

Navigating the Data Universe Aboard the Starship Enterprise Guide

Jay Iyengar, Independent consultant, Evanston, IL

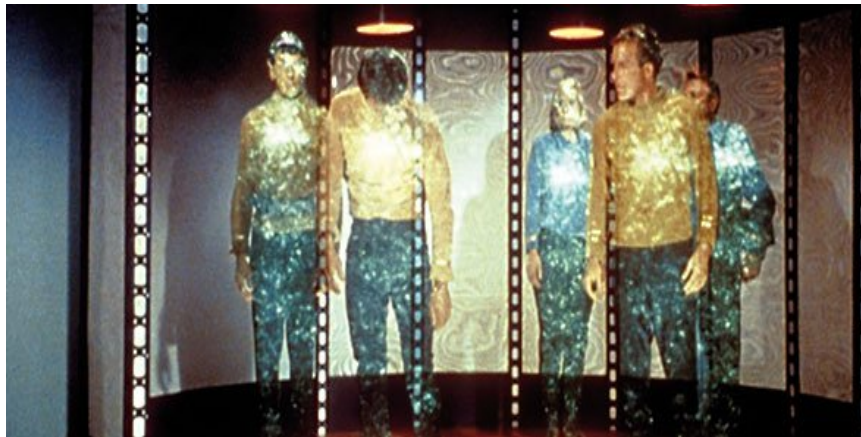
ABSTRACT

SAS® Enterprise Guide® is a standalone client application that gives the user access to the power of Base SAS, along with other SAS modules. With Enterprise Guide, one can easily perform the full range of data processing and analysis tasks, from reading, to manipulation, to reporting and graphics. Although EG is oriented towards the analyst with its point and click interface, it includes aspects useful to both the programmer as well as the analyst. This paper gives an overview of each task area, as well as focuses on the pros and cons of using EG.

INTRODUCTION

SAS® Enterprise Guide® (EG) has an intuitive, straight – forward graphic interface that those who are unfamiliar with technical SAS constructs can understand. Projects in EG are visually represented with a project tree flowchart diagram. Each task in EG has a built-in wizard which takes the user through a series of steps in the task process. As EG executes the task, it develops and runs the SAS code for that task in the background. Enterprise Guide is a good complement, but not a replacement to SAS Display Manager (SAS). Some tasks are performed more effectively by SAS, instead of EG, or by Base SAS programming, instead of EG's task wizards. Adding EG to your organization's SAS suite gives you a powerful set of tools to perform the multitude of tasks confronted by an analyst or programmer.

I. DATA ACCESS



"Beam Us Up Scotty!"

Enterprise Guide (EG) has capabilities for reading in external data files and converting them to SAS data sets. The Import Data wizard in EG consists of 4 steps where you define the external path and filename, fields, variable attributes and the output data set, as well as other options. The wizard has data access tools for reading several types of external data files, including PC files, text, csv, and SPSS/Stata data files.

The wizard provides the user the option of whether to use a DATA step with an INFILE and INPUT statement, or the IMPORT Procedure using the SAS/Access module. If the external file is an Excel worksheet, the user can decide whether to specify a range of cells to import, or the entire worksheet. The field attributes that you're allowed to define and modify include the variable name, type (numeric or character), length, informat, and format. Figure 1.1 below shows the Import Data task wizard.

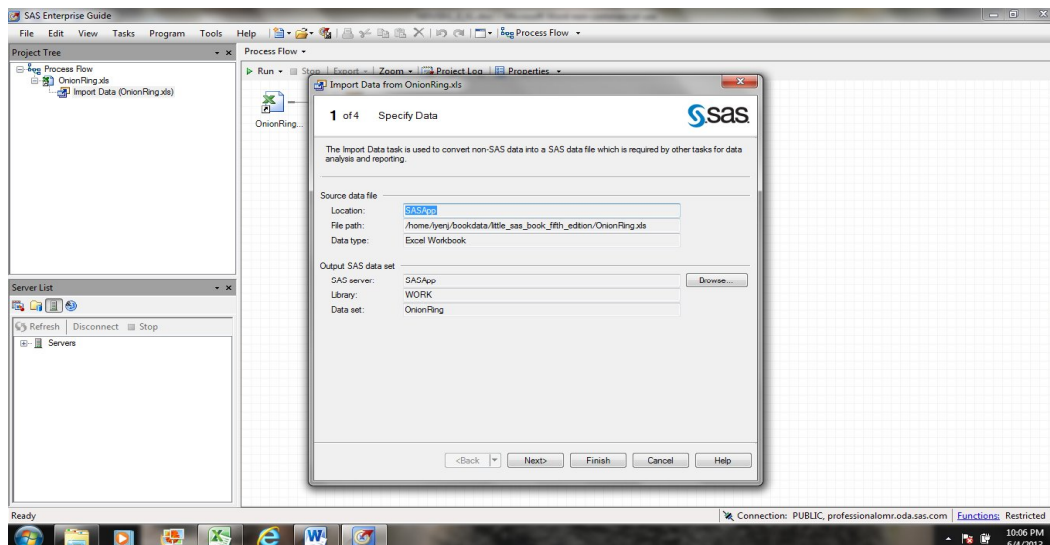


Figure 1.1

I would contend that SAS EG provides enhanced data access abilities beyond those provided by the Import or Export wizards in SAS. The Import/Export wizards don't give the user the ability to select which fields to include or exclude from the external file in the resulting SAS data set. With the Import/Export wizards, the user can select which worksheet to import from an Excel workbook, but not a specific range of cells. Although, you can do this in SAS using DDE, and using DDE with EG can be tricky. But the ability to modify variable attributes (length, type) using a point-and-click interface to convert data, is unique to EG. Table 1.1 below lists the various SAS\Access for PC files file types that can be read in using EG, the SAS code used and the SAS module needed in your installation.

SAS Task	SAS Procedures Used	SAS Module
Import Data (Microsoft Excel\Access, text files)	Proc IMPORT, Data Step	BASE, SAS/Access
Import JMP file		
Import SPSS file		
Import Stata file		

Table 1.1

EG also has the ability to read and access data from relational database systems and platforms. This is another strength of EG in accessing data. Enterprise Guide has the ability to connect to databases in Oracle, DB2, SQL Server, and MySQL formats. It also supports ODBC and OLEDB. To read relational databases using the SAS\Access procedures, you need to obtain the specific license 'SAS\Access for Relational Databases'. It's often the case that large data sets are stored in one of these database platforms. Thus, if you're dealing with big data, you can choose between EG and SAS as your data access tool. Whether you're dealing with PC files or relational databases, the tools EG uses rival that of SAS.

VIEWING YOUR DATA

Once you define a project in Enterprise Guide, the next step is to add data to your project. When you open a SAS data set and add it to your project, EG automatically displays the data set in a table structure. Each project can contain many data sets, and you can switch between views of each one by clicking on that data set under the project tree. Figure 1.2, below, shows the display of a SAS data set in EG, once it's been added to a project.

	ID	Name	RestHR	MaxHR	RecHR	TimeMin	TimeSec	Tolerance	Year
1	2501	Bonaventure,	78	177	139	11	13	I	1999
2	2544	Jones, M	79	187	136	12	26	N	1999
3	2552	Roberson, P	68	158	139	15	41	D	1999
4	2568	Eberhardt, S	72	182	122	16	49	N	1999
5	2571	Nunnally, A	65	181	141	15	2	I	1999
6	2578	Cameron, L	75	158	108	14	27	I	1999
7	2579	Underwood, K	72	165	127	13	19	S	1999
8	2588	Ivan, H	70	182	126	15	41	N	1999
9	2595	Warren, C	77	170	136	12	10	S	1999

Figure 1.2

The table has properties similar to an Excel worksheet. By highlighting a column or row of the table, you're able to modify its layout. If you right-click you can hide or unhide a column, or adjust the column width. Similarly, you can hide or show a row, or adjust the height of a row of the table by right-clicking. In version 5.1 of EG, there's a new tool available called the Data Explorer, which allows you to examine your data before you add it to a project.

In EG, the data set is opened in read-only mode. EG creates a pointer to the data, but does not make a new copy of it. This is different than viewing your data in SAS. In SAS, to view a data set you're required to open a data set in the View Table window. If you want to have multiple data sets open simultaneously, each data set is opened in a separate View Table window. Traditionally in SAS, if you wanted to view a record by record listing of your data, you'd use the PRINT procedure.

II. DATA VALIDATION

Validating data is a project phase and responsibility of SAS programmers in many work environments. Once data has been collected, it's now necessary to prepare the data to get it in shape for analysis or reporting. The first step in data validation is to examine the data and to assess the condition it's in. Thus, you need, to know the variables in your data set and their attributes, and to examine data values for each variable. This involves producing frequency tables for character variables to detect invalid values, and running descriptive statistics for numeric variables to look at the range of values and outliers.



"Spock, where's that power you promised?"

In Enterprise Guide, there are several wizards used to perform these tasks. To output the contents of a SAS data set you would use the Data Set Attributes wizard. The One-way Frequencies wizard is used to produce frequencies using PROC FREQ. The Summary Statistics wizard produces descriptive statistics (Mean, Std dev.) for numeric variables, using PROC MEANS. Each of the wizards gives the option of creating a SAS data set from the output results, just like the SAS Procedure.

DATA MODIFICATION

Once you've examined your data, you can document what variables need to be edited. Then you can take the appropriate steps to modify your data. There are many titles used to describe this process; data cleaning, data scrubbing, quality control, recoding data, editing data. Table 1.2 below lists the tasks in SAS Enterprise Guide that are relevant to data manipulation, and the BASE SAS procedures that EG uses to perform these tasks.

SAS Task	BASE SAS Procedures Used
Append Table	SQL
Filter and Sort	
Query Builder	
Sort Data	SORT
Transpose	TRANSPOSE

Table 1.2

If you're using Enterprise Guide, and you wanted to recode variables on a SAS data set, you would use the Query Builder task wizard (see Figure 1.3 below). The Query Builder gives you the power to build and execute select queries. In order to build and run this query, EG harnesses the power of the SQL procedure. Just as if you're writing a PROC SQL Select statement, you select columns from an existing SAS data set, and compute new columns using the task wizard. If you want to recode a variable, the Query Builder uses the case-logic expression.

In Base SAS programming, the primary tool for manipulating data is the DATA step. The advantage of using Base SAS coding in data manipulation is that it provides efficient programming structures. Some of these structures are absent from EG's toolset. EG doesn't have wizards which write SAS code to define and reference ARRAY's, Do-End Groups, and Iterative Do-Loops.

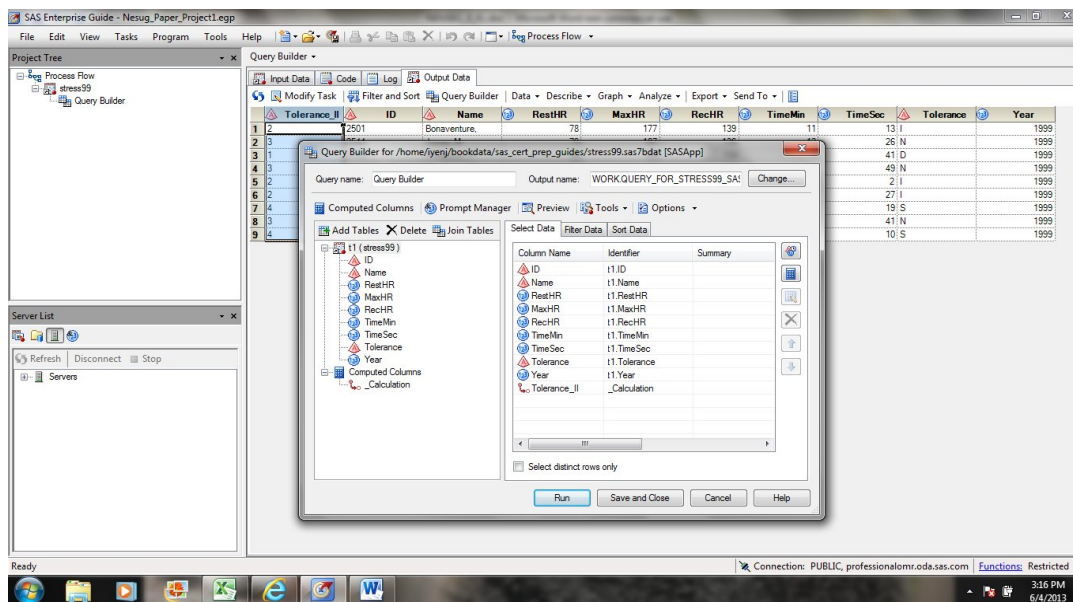


Figure 1.3

To recode variables using BASE SAS, you would probably code a series of If-Then-Else conditional logic in a DATA step. Although you also have the option of using Select-When constructs, or even use PROC SQL if you want. Practically, if you're working with disorganized, messy data, you're probably going to make a lot of edits or recodes to your data set. Going through the task wizard to define each condition to recode is cumbersome. Therefore, coding using DATA step programming is probably better suited for it.

COMBINING DATA SETS

Enterprise Guide has facilities for combining data sets both horizontally and vertically. EG uses the Query Builder wizard for combining data sets horizontally. The Query Builder wizard performs a join or merge by coding and executing a PROC SQL join query. The user defines the type of join that is desired, and the join condition/filter for the join. Available join types are Inner, Outer (Left Join, Right Join, Full Join), or Natural Join. Figure 1.4 below shows the Query Builder wizard steps to process a join.

When you click on 'join properties', you're taken to a window which illustrates the resulting set of data by choosing a join type, through the use of Venn diagrams. Venn diagrams facilitate an understanding of join methodology that is helpful instruction for those without SQL or relational database knowledge. This applies to traditional SAS DATA Step programmers as well as business analysts.

An advantage of using PROC SQL to combine your data is that you're not required to presort your data before merging. In SAS, if you code a DATA Step match-merge, each data set needs to be sorted on the by variable. This usually means you're required to run the Sort procedure on each data set, or use an index, unless your data is in default sorted order.

Because EG uses PROC SQL joins to combine data it has limitations about specific subsets of data it's able to produce. PROC SQL joins are able to output matching records between tables (Inner joins), or are able to produce sets of records found in one table or the other (Left Join or Right Join), but not non-matching records. Since a SQL join is unable to output unmatched records, (those found in one table, but not the other), EG doesn't support this capability. This is one advantage of using the DATA Step match merge to combine your data. Although, PROC SQL supports the EXCEPT SET operator which produces the exceptions from two queries.

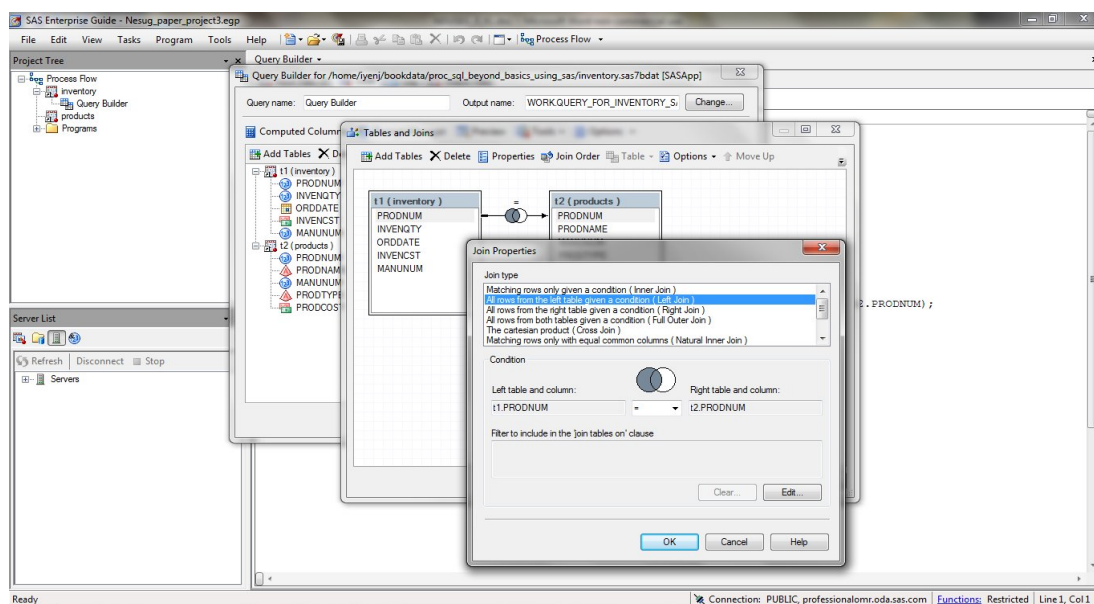


Figure 1.4

EG uses the Append Table wizard to combine data sets vertically. The Append Table wizard uses the PROC SQL outer union set operator to concatenate SAS data sets. Base SAS has multiple methods for appending data sets. A programmer has the choice of using PROC APPEND, the DATA Step Set statement, or PROC DATASETS. This is in addition to the PROC SQL set operators. Thus, Base SAS provides more options for combining data sets together than the Enterprise Guide task wizards.

III. REPORTING

SAS EG has several wizards for producing reports, and each wizard utilizes a different SAS reporting procedure. EG is capable of generating simple data listings, complex reports, and complex tables. It's capable of producing both detail and summary reports. Table 1.3 below lists the reporting task wizards in EG, and the SAS procedures used for each wizard.

SAS Task	BASE SAS Procedures Used
List Data	PRINT
List Report Wizard	REPORT
Summary Tables	TABULATE

Table 1.3

The List Data wizard provides options that access the full power of PROC PRINT. Included in this wizard are results sub-wizards which use the Output Delivery System (ODS) to set the format of the report. Just as in ODS, you select different destinations which are output formats for your report, such as HTML, PDF, or RTF (Rich Text Format).



"Let them eat static."

The List Report wizard is appropriate to produce complex reports. It uses the Report procedure to generate the report in the background. PROC REPORT is one of the more nuanced, extensive procedures in Base SAS. EG reduces the complexity of PROC REPORT through organizing report production into sections and user-friendly templates. The wizard lets you select the options which define the dimensions of the report layout.

Initially, you select variables to include as columns in the report. Then you choose variables to stratify your data. There are different stratification levels, for selecting a variable to group the rows of the report, or creating a separate table of the report based on each value of a categorical variable. As you select options, the wizard displays a preview of the report, so you see how its appearance is affected. Figure 1.5 shows the List Report wizard.

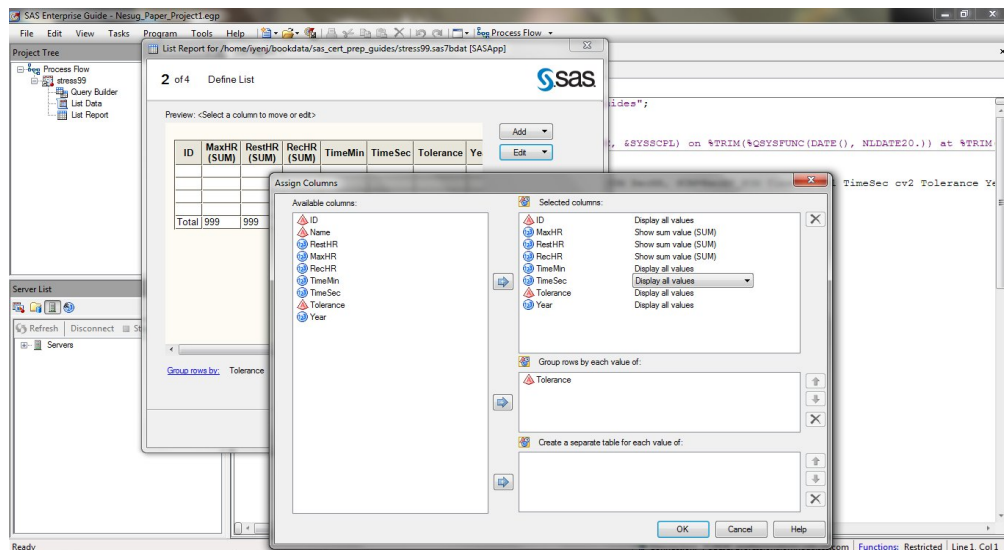


Figure 1.5

In terms of bells and whistles, the List Report wizard is essentially equivalent to the REPORT procedure in Base SAS. The only real drawback is that the wizard doesn't give you the ability to compute new report variables, using compute blocks. Base SAS supports using PROC REPORT output in conjunction with the Output Delivery System. However, using the EG List Report wizard, users aren't able to send report output to different destination formats with ODS. But, EG does provide the capability of exporting your report from the results window to an HTML or PDF file.

The reporting wizards are good for creating one-time or ad-hoc reports. But frequently it's the case there's a need for standard reports which can be automated and produced at regular intervals. EG has stored processes which are reusable SAS programs, which accept prompts that are macrovariables. However, stored processes aren't translatable into macros without substantial manual editing.

This is a case in point of an advantage SAS has over Enterprise Guide. EG doesn't have task wizards for defining and coding SAS Macros. You can write and submit Macro code in EG. But, if you save the code as a separate SAS program, it's no longer integrated into the EG project. I've found it to be effective to define a report template using EG, and then export the code that's generated to a SAS program. Then you can parameterize that code, automate your report, and put it into production.

IV. GRAPHICS

SAS Enterprise Guide uses SAS\Graph to generate standard and custom graphics. The commonly used graphs that Enterprise Guide supports include bar charts, line charts, bar-line charts, and pie charts. With the exception of some of the Statistical Graphics procedures (SG), all of the capabilities of SAS\Graph can be harnessed using Enterprise Guide. There are multiple wizards used for bar, line, and pie charts. Figure 1.6 shows the Bar Chart wizard in EG.

One of the challenges of developing graphs using SAS\Graph code is the technical and analytical components of graphs which need defining and declaration. Graph objects consist of x and y axes, axis labels, axis values, major tick marks, minor tick marks, data labels, and legends. Bar-line graphs have more than one y axis which adds another level of complexity. Producing a bar or line graph requires consulting technical documentation to lookup the detailed SAS\Graph code which corresponds with each of these options.

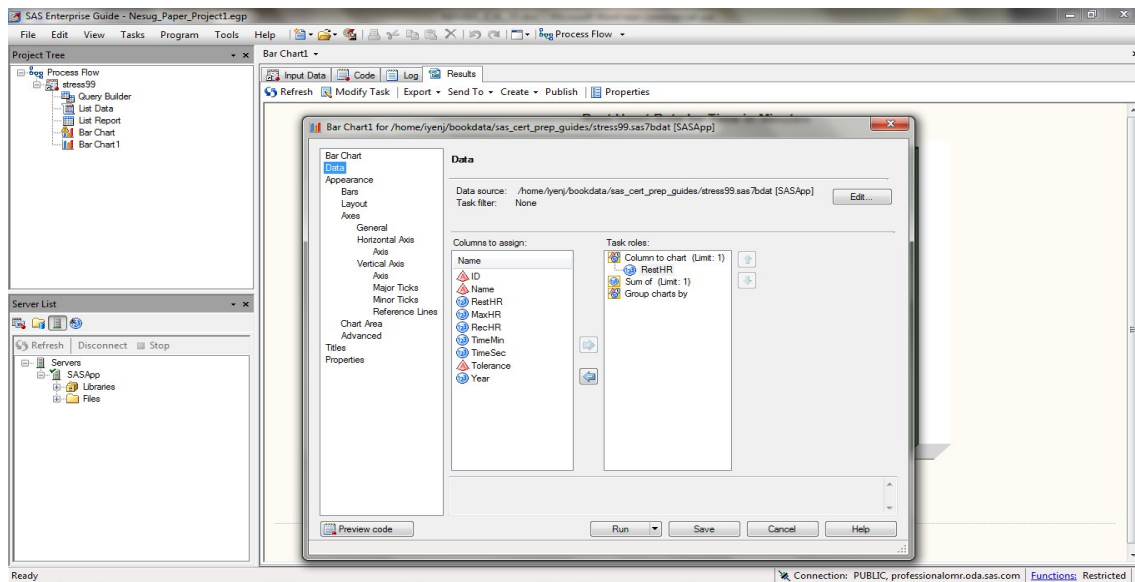


Figure 1.6

The wizards make tedious graph production an easy straight-forward exercise. Through the use of intuitive, straight-forward wizards, EG takes the technical hassle out of graph production. In the Bar Graph wizard, there are separate windows for each graph axis.

On each axis window, there are checkboxes with non-technical descriptions of that graph option, so that the user can understand the formatting options and apply the appropriate labels, values, and other objects to that axis. The windows provide a visual template of the graph which illustrates how the output and resulting graph changes with respect to selection of specific options.



"Fascinating, Captain!"

EG allows you to save the graph in several standard formats. You can export the graph to an HTML, XML, or PDF file. Of course, you can always export the SAS\Graph code that's generated for you, and save it as a SAS program (.SAS file) if you want to execute it in EG, or the editor in SAS Display Manager. If you choose to go this route, the graph will be opened in a graph results window, and can be saved as an image using the .png format. Table 1.4 below is an exhaustive list of the graph tasks in Enterprise Guide, the SAS\Graph procedure that's run in the background, and the module of SAS needed.

SAS Task	SAS Procedures Used	SAS Module
Area Plot	GPLOT	Base SAS, SAS/GRAPH
Bar Chart	GCHART	
Bar-Line Chart	GBARLINE	
Box Plot	GPLOT	
Bubble Plot	GPLOT	
Characterize Data	CONTENTS, UNIVARIATE, FREQ, SGPLOT	
Contour Plot	GCONTOUR, G3GRID	
Donut Chart	GCHART	
Line Plot	GPLOT	
Map Chart	GMAP	
Pie Chart	GCHART	
Radar Chart	GRADAR	
Scatter Plot	GPLOT, G3D	
Scatter Plot Matrix	SGSCATTER	
Surface Plot	G3D	
Tile Chart	GTILE	

Table 1.4

V. CONCLUSION

SAS® Enterprise Guide® (EG) is a valuable addition to standard SAS software, and has a strong suite of tools that enable it to perform the full range of data tasks. Navigating the data universe using EG requires paying attention to the specifics of the software's built-in processes. The effectiveness of using EG to execute a specific task depends on the project, the type of task, and the quality of the data. EG rivals traditional SAS for data access, reporting, and graphics project functions. Its limitations are in manipulating data, and automating processes and putting them into production.



"Live long, and prosper."

REFERENCES

Shapiro, Mira and Kirk Paul Lafler (2011), "Point-and-Click Programming Using SAS® Enterprise Guide®," Proceedings of the 2011 SAS Global Forum (SGF) Conference.

Hemedinger, Chris (2008), "Find Out What You're Missing: Programming with SAS® Enterprise Guide®" SAS Institute Inc. Cary, NC, March 2008.

SAS Institute Inc. "SAS Tasks and Associated SAS Procedures and Licenses". <http://support.sas.com/documentation/onlinedoc/guide/sastasksandprocs.htm>.

Miller, Oliver "50 Quotes from the 'Star Trek' movies, in order of awesomeness". <http://thoughtcatalog.com/2013/quotes-from-the-star-trek-movies-in-order-of-awesomeness>.

ACKNOWLEDGEMENTS

The author would like to thank Brian Varney and Misty Johnson, MWSUG 2014 SAS 101 Co-Chairs, for accepting my abstract and paper; Craig Wildeman, Operations Co-Chair, Cindy Lee, Academic Chair, and the MWSUG Conference Committee for organizing a great conference!

TRADEMARK CITATION

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 registered trademarks or trademarks of their respective companies.

Jay Iyengar is an independent consultant, trainer and SAS Certified Professional. He has been using SAS since 1997. His industry experience includes Government, Healthcare and Clinical/Pharmaceutical. He presented a paper at SESUG 2011, entitled 'Can You Decipher the code? If you can, maybe you can break it'. He attended last year's Wisconsin-Illinois SAS Users Group (WILSU) He has attended several conferences of the Northeast SAS Users Group (NESUG), and regularly attended meetings of the DC SAS Users Group (DCSUG).

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

Contact Information

Jay Iyengar
Independent Consultant
800 Hinman Ave.,
Evanston, IL 60202
Email: lyenJ@aol.com

LinkedIn: <http://www.linkedin.com/pub/jay-iyengar/31/897/577>