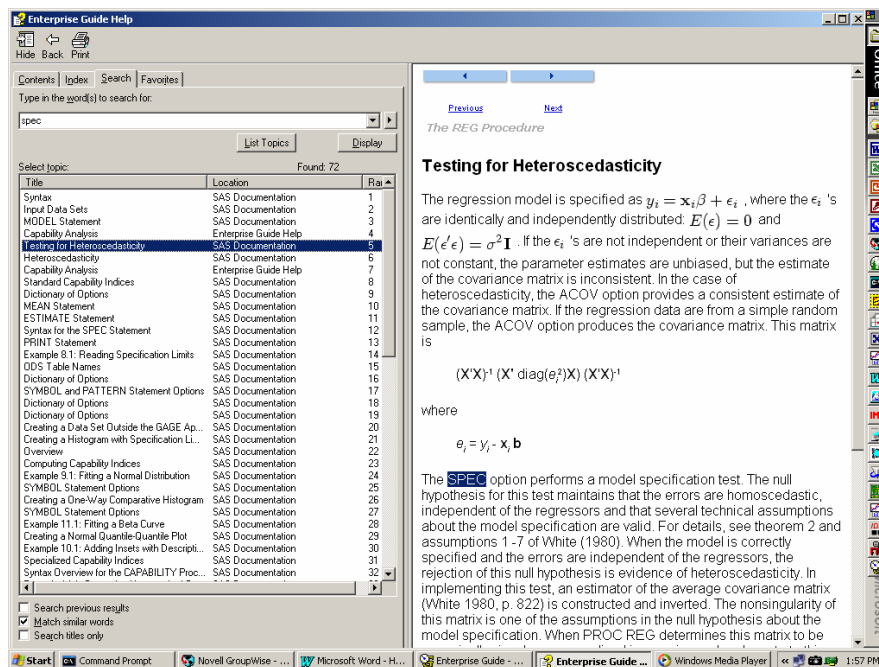
To perform tests for heteroskedasticity (non-constant error variance) and collinearity, we return to the **Linear Regression** task under **earnings** and double-click on its icon. To avoid redoing the plots, click on the **Plots** tab and uncheck any selected plots. Then go to the **Statistics** tab and select the **Variance inflation values** and **Heteroskedasticity test** boxes and click **Finish**.

Collinearity, near linear relationships among the explanatory variables, leads to inflated variance estimates. The Variance Inflation Factors, VIFs, measure the degree of variance inflation associated with each explanatory variable. Scroll through the output to view the VIFs (at the right on the Parameter Estimates table). The VIF for the i^{th} variable is defined as $VIF_i = 1/(1-R_i^2)$ where R_i^2 is the R-square obtained from regressing the i^{th} explanatory variable on the other explanatory variables. While there is no particular cutoff point for a large VIF, all reported values in this case are close to 1, and so collinearity is not a problem here.

When the assumption of homoskedasticity, constant error variance, is violated, the OLS parameter estimates remain unbiased and consistent, but the standard error estimates are now biased. To see exactly what option for PROC REG is being used for the heteroskedasticity test, we can check the code generated by EG. Double-click on the **Code** icon under the **Linear Regression** task. The relevant part of this code is shown below.

	FOOTNOTE "Generated by the SAS System (&_EGSERVERNAME, &SYSSCPL) on %SYSFUNC(DATE	
	TITLE1 "Linear Regression";	
Var	TITLE2 "Results";	ice
	PROC REG DATA=SASUSER.QURY6216;	on
Inte	/* Regression model */	0
	Linear_Regression_Model: MODEL growth_earnings = local_tax_level climate_qual:	
loc	/	.45
	/* Other diagnostic statistics */	
	VIF SPEC	
	;	
clin		29
	/* output dataset to save predictions and other statistics */	
coa	OUTPUT OUT=SASUSER.PRED3401 (LABEL="Linear regression predictions and statist	50
	PREDICTED=predicted_growth_earnings	
	RESIDUAL=residual_growth_earnings	
me	STUDENT=student_growth_earnings	57
	RSTUDENT=rstudent_growth_earnings	
	;	
edu	RUN;	93
	QUIT;	

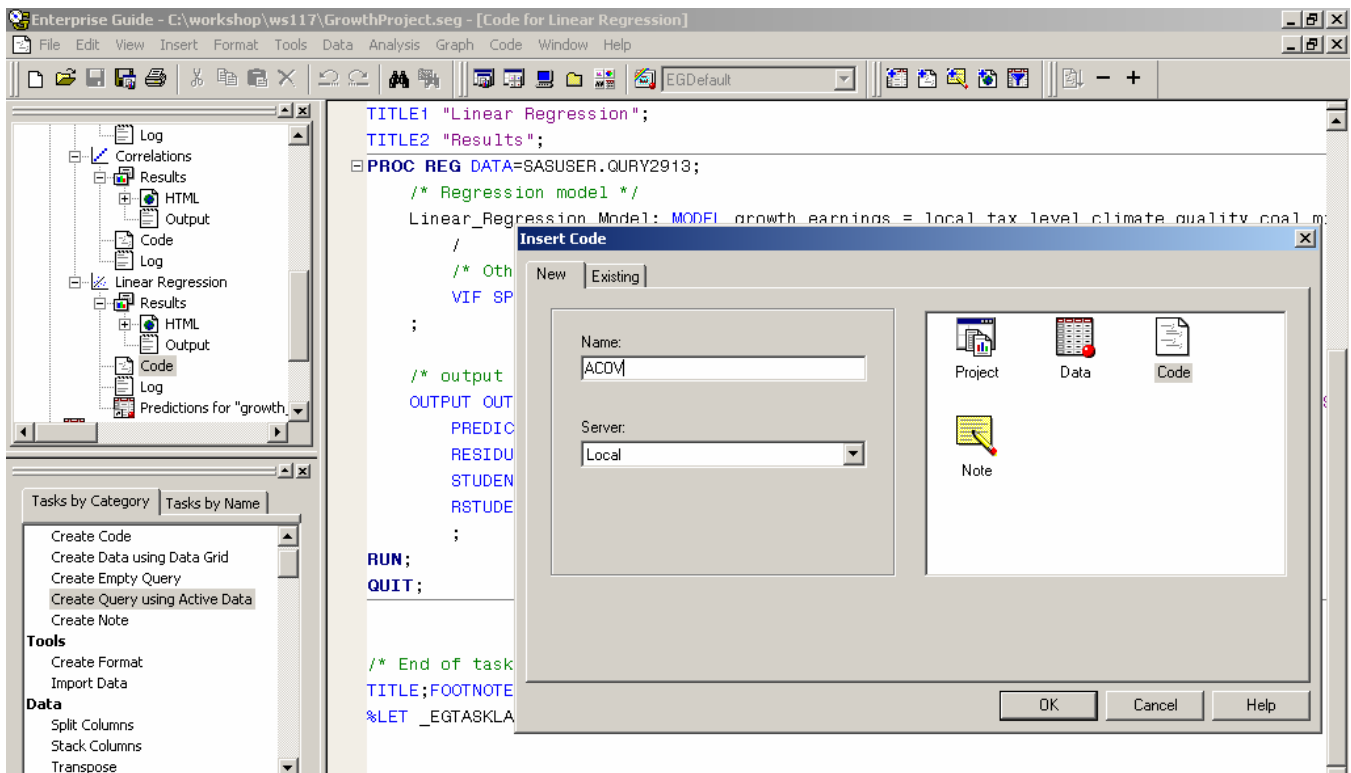
Note the option SPEC on the MODEL statement. We could learn more about this option by clicking on **Help**→**Enterprise Guide Help**, clicking on the **Search** tab and entering "spec" as the search term. On the long list of results, double-click on the entry near the top labeled "Testing for heteroskedasticity".



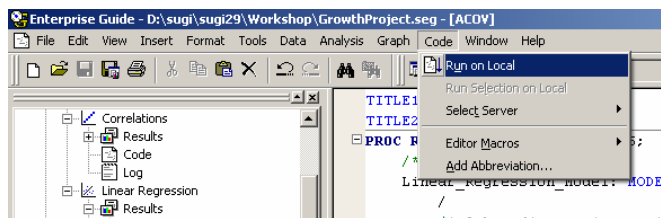
Reading the Help information, we learn of an option called ACOV in PROC REG that produces White's consistent estimate of the covariance matrix that is valid under heteroskedasticity. In order to have EG perform a task not available on its menus, we can enter our own SAS code for execution. In this case, we only need to add a single option to the MODEL statement, so we can copy-and-paste from the **Code** window under the **Linear Regression** task to our own Code window. After closing the Help window, we begin by double clicking on the Code window under the Linear Regression task and using the mouse to mark and **Edit** → **Copy** the PROC REG code through the MODEL statement:

```
PROC REG DATA=SASUSER.QUERY6216;
  /* Regression model */
  Linear_Regression_Model: MODEL growth_earnings = local_tax_level climate_quality
  coal_mines mean_annual_earnings_log educ_srvcs
  /
  /* Other diagnostic statistics */
  VIF SPEC
;
```

Now click on **Insert** → **Code** from the main toolbar, select the **New** tab and name the new code icon ACOV:



Click **OK**. Now paste the PROC REG code into the **ACOV** window and add the option ACOV to the MODEL statement (with the other options; for example, right after SPEC). Then click **Code** → **Run on local**:

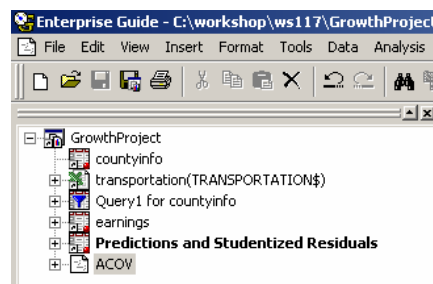


The output now contains White's consistent estimate of the covariance matrix that accounts for non-constant error variance.

We're now ready to create out-of-sample predictions using the estimated model.

OUT-OF-SAMPLE PREDICTIONS

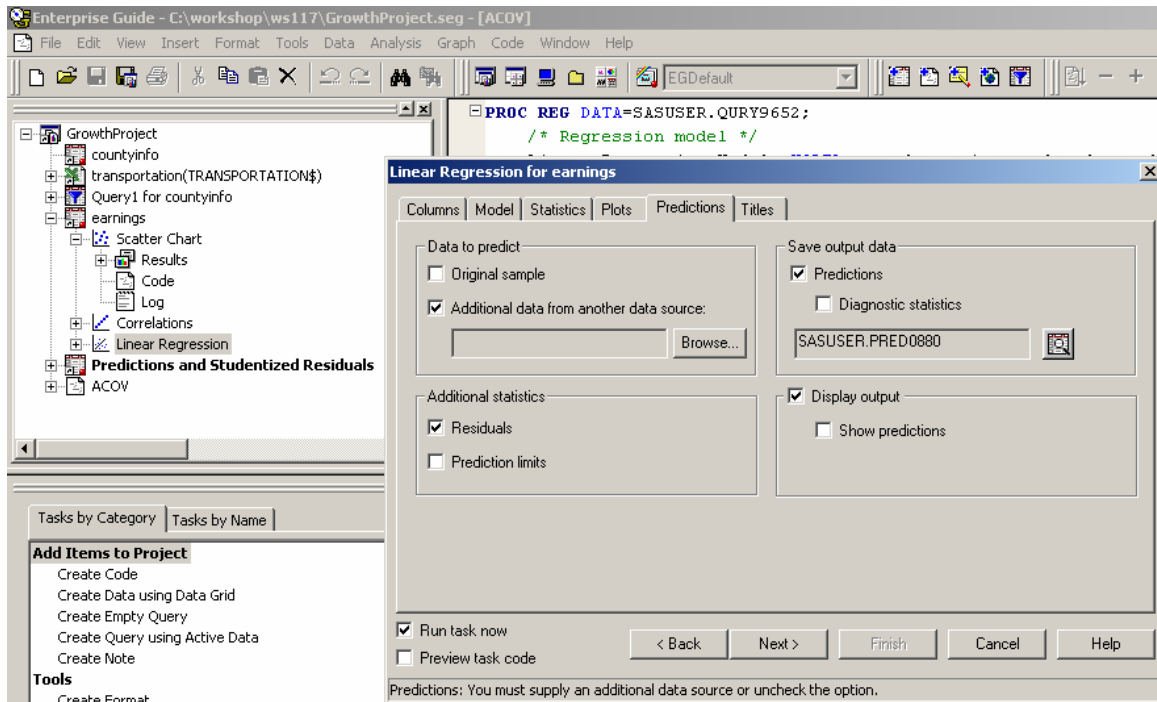
This would be a good time to clean up the Project Window by collapsing all the icons.



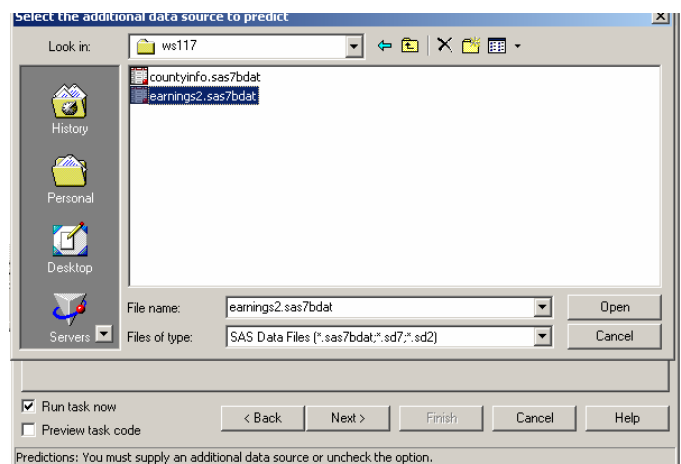
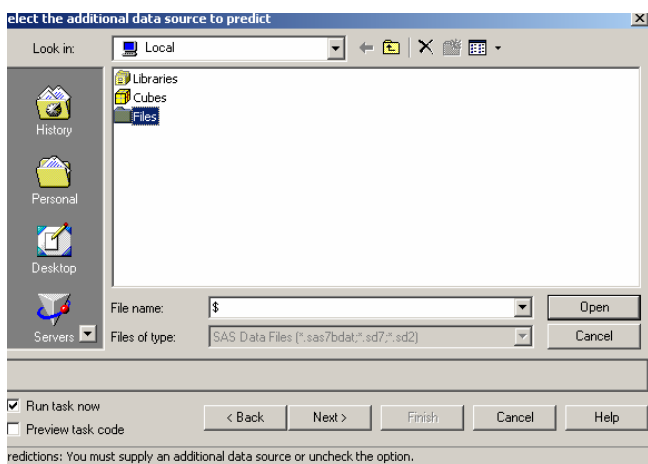
Another SAS data set, called earnings2, contains data for the explanatory variables in our model, but has missing values for

the dependent variable, growth_earnings. The goal is to use our estimated regression model to calculate predicted values for growth_earnings for this data set.

We begin by double-clicking on the **Linear Regression** icon, which is under the earnings icon in the Project window, and clicking on the **Predictions** tab. Click in the box under "Data to predict" labeled "**Additional data from another data source**", deselect the "**Original sample**" option, and, finally, click on **Browse**.



In the window labeled "**Select the additional data source to predict**", double-click on **Files** and navigate to the folder C:\workshop\ws117 and select the file earnings2.sas7dbat. Click on the **Open** button.



When the **Predictions** dialog reappears, click on **Finish**. A new icon labeled "Predictions for growth_earnings ..." appears under the Linear Regression task. Double-click on this icon to view predicted values for growth_earnings for the 128 observations in this data set.

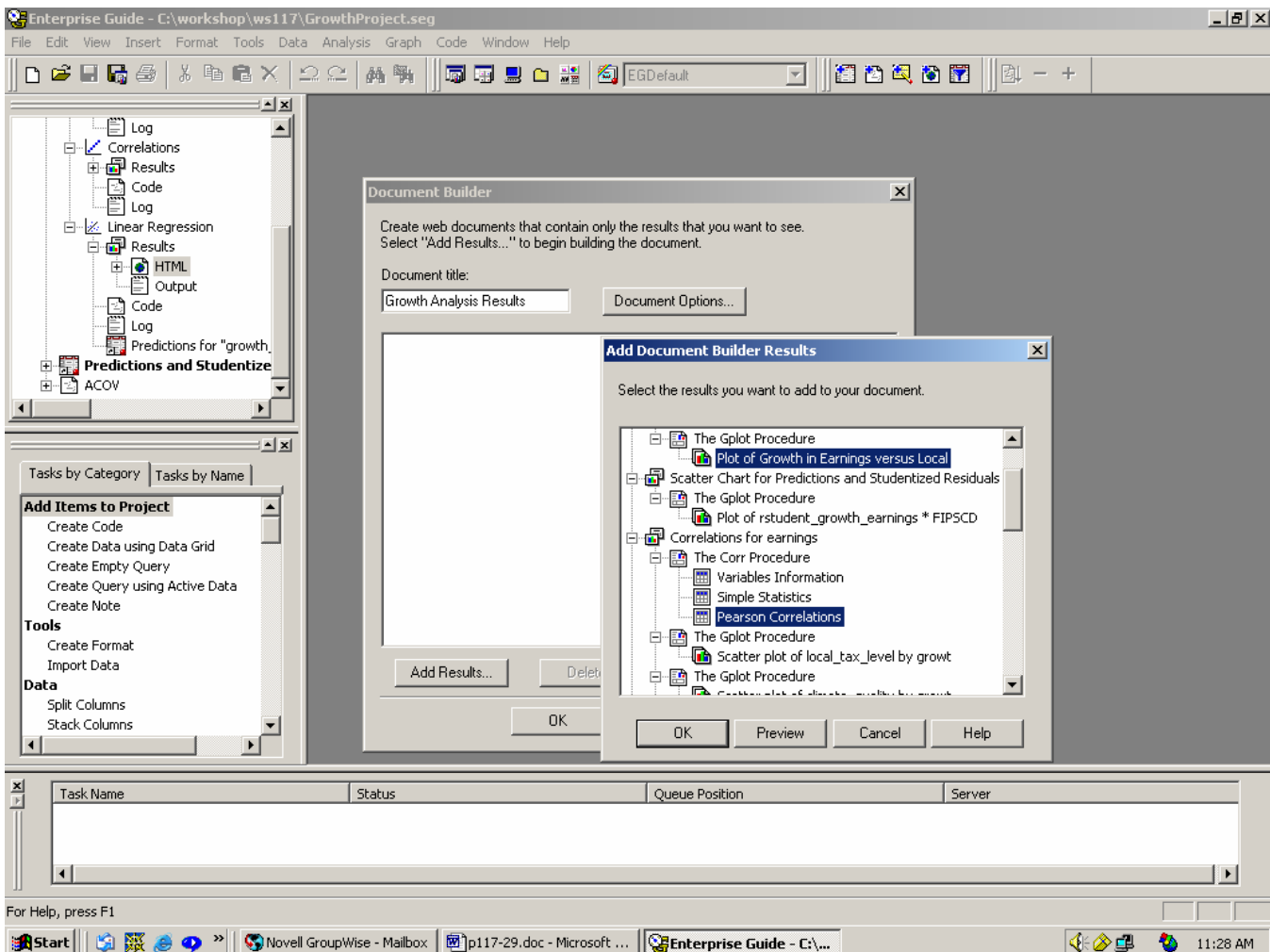
	growth_earnings	predicted_growth_earnings	residual_growth_earnings	student_growth_earnings	rstudent_growth_earnings
1	.	-11.18570082	.	.	.
2	.	-4.06107716	.	.	.
3	.	8.1444526226	.	.	.
4	.	-3.710559693	.	.	.
5	.	8.9538840161	.	.	.
6	.	1.2070404177	.	.	.
7	.	8.5850431785	.	.	.
8	.	1.5653911844	.	.	.
9	.	-1.572444444	.	.	.
10	.	4.8666176301	.	.	.
11	.	5.4381464711	.	.	.
12	.	5.0630324878	.	.	.
13	.	8.5781624076	.	.	.
14	.	7.6436847759	.	.	.

This concludes the modeling part of this workshop.

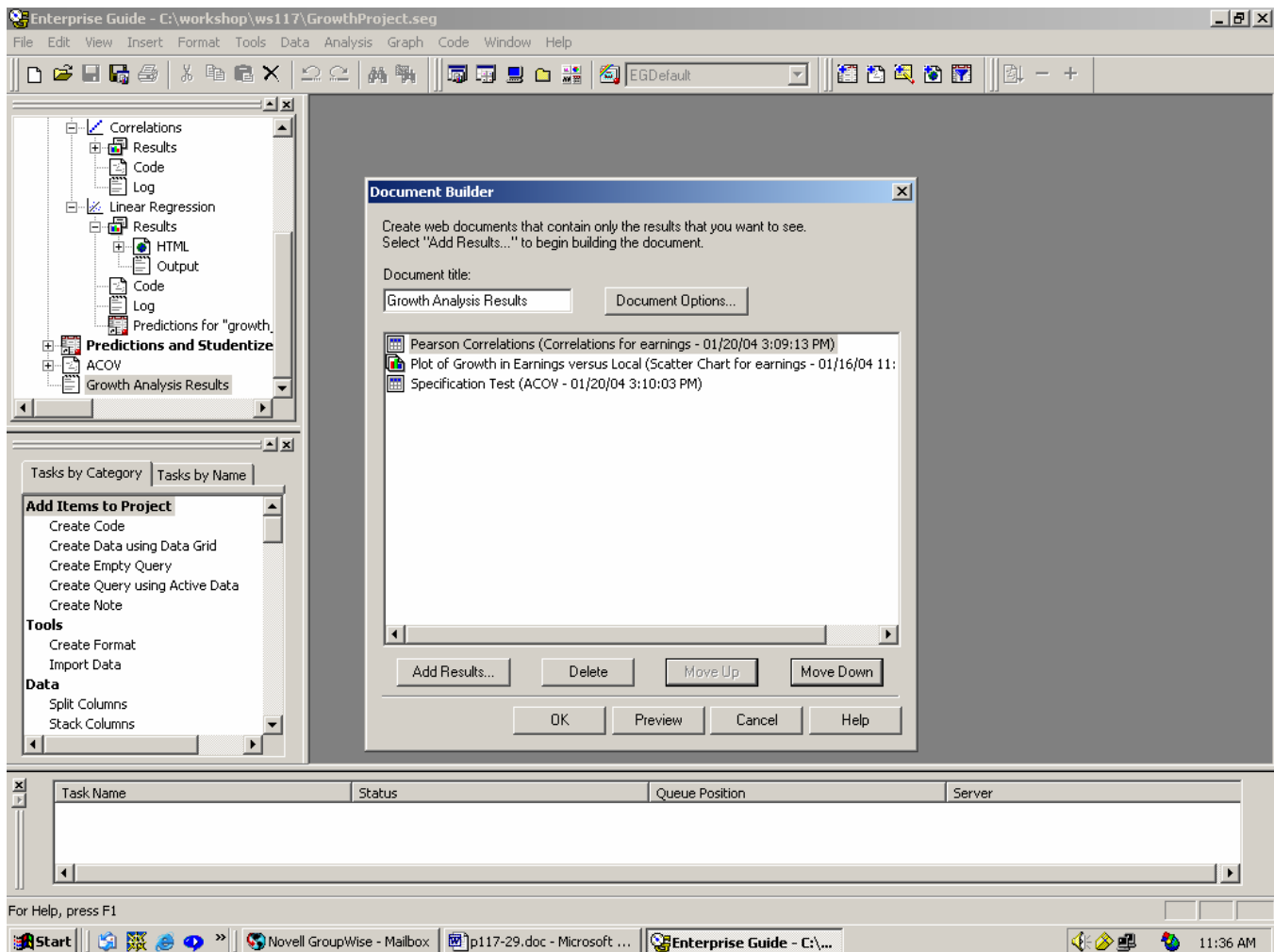
CREATING A DOCUMENT OF RESULTS

The tasks that we've run have generated a lot of output; some of it for our own exploratory use, other of it suitable for final presentation. We'd like to extract those that fall into the second category and put them together into a report. The Document Builder in Enterprise Guide creates a document definition of instructions for combining HTML results from multiple tasks.

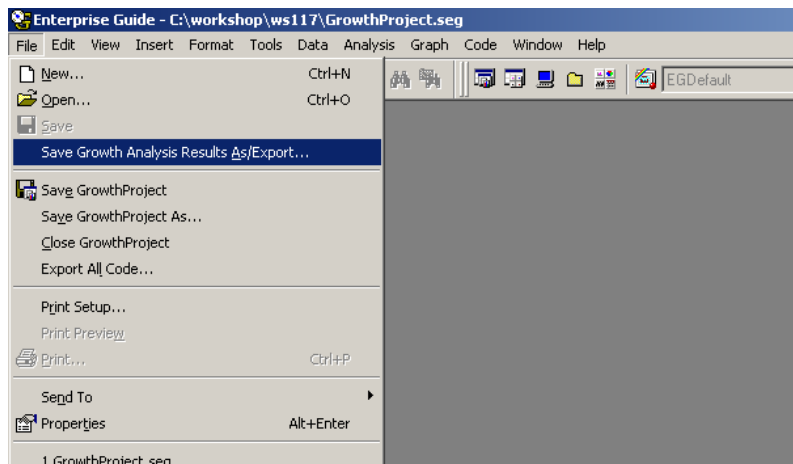
Click on **Tools → Document Builder...**. Type **Growth Analysis Results** for a title, and click on **Add Results...**. Holding down the **Ctrl** key, highlight items you'd like to include in a final report. For example, select **Plot of Growth in Earnings versus Local** in the Scatter Chart task, the **Pearson Correlations** table in the Correlations task, and **Specification Test** in the ACOV task. Click on the **OK** button.



You'll be presented with a list of the items currently in the document, and can add further items or delete some of those already included. We can also rearrange the presentation of the output. Highlight the **Pearson Correlations** results and click on the **Move Up** button to present these first. Click on the **Preview** button to see what the document looks like so far, scroll through the display and then close the browser.



We don't have to add items one table at a time. Click on the **Add Results** button and select the **Linear Regression** task. All of the result items under the task will be included. Click on **Preview** to confirm this, then close the browser and click on the **OK** button. Note that we have just defined the items to be included in the document; to publish it, click on **File → Save Growth Analysis Results As/Export...** You can reformat the document at other times; double-click on its icon in the Project Window, click on the **Document Options**, select another style from the drop-down list (for example, **BarrettsBlue**), click on **OK** and then preview the result. EG also has a nice style editor for customizing your own (accessed through the **Customize Styles** button on the toolbar).



CONCLUSION

Enterprise Guide provides a powerful interface to the suite of tools in the SAS System for conducting statistical analyses. The menus and selection dialogs make it easier to find the correct options available in the analytic procedures. The organization into projects helps group related tasks and the data being examined. EG's Document Builder aids in putting the results together into a final report. An analyst does not need to be a SAS programmer to accomplish these goals.

This workshop has only touched on a few of the capabilities of this software tool. Users are encouraged to explore other options available in the menu system and dialog boxes.

REFERENCES

Factors Associated with Rural Economic Growth, Lessons from the 1980's, Kusmin, Lorin D., Redman, John M., and Sears, David W., U.S. Department of Agriculture, ERS Technical Bulletin Number 1850, 1996.

Accomplishing Tasks in SAS® Using Enterprise Guide® Software Course Notes, SAS Institute Inc., Cary, NC, 2002.

ACKNOWLEDGMENTS

The authors thank Lorin Kusmin of the Economic Research Service for providing the data used in the examples for this workshop.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Charles Hallahan
Economic Research Service, USDA
1800 M Street NW, Room S2070
Washington, DC 20036-5831
Work Phone: (202) 694-5051
Fax: (202) 694-5781
Email: hallahan@ers.usda.gov

Linda Atkinson
Economic Research Service, USDA
1800 M Street NW, Room S5015
Washington, DC 20036-5831
Work Phone: (202) 694-5046
Fax: (202) 694-5781
Email: atkinson@ers.usda.gov

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.