IT-82

SAS-Enterprise Guide® for Institutional Research and Other Data Scientists

Claudia W. McCann, East Carolina University

Abstract

Data requests can range from on-the-fly, need it yesterday, to extended projects taking several weeks or months to complete. Often institutional researchers and other data scientists are juggling several of these analytic needs on a daily basis, i.e., taking a break from the longitudinal report on retention and graduation to work on responding to a USN&WR survey to answering the simple 5 minute data query question from an administrator.

SAS Enterprise Guide is a terrific tool for handling multiple projects simultaneously. This interactive tutorial is designed to walk the data analyst through the process of setting up a project, accessing data from several sources, merging the datasets, and running the analyses to generate the data needed for the particular project.

Specific tasks covered are pulling SAS datasets and Excel files into the project, exploring several facets of the ever-so-powerful Query Builder, and utilizing several quick and easy descriptive statistical techniques in order to get the desired results.

SAS-Enterprise Guide® for Institutional Research and Other Data Scientists

Claudia W. McCann, East Carolina University

The SAS-Enterprise Guide (SAS-EG) software gives you the power of SAS with a windows point and click interface.

When you open SAS-EG for the first time, your screen will look something like this:

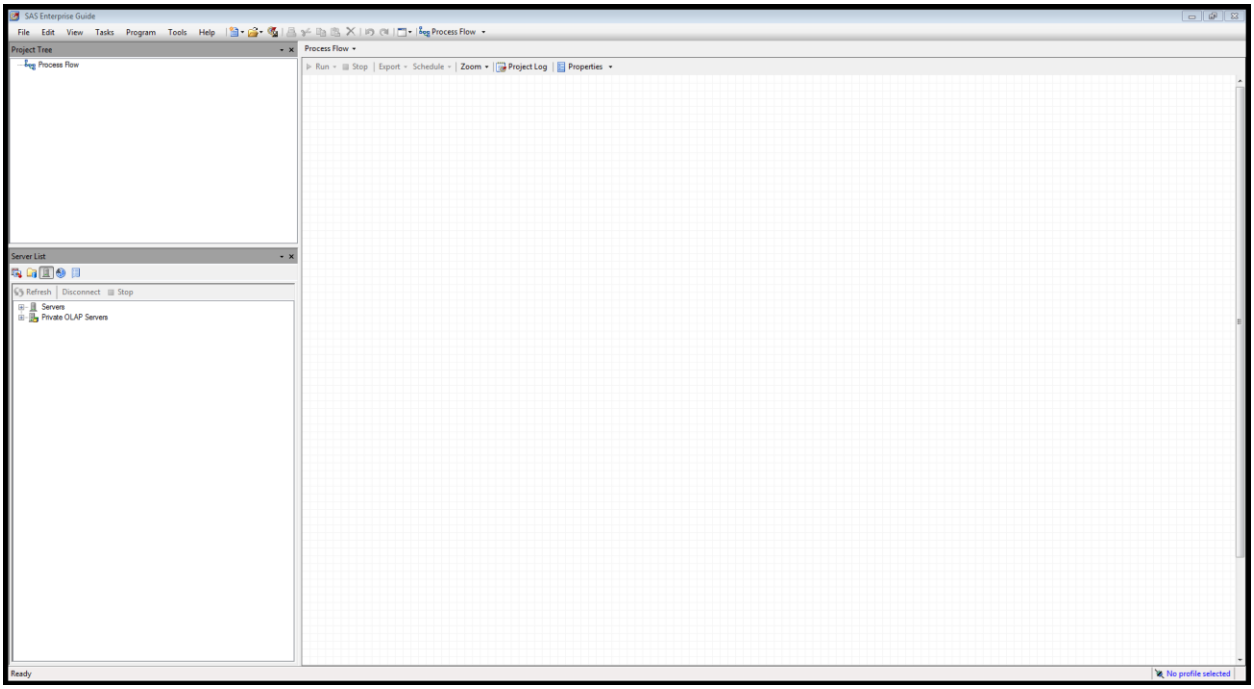


Figure 1. Opening Screen for a SAS-EG Project

As you work with SAS-EG, you will quickly learn what panes you want open and how you want them arranged. There are lots of view options. For this demonstration, we are going to change the view to my own personal preference which is to turn off the project tree and move the Server List to the right.

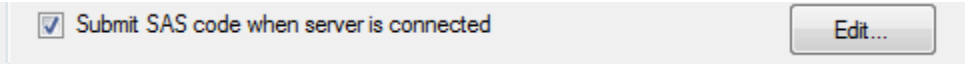
➔ Close Project Tree
➔ On Server List, click down arrow and dock right

Now have two panes open, the Process Flow which is your workspace and the Server List.

A valuable tool for me is to have a server access program run each time I open SAS-EG. The files I access most often are in libraries other than

those that open as default in SAS-EG. To run a program each time SAS-EG opens:

➔ Click on Tools
➔ Click on Options
➔ Click on SAS Programs
➔ Click 'on' the Submit SAS code



➔ Click on Edit to enter or modify SAS code.

The following is an example of the code I run when at my institution.

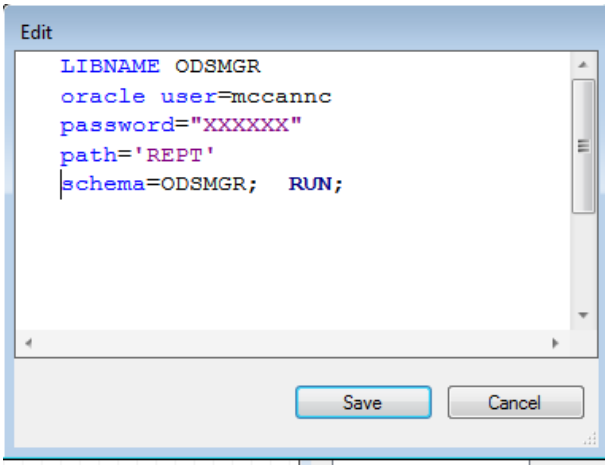


Figure 2. Sample of Code to Submit at SAS-EG Start-up

The following is a typical request one might receive during the course of the day. I actually received this request while working on this presentation.

Attached is the most recent table of reported certification rates. Dr. Benfield, do you have reports for 2013/14 from FNP/AGNP and CNS and NNP? They should have results for 2013, may or may not for 2014.

Claudia, can you insert the number of graduates for these concentrations?

Thanks so much.

Here is a portion of the table:

Year	Certification Organization	Certification Exam (by specialty area)	# Graduates	# Students Taking Exam	Certification Pass Rate
2009	ANCC	Family Nurse Practitioner**		15	100%
2010	AANP & ANCC			31	87.1%
2011				30	93.3%
2012				28	95.5%
2013					
2009	ANCC	Adult/Gerontology Nurse Practitioner**		14	92.9%
2010	AANP & ANCC			34	94.1%
2011				26	92.3%
2012				23	82 %
2013					
2009	American Midwifery Certification Board	Certified Nurse Midwife		7	57.0%
2010				7	57.0%
2011				8	100%
2012				12	92%
2013				8	87.5%

Figure 3. Typical Data Request

Our job is to fill in the # Graduates column. A couple of data knowledge points you may need as we proceed:

- The year variable is calendar year
- Certification Rates applies to graduate level students
- Specialty Area is First_Concentration_Desc and/or Program_Desc

Our first step is to bring data into this project. There are two ways to do this.

1. Using the Server List
   → Expand the folders
   → Double click the file you want to open
2. Using the Workspace
   →Click on File
   →Click Open
   →Navigate to the data file you want to open, double click

First, open the data file called graduation_data.sas7bdat using either one of the methods described above. The file is located on:

C:|SESUG14|IT|McCann

As with any software, it is always a good idea to save your work periodically. Let's save this project now:

➔ Click on File at the top left.
➔ Click on Save Project as
➔ Save as *SESUG 2014 SAS-EG Project* in *C:|SESUG14|IT|McCann*

Now, back to the data file:

Using the **Query Builder**, we are going to:

➔ Change the Query Name to 'MSN and PMNC Graduates'
➔ Select the following variables

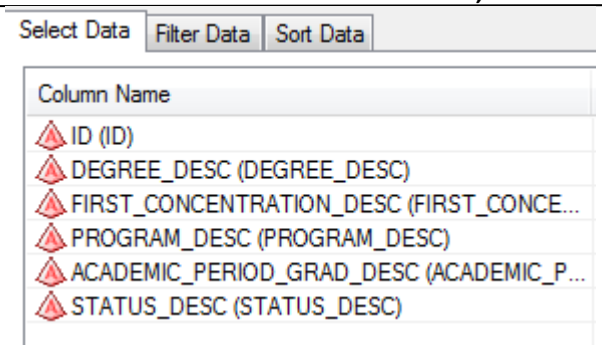


Figure 4. Select Data

Filter the data as follows:

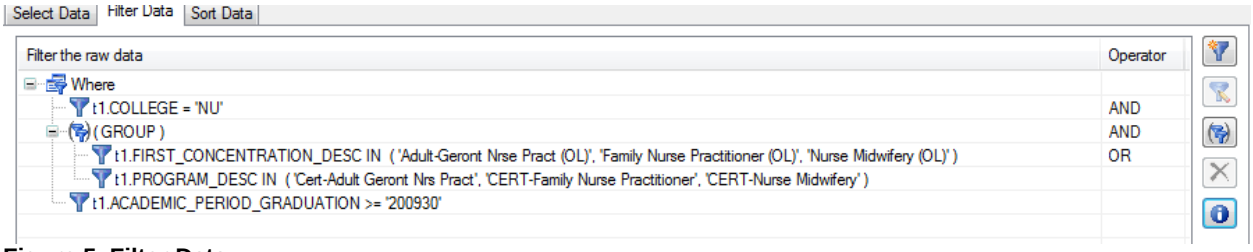


Figure 5. Filter Data

➔ Run the Query
(Should have 352 records)

Run PROC FREQs to be sure we have 'good' data

➔ Click on Describe
➔ Click on One-Way Frequencies
➔ Click and drag ID and Status_Desc over as Analysis variables
➔ Click on Statistics and then on Frequencies only

➔ Click on Results and then change the Order output data by: to Descending frequencies

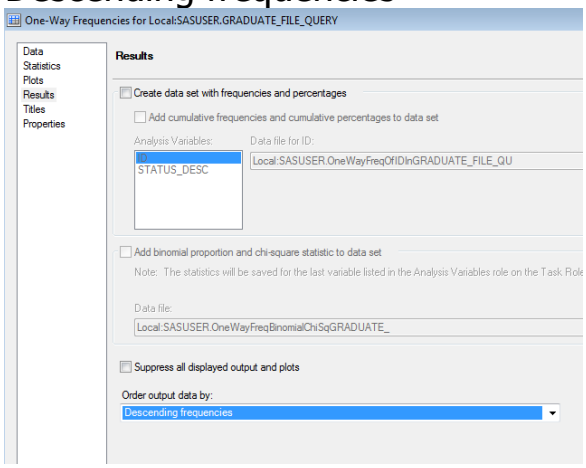


Figure 6. Order Ouput Data

➔ Click on Titles. (you can change the title that appears on your output here)

➔ Click on Properties, then on Edit, and change the Label to 'Data Verification'

➔ Click on Run

You will notice there are two duplicate IDs. This warrants investigating, i.e., did the student earn two separate degrees or is there a problem with the data?

Scrolling down further, you will notice the Status variable has Degree Awarded and Applied for Graduation. We only want Degree Awarded.

➔ Close the output by clicking on the small **x** in the upper right hand corner.

Modify our Query

➔ Right click on the MSN and PMNC Graduates icon

➔ Click on modify

➔ Check the Select Distinct rows only box (removes duplicates)

➔ Click on Sort Data, drag ID over into Column Name. (In order to verify the duplicate IDs, we will sort the output by ID and then find those records)

➔ Adjust the filter data by dragging STATUS across, click on the down arrow to the right, click on Get Values, click on G, click Finish. (pulls in only records of those who have graduated)

➔ Click Run again. The following prompt appears:

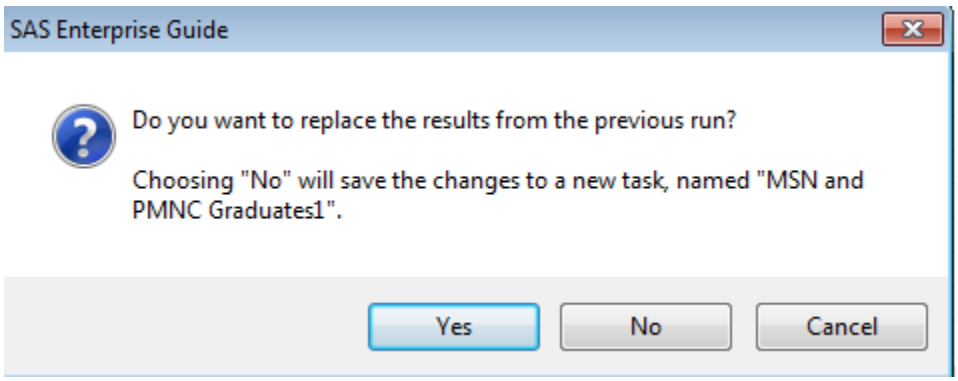


Figure 7. Replacement Prompt

In this case, we want to click Yes. Later in the demonstration, we will use a scenario where we want to click no.
(Should have 300 records this time)

➔ Rerun the PROC FREQ by right clicking the Data Verification icon, clicking on Run, and reviewing the output.

Now that we have achieved T-LAR (That Looks About Right) we want to run a second Query Builder to recode Academic_Period_Grad_Desc into calendar year.

- ➔ Open the Query for Graduation Status (double click on the icon)
- ➔ Click on Query builder
- ➔ Change the Query name to 'Generating Grad Year'
- ➔ Pull all variables over into the Select Data pane by clicking and holding the file name and dragging into the pane
- ➔ Click the Computed Columns icon
- ➔ Click New
- ➔ Click Recoded Column and Next
- ➔ Click on Academic_Period_Grad_Desc and Next
- ➔ Click on Add
- ➔ On the Replace a Condition
  - select *contains* as the Operator
  - type *2009* in the Value box
  - type *2009* in the With this value box
  - Click OK.
- ➔ Click on 'Add' and 'Replace a Condition' for 2010, 2011, 2012 and 2013.
- ➔ Click Next
- ➔ Rename the Identifier and Column name to 'Calendar Year Graduated', click Finish, and then Close. This new variable now appears in your Select Data.
- ➔ Run the Query

Run a One-Way Frequency on the new variable.

Calendar Year Graduated	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2009	52	17.33	52	17.33
2010	73	24.33	125	41.67
2011	57	19.00	182	60.67
2012	64	21.33	246	82.00
2013	46	15.33	292	97.33
Spring 2014	8	2.67	300	100.00

Figure 8. Frequency Output

We now have all the variables we need to fill out the table. To get the breakout needed:

- ➔ Open the dataset out of Generating Grad Year
- ➔ Click on Describe
- ➔ Click on **Summary Tables Wizard**
- ➔ Click Edit… to set overall Task Filters using the edit button
  - In the first pull down box, select First_Concentration_Desc
  - In the second pull down box, select Equal to
  - Click on the  button and select Family Nurse Practitioner and then OK
  - Click on the last pull down box and select OR
  - Repeat these steps to pull in Program_Desc, Equal to, CERT-Family Nurse Practitioner.
- ➔ Click OK.
- ➔ Click Next>
- ➔ Skip the Analysis Variables pane (2 of 6) by clicking Next>
- ➔ Add First_Concentration_Desc and Program_Desc Columns
- ➔ Add Calendar Year Graduated to Rows
- ➔ Finish

Summary Tables

	FIRST_CONCENTRATION_DESC		Total
		Family Nurse Practitioner (OL)	
	PROGRAM_DESC	PROGRAM_DESC	
	CERT-Family Nurse Practitioner	MSN-Nursing	
	Frequency	Frequency	Frequency
Calendar Year Graduated			
2009	2	24	26
2010	4	28	32
2011	3	22	25
2012	2	27	29
2013	6	17	23
Total	17	118	135

Generated by the SAS System ('Local', W32_7PRO) on July 17, 2014 at 1:38:03 PM

Figure 9. Summary Tables Output

We can now fill in the Family Nurse Practitioner # Graduates column on our table.

Next we need to get the comparable data for Adult/Gerontology Nurse Practitioner and for Certified Nurse Midwife. First, let's rename the existing Summary Tables icon.

➔ Right Click on the Summary Tables icon
➔ Click on Rename
➔ Type in FNP Grads

Now, to run the same table for Adult/Gerontology

➔ Right click on the FNP Summary Tables icon
➔ Click Modify FNP
➔ Click Edit… and change the task filters to Adult/Gerontology
➔ Click OK
➔ Click Finish

Here is a situation where you would answer No to the replace prompt.

We can now fill in the Adult/Gerontology Nurse Practitioner # Graduates column on our table.

➔ Close the Summary Tables and rename to AGNP

Repeat this process for Nurse Midwifery by modifying the Summary Tables AGNP, changing the selection edits to Nurse Midwifery, and Finish.

Let's save the project and close it.

➔ Click on File in upper left hand corner and select save SESUG 2014 SAS-EG Project
➔ Click on File in upper left hand corner and select Close Project

< class break >

To reopen the project, SAS-EG has a nice feature under File that lets you reopen recent projects.

➔ Click on File
➔ Click on Recent Projects (toward the bottom of the menu)
➔ Select SESUG 2014 SAS-EG Project.egp

Often time as projects develop, you want to move icons into more meaningful arrangements. The SAS-EG default is to auto arrange the icons. You turn this off simply by:

➔ Right clicking on any blank space in the process flow pane or grid
➔ Click off Auto Arrange

Now we want to move all the icons lower down.

➔ Click and hold, draw a box around all the icons. All of the them should be highlighted
➔ Click and hold again and pull the entire set down leaving about 4 inches at the top.

Often we need to match data between two or more tables. This is a pretty straightforward process in SAS-EG using the Query Builder.

Import the Excel file called *BSN Cohort Fall 2011.xlsx* using either one of the methods described earlier. The file is located on:

C:|SESUG14|IT|McCann

➔ Double click on the file to open it. SAS-EG then prompts you for some import information.
➔ Click Next

On Screen 2 of 4, be sure this box is checked

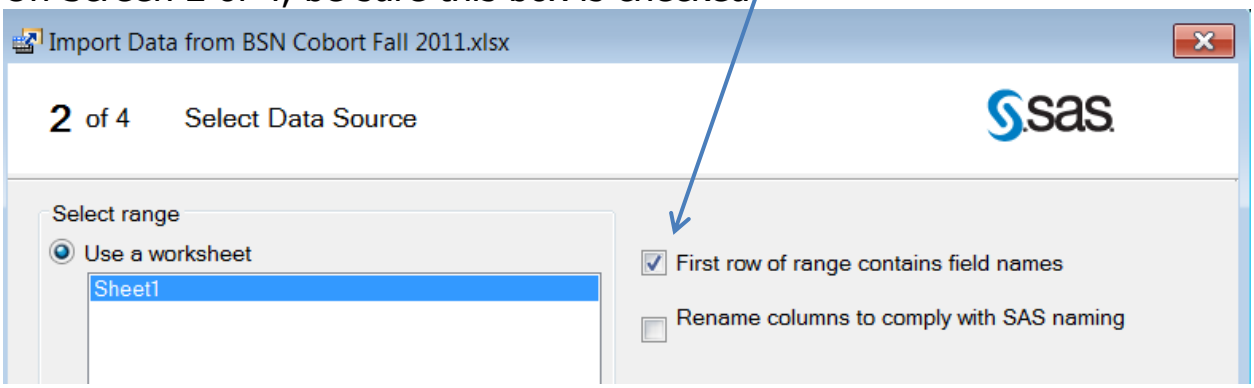


Figure 20. Import Data

➔ Click Next. Screen 3 of 4 allows you to make all sorts of decisions and modifications to your data. For now,
➔ Click Next and then Finish

Import *enrollment_fall_2012.sas7bdat* using either one of the methods described earlier. The file is located on:

C:|SESUG14|IT|McCann

In order to join files, there needs to be a common identifier or identifiers. On our other files, the variable ID is used as a unique identifier. We are also going to make several adjustments to this file to make it more functional and easier to read.

First of all, the ITEM006 field needs identifiers to know what the numbers mean. We can easily run a **proc format** within SAS-EG to apply value labels to the numbers.

- ➔ Click on Tasks
- ➔ Click on Data
- ➔ Click on Create Format…
- ➔ Type EnrlStat as the Format name: The Library: option lets you save the format temporarily in the Work library or you can save it permanently to your SASUSER library. We will write the format temporarily in Work.
- ➔ Click on Define Formats in the leftmost box.
- ➔ Click on New and type in the following:

Label	Discrete
New Student	1 (Then Click New again)
New Transfer Student	2 (Click New again)
Continuing Student	3
Returning Student	4

- ➔ Click Run

Your log should look something like:

```
33          PROC FORMAT
34           LIB=WORK
35          ;
36
36        !   VALUE $EnrlStat
37                  "1" = "New Student"
38                  "2" = "New Transfer Student"
39                  "3" = "Continuing Student"
40                  "4" = "Returning Student";
NOTE: Format $ENRLSTAT has been output.
41          RUN;
```

Figure 31. Log from PROC FORMAT

Next we want to Query Build this dataset so that we are only looking at Nursing Students. You will notice that the N for this dataset is 26,947 which is the enrollment count of all students for Fall 2012. We only want to look at students enrolled in the College of Nursing (CON).

Make a smaller dataset by:

- ➔ Open the enrollment_fall_2012 dataset
- ➔ Click on Query Builder
- ➔ Change the Query Name to CON Enrollment Fall 2012
- ➔ Move ITEM100 and ITEM006 to the Select Data box.
  We are going to edit these fields so they are easier to use.
  - Click on Item100 to highlight it and then on the properties icon to the right: 
  - Change the Column Name to ID and click OK
  - Click on ITEM006 and then on the properties icon.
  - Change the Column Name to Enrollment Status
  - Click on the Change button next to the Format Box
  - Click on User Defined and then on the EnrlStat format that we just created and click OK
- ➔ Click on the Filter Data tab and drag ITEM80A over.
- ➔ Change the Operator to Between
- ➔ Put in a Start Value of 513800
- ➔ Put in an End Value of 513899
- ➔ Click Finish
- ➔ Run the query.

There should be 1,271 records in the dataset. This is the count of students enrolled in the college for the Fall 2012 semester.

We now have two files that we want to join. The question we are trying to answer is what is the retention rate of the Fall 2011 BSN Cohort to the Fall 2012 semester.

- ➔ Double click the cohort file (Data Imported…) that we created from the excel file.
- ➔ Once the file is open, Click on Query Builder
- ➔ Change Query Name to Merge
- ➔ Click on Add Tables. The SAS-EG default is to make all files in the project available for adding to the query. This is where having renamed the queries pays off.
- ➔ Expand the Source column and click on the Name associated with CON Enrollment Fall 2012.
- ➔ Click on Open. Two tables, t1 and t2 should be listed on the left.
- ➔ Pull ID from the Cohort file (t1) over to the Select Data area.
- ➔ Double click on Join Tables
- ➔ Change the join to include all records from t1.
  - Right click on the join symbol
  - Click on Properties
  - Select on 'All rows from the left table …(Left Join)'
  - Click OK
- ➔ Click Close

We now want to recode the variable Enrollment Status to be Enrolled or Not Enrolled.

- ➔ Double Click on Computed Columns
- ➔ Click on New
- ➔ Click on Recoded column
- ➔ Click Next
- ➔ Click on t2.Enrollment Status
- ➔ Click Next
- ➔ Click on Add
- ➔ Click on the pull down menu arrow next to add
- ➔ Click on Get Values
- ➔ Click on 1, hold down the shift key and click on 4. This will select all the values. Click on OK.

➔ Type 'Retained' in the With this value: box.
➔ Click OK
➔ Under Other values, click on Specify a value and then type 'Not Retained' in the box.
➔ Click Next
➔ Change the Identifier and Column Name to Retention Status
➔ Click Finish and Close

Since we are only interested in pulling retention information on new pre-licensure BSN cohort members, we need to filter out other types of enrolled students.

➔ Click on the Filter Data tab
➔ Drag over t1.New Student Type and click on the pull down menu arrow beside the Value: box.
➔ Click on Get Values and select New Pre-Licensure and click finish.

We are now ready to run this query.

➔ Click Run

You should have 131 records. To determine retention rate:

➔ Click on the pull down menu arrow to the right of Describe
➔ Select One-Way Frequency
➔ Pull Retention Status over as the Analysis Variable
➔ Click Run

Your output should look like this.

Retention Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Not Retained	4	3.05	4	3.05
Retained	127	96.95	131	100.00

Figure 42. One-Way Frequency

Our retention rate for this cohort is 96.95%.

But who are the four not retained. There are several quick ways to determine who was not retained from within SAS-EG such as sorting the dataset on the variable and looking at the records or limiting a one-way frequency on ID to not retained but, I want to share how useful Excel Pivot Tables are in situations such as these.

Excel Pivot Tables

- ➔ Reopen the last dataset (out of the Merge query build)
- ➔ Click on Send To
- ➔ Click on Microsoft Excel (may be blinking on your windows tool bar)
- ➔ Highlight entire spreadsheet by clicking on upper left hand box:

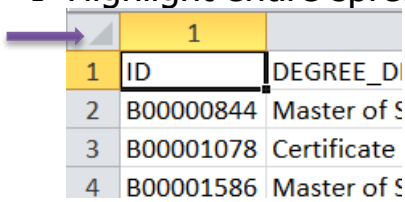


Figure 53. Excel Spreadsheet

- ➔ Click on the Insert tab at the top of the spreadsheet
- ➔ Click on Pivot Table
- ➔ Click on OK

With Excel you can quickly click and drag to generate tables.

- ➔ Drag the ID field into the values box
- ➔ Drag the Retention Status field into the Row Labels box.

- ➔ Now, if you double click on the 4 in the table…

One other task that I find much easier to do in Excel than in SAS is parsing; converting a single column into two or three, for example, changing a birthdate into a birth year, birth month, and birth day.

- ➔ Open the data file demographics.sas7bdat from the Server List or from File → Open → Data.

➔ Click on Send To and then Microsoft Excel
➔ Open the Excel file (blinking on the windows tool bar)
➔ Highlight the Birthdate Column
➔ Click on the Data tab
➔ Click on Text to Columns
➔ Click on Fixed width if it is not already selected
➔ Click next
➔ Create break lines to separate to separate year, month, and day.

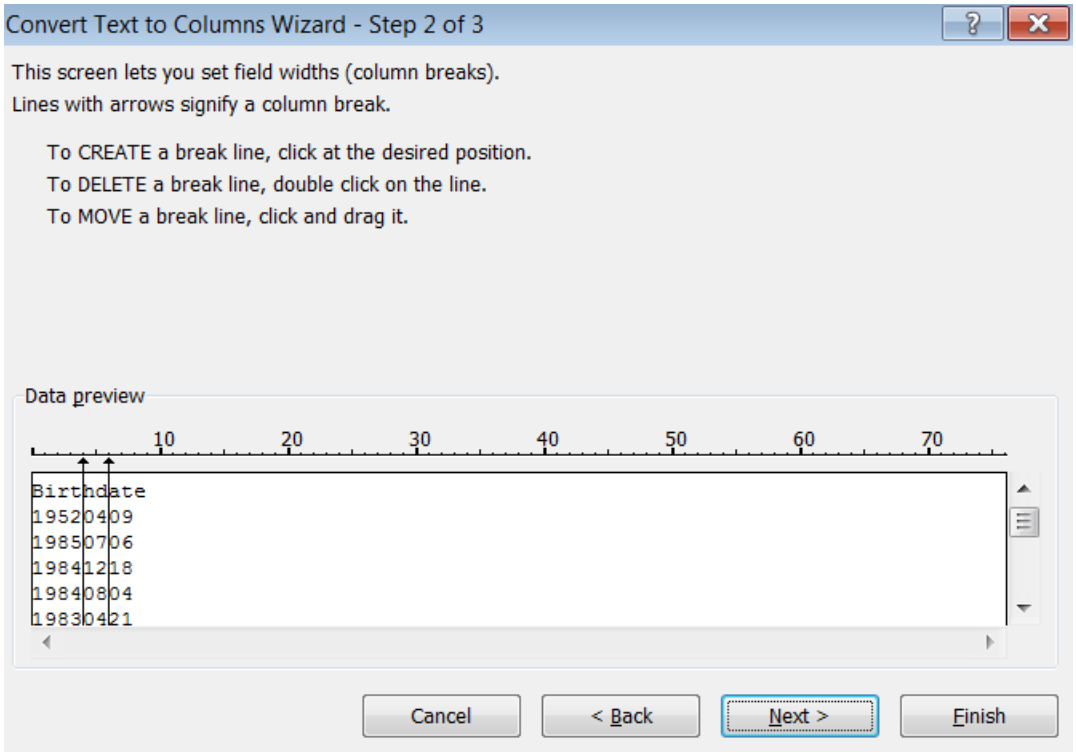


Figure 14. Text to Columns in Excel

➔ Click Next
➔ Click Finish
➔ Change the Headers in Row 1 to Byear, Bmonth, Bday.

This file can then be saved and imported back into SAS-EG with birthdate in an easier format to use such as generating an age variable.

CONTACT INFORMATION

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

Name:	Claudia W. McCann
Enterprise:	East Carolina University College of Nursing
Address:	2132 Health Sciences Building
City, State ZIP:	Greenville, NC 27817
Work Phone:	(252) 744-6476
E-mail:	mccannc@ecu.edu