

Paper 289-2012

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®

Mira Shapiro, Analytic Designers LLC, Bethesda, MD

Kirk Paul Lafler, Software Intelligence Corporation, Spring Valley, CA

ABSTRACT

Importing Microsoft Excel files into SAS can often be a challenge. Perfectly formatted Excel files with labels in the first row and idiosyncrasy-free, clean data is not usually the norm. We will show how to overcome many of the obstacles associated with creating SAS data sets from Excel workbooks by using various combinations of SAS Enterprise Guide 4.3's features. The import wizard, generated code, code suggestion mechanism, options, and the ability to preview the first section of a CSV file will all be shown as mechanisms for creating analytic data sets from Excel input.

INTRODUCTION

Analytics professionals rely on working with reliable and clean data sources. Data is often delivered in Excel spreadsheet format. Many times, the spreadsheets were developed for another purpose, and sending the data to be included in an analytic database was not considered in the design or formatting strategy. SAS and SAS Enterprise Guide provide the capabilities to assess and overcome many of the common issues associated with reading Excel spreadsheets into a SAS Dataset.

In this discussion, we will look at two common issues that arise and examine how Enterprise Guide can be used to properly import your data to SAS.

DATA USED IN EXAMPLES

The data used in all the examples in this paper consist of a selection of movie classics. The Movies CSV file and Microsoft Excel file consist of six columns: title, length, category, year, studio, and rating. In addition there are versions of these files that include an extra row at the beginning and character data in what is intended to be a numeric field. Illustrations of the data follow:

MOVIES Microsoft Excel File and CSV File

	A	B	C	D	E	F
1	Title	Length	Category	Year	Studio	Rating
2	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
3	Casablanca	103	Drama	1942	MGM / UA	PG
4	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
5	Coming to America	116	Comedy	1988	Paramount Pictures	R
6	Dracula	130	Horror	1993	Columbia TriStar	R
7	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
8	Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
9	Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
10	Jaws	125	Action Adventure	1975	Universal Studios	PG
11	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
12	Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
13	Michael	106	Drama	1997	Warner Brothers	PG-13
14	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
15	Poltergeist	115	Horror			
16	Rocky	120	Action Adventure			
17	Scarface	170	Action Cops & Robber			
18	Silence of the Lambs	118	Drama Suspense			
19	Star Wars	124	Action Sci-Fi			
20	The Hunt for Red October	135	Action Adventure			
21	The Terminator	108	Action Sci-Fi			
22	The Wizard of Oz	101	Adventure			
23	Titanic	194	Drama Romance			

Title, Length, Category, Year, Studio, Rating
 Brave Heart, 177, Action
 Adventure, 1995, Paramount Pictures, R
 Casablanca, 103, Drama, 1942, MGM / UA, PG
 Christmas Vacation, 97, Comedy, 1989, Warner
 Brothers, PG-13

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

MOVIES with extra row Microsoft Excel File and CSV File

	A	B	C	D	E	F	G	H
2	SAS Enterprise Guide will help you to examine the data to determine where you can find the column headings and the data.							
3								
4	Title	Length	Category	Year	Studio	Rating		
5								
6	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R		
7	Casablanca	103	Drama	1942	MGM / UA	PG		
8	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13		
9	Coming to America	116	Comedy	1988	Paramount Pictures	R		
10	Dracula	130	Horror	1993	Columbia	R		
11	Dressed to Kill	105	Drama Mysteries	1980	Filmways	R		
12	Forrest Gump	142	Drama	1994	Paramount	PG-13		
13	Ghost	127	Drama Romance	1990	Paramount	PG-13		
14	Jaws	125	Action Adventure	1975	Universal	PG		
15	Jurassic Park	127	Action	1993	Universal	PG-13		
16	Lethal Weapon	110	Action Cops & Robber	1987	Warner Br	R		
17	Michael	106	Drama	1997	Warner Br	PG-13		
18	National Lampoon's Vacation	98	Comedy	1983	Warner Br	PG-13		
19	Poltergeist	115	Horror	1982	MGM / UA	PG		
20	Rocky	120	Action Adventure	1976	MGM / UA	PG		
21	Scarface	170	Action Cops & Robber	1983	Universal	R		
22	Silence of the Lambs	118	Drama Suspense	1991	Orion	R		
23	Star Wars	124	Action Sci-Fi	1977	Lucas Film	PG		
24	The Hunt for Red October	135	Action Adventure	1989	Paramount	PG		
25	The Terminator	108	Action Sci-Fi	1984	Live Enter	R		
26	The Wizard of Oz	101	Adventure	1939	MGM / UA	G		
27	Titanic	194	Drama Romance	1997	Paramount	PG-13		

Sometimes people add some comments at the top of the spreadsheet and you need to figure out where your titles and data start,,,,,
SAS Enterprise Guide will help you to examine the data to determine where you can find the column headings and the data,,,,,
,,,,,
Title,Length,Category,Year,Studio,Rating
,,,,,
Brave Heart,177,Action
Adventure,1995,Paramount Pictures,R
Casablanca,103,Drama,1942,MGM / UA,PG
Christmas Vacation,97,Comedy,1989,Warner
Brothers,PG-13

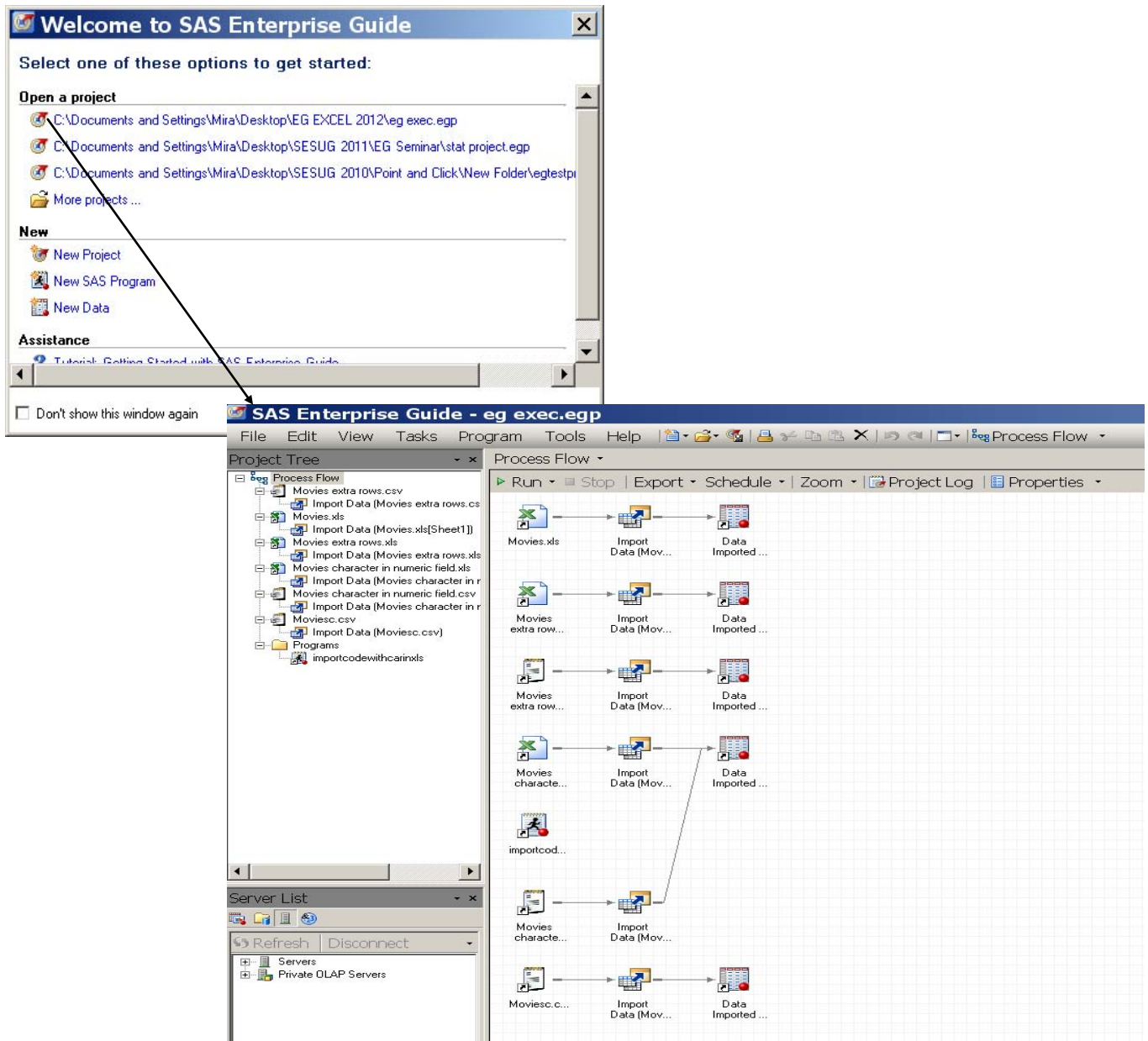
MOVIES character in numeric field Microsoft Excel File

	A	B	C	D	E	F
1	Title	Length	Category	Year	Studio	Rating
2	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
3	Casablanca	103	Drama	1942	MGM / UA	PG
4	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
5	Coming to America	116	Comedy	1988	Paramount Pictures	R
6	Dracula	130	Horror	1993	Columbia TriStar	R
7	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
8	Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
9	Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
10	Jaws	125	Action Adventure	1975	Universal Studios	PG
11	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
12	Lethal Weapon	x23	Action Cops & Robber	1987	Warner Brothers	R
13	Michael	106	Drama	1997	Warner Brothers	PG-13
14	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
15	Poltergeist	115	Horror	1982	MGM / UA	PG
16	Rocky	120	Action Adventure	1976	MGM / UA	PG
17	Scarface	170	Action Cops & Robber	1983	Universal Studios	R
18	Silence of the Lambs	118	Drama Suspense	1991	Orion	R
19	Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
20	The Hunt for Red October	135	Action Adventure	1989	Paramount Pictures	PG
21	The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R
22	The Wizard of Oz	101	Adventure	1939	MGM / UA	G
23	Titanic	194	Drama Romance	1997	Paramount Pictures	PG-13

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

OPENING ENTERPRISE GUIDE AND YOUR PROJECT

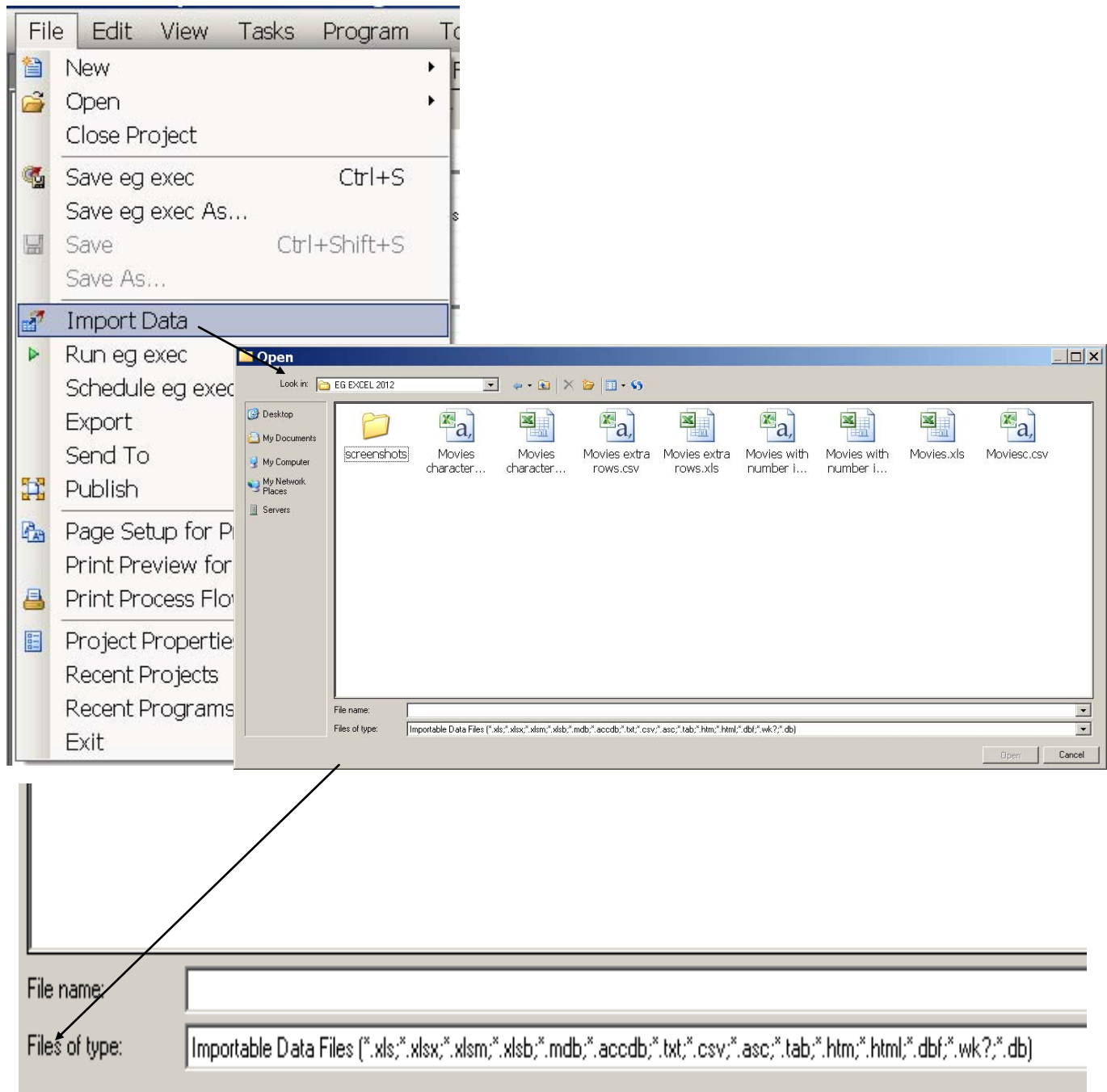
In order to begin, you need to open Enterprise Guide. When the “Welcome to SAS Enterprise Guide” window appears, you can choose to create a new project or select an existing project. Here we are choosing an existing project.



Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

IMPORTING YOUR DATA

In order to import your data, choose **File** and then **Import Data** from the tool bar. Then use the **Open** dialog box to navigate to the file of interest. Note that there is a list of importable file types in the **Files of type:** section of the **Open** dialog box. These file types include Excel, text and database types.



Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

IMPORTING A “PERFECTLY FORMATTED” EXCEL SPREADSHEET

Once you have navigated to the file of interest and clicked **Open**, the Enterprise Guide **Import Data** wizard will guide you through the import process. The first step allows you to verify the data input source and choose your output source. Note that unless you change the destination library, that the resulting SAS data set is placed in the work library and is temporary and will not be saved after closing Enterprise Guide. You may change the Output destination to another permanent SAS library. If you choose to have the resulting data set reside only in a temporary library, it will be recreated the next time you run the program in the saved Enterprise Guide Project.

Step 1 of the Import Process

Specify the output library

The data set name defaults to the name of the incoming data set.

Click **Next** to continue to use the wizard to apply options to the import process or click **Finish** to complete the import.

Step 2 of the Import Process

Step 2 of the **Import Data** process provides multiple options for manipulating the incoming Excel spreadsheet. You may indicate whether the first row contains field names, and you may choose to rename the columns to comply with SAS naming conventions. If the incoming field names include spaces, it is recommended that you choose **Rename the columns to comply with SAS naming conventions** so that they can be directly used in SAS programming statements. In this step you may also choose to import only a section of the spreadsheet, either by name or by cell ranges.

Check to use the first row of the spreadsheet for variable names.

Rename the columns using SAS naming conventions. Important: if your column names have imbedded blanks. This facilitates SAS programming

This section gives the capability to select a range of cells either by name or by cell range specification.

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

Step 3 of the Import Process

Step 3 of the **Import Data** process allows the user to examine and change attributes. Modifications can be made to the Name, Label, Type and Informats and Formats. Since in this example we are working with our “perfectly formatted” spreadsheet, we will not make any changes.

Incl	Source Name	Name	Label	Type	Source Informat	Len	Output Format	Output Informat
<input checked="" type="checkbox"/>	Title	title	Title	String	\$CHAR27.	27	\$CHAR27.	\$CHAR27.
<input checked="" type="checkbox"/>	Length	length	Length	Number	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	Category	category	Category	String	\$CHAR20.	20	\$CHAR20.	\$CHAR20.
<input checked="" type="checkbox"/>	Year	year	Year	Number	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	Studio	studio	Studio	String	\$CHAR18.	18	\$CHAR18.	\$CHAR18.
<input checked="" type="checkbox"/>	Rating	rating	Rating	String	\$CHAR5.	5	\$CHAR5.	\$CHAR5.

☒ Include field in output data set

Name:

Label:

Type:

Source attributes

Source informat:

Output attributes

Length:

Input format:

Output format:

OK Cancel

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

Step 4 of the Import Process

Step 4 of the **Import Data** process provides three important options. The first enables embedding the data in the generated SAS code. It is entirely up to the user whether they will find this useful. Secondly, if your SAS license includes SAS/ACCESS Interface to PC Files, Enterprise Guide will use this capability to enhance and speed up the import process. The additional option will allow SAS Enterprise Guide to “clean up” the text files.

Option to embed data in the generated SAS code

If you have a SAS license for SAS/ACCESS Interface to PC, Enterprise Guide will use this function to import files.

Remove problematic characters in text-based files

Click **<Back** to change some of your selected functions or **Finish** to complete the import.

The data was correctly imported into SAS Enterprise Guide.

Import Data (Movies.xls[Sheet1])

	Title	Length	Category	Year	Studio	Rating
1	Brave Heart	177	Action Adventure	1995	Paramount Pictu...	R
2	Casablanca	103	Drama	1942	MGM / UA	PG
3	Christmas Vacati...	97	Comedy	1989	Warner Brothers	PG-13
4	Coming to Ameri...	116	Comedy	1988	Paramount Pictu...	R
5	Dracula	130	Horror	1993	Columbia TriStar	R
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
7	Forrest Gump	142	Drama	1994	Paramount Pictu...	PG-13
8	Ghost	127	Drama Romance	1990	Paramount Pictu...	PG-13
9	Jaws	125	Action Adventure	1975	Universal Studios	PG
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
11	Lethal Weapon	110	Action Cops & R...	1987	Warner Brothers	R
12	Michael	106	Drama	1997	Warner Brothers	PG-13
13	National Lampoo...	98	Comedy	1983	Warner Brothers	PG-13
14	Poltergeist	115	Horror	1982	MGM / UA	PG
15	Rocky	120	Action Adventure	1976	MGM / UA	PG
16	Scarface	170	Action Cops & R...	1983	Universal Studios	R
17	Silence of the La...	118	Drama Suspense	1991	Orion	R
18	Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
19	The Hunt for Red...	135	Action Adventure	1989	Paramount Pictu...	PG
20	The Terminator	108	Action Sci-Fi	1984	Live Entertainme...	R
21	The Wizard of Oz	101	Adventure	1939	MGM / UA	G
22	Titanic	194	Drama Romance	1997	Paramount Pictu...	PG-13

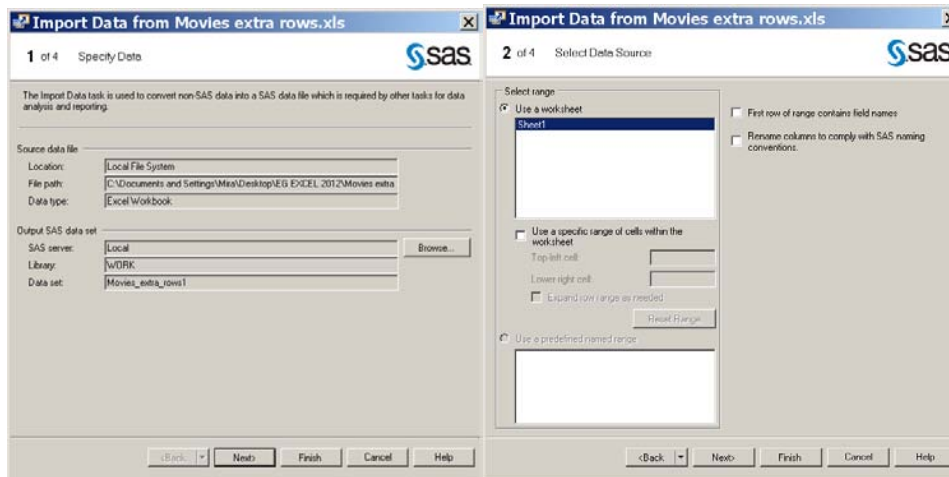
Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

IMPORTING AN EXCEL SPREADSHEET WITH EXTRA ROWS

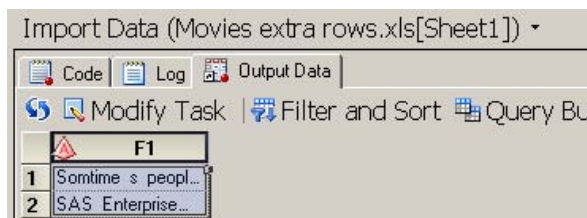
The file, Movies extra row.xls , contains blank lines and some extra lines of text at the top. Unlike the “perfectly formatted” Excel file, we will need to examine the file to understand the best way to import the data into SAS. With a small amount of pre-processing, the Enterprise Guide **Import Data** wizard guides the user through the process without the need for programming.

What doesn't work and Why Not!

For a first try we navigate to the Excel spreadsheet with the extra rows. Since we know the first row doesn't have field names, we leave that box unchecked and click the **Finish** button and examine the results.



The result of this process is the following SAS data set, which is clearly not what is expected or a correct representation of the Movies data set.



In order to understand what happened, we will take a look at the code that SAS Enterprise Guide created. It is clear from examining the code, even for someone who is not an experienced programmer, that somehow the text in the file, was interpreted as data and we did not properly import our Movies data set.

```
DATA WORK.Movies_extra_rows1;
  LENGTH
    F1 $ 124
  FORMAT
    F1 $CHAR124. ;
  INFORMAT
    F1 $CHAR124. ;
  INFILE 'C:\Documents and Settings\Mira\Local
Settings\Temp\SEG4020\Movies_extra_rows-
e6b3a708621943489f6c343395b6c302.txt'
    LRECL=124
    ENCODING="WLATIN1"
    TERMSTR=CR LF
    DLM='7F'x
    MISSOVER
    DSD ;
  INPUT
```

THIS CODE DOES NOT ACHIEVE THE DESIRED RESULTS!

The text at the top of the file is seen as the only variable "F1" and is determined to be character and 124 characters in length.

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

What WORKS to import the Movies Data Set with extra rows, and Why!

Movies extra row.xls was opened in Excel and saved as a comma separated value (CSV) file, resulting in the file illustrated above in the Data Used section. Now we will go back to the FILE->IMPORT and point to the Movies extra row.csv data set and proceed.

After clicking the **Next** button the **Import Data** wizard displays a sample of the file and provides options for directing how the data is imported. Looking at the display we can see that the field names appear in line 4 and the data begins in line 6, so we fill in these values in the associated boxes. In this case we won't limit the number of records read, but that is an option in this dialog.

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

Code created by Enterprise Guide to import the Movies extra rows.csv file.

```
DATA WORK.Movies_extra_rows;
  LENGTH
    title           $ 27
    length          8
    category        $ 20
    year            8
    studio          $ 18
    rating          $ 5 ;
  LABEL
    title           = "Title"
    length          = "Length"
    category        = "Category"
    year            = "Year"
    studio          = "Studio"
    rating          = "Rating" ;
  FORMAT
    title           $CHAR27.
    length          BEST3.
    category        $CHAR20.
    year            BEST4.
    studio          $CHAR18.
    rating          $CHAR5. ;
  INFORMAT
    title           $CHAR27.
    length          BEST3.
    category        $CHAR20.
    year            BEST4.
    studio          $CHAR18.
    rating          $CHAR5. ;
  INFILE
    'C:\Documents and Settings\Mira\Local Settings\Temp\
SEG3080\Movies_extra_rows-fe04b331a3b748219dc38d9689d2da57.txt'
    LRECL=72
    ENCODING="WLATIN1"
    TERMSTR=CRLF
    DLM='7F'x
    MISSEVER
    DSD ;
  INPUT
    title           : $CHAR27.
    length          : ?? BEST3.
    category        : $CHAR20.
    year            : ?? BEST4.
    studio          : $CHAR18.
    rating          : $CHAR5. ;
RUN;
```

Our Resulting Correct Data Set!

This time, using the CSV file and pointing to the correct lines for the column headings and data, we have properly imported our Movies data set, avoiding the blank lines and extra text.

Import Data (Movies extra rows.csv) ▾

Code Log Output Data

Modify Task Filter and Sort Query Builder Data Describe Graph A

	title	length	category	year	studio	rating
1	Brave Heart	177	Action Adventure	1995	Paramount Pictu...	R
2	Casablanca	103	Drama	1942	MGM / UA	PG
3	Christmas Vacati...	97	Comedy	1989	Warner Brothers	PG-13
4	Coming to Ameri...	116	Comedy	1988	Paramount Pictu...	R
5	Dracula	130	Horror	1993	Columbia TriStar	R
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
7	Forrest Gump	142	Drama	1994	Paramount Pictu...	PG-13
8	Ghost	127	Drama Romance	1990	Paramount Pictu...	PG-13
9	Jaws	125	Action Adventure	1975	Universal Studios	PG
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
11	Lethal Weapon	110	Action Cops & R...	1987	Warner Brothers	R

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

IMPORTING AN EXCEL SPREADSHEET WITH A CHARACTER IN A NUMERIC FIELD

More often than not, we are presented with spreadsheets that contain incorrect characters in a numeric field. The good news is that these kinds of errors are easy to spot. The bad news is that the field then gets converted to character rather than numeric upon import. Once the error is spotted there are multiple options to address this problem. In this example, it is assumed that we want to encode the value as missing and maintain the numeric format in the column.

Enterprise Guide Makes This Easy!

We begin the import process as in the previous two examples.

The first screenshot shows the 'Specify Data' step (1 of 4) of the 'Import Data from Movies character in numeric field....' wizard. It includes fields for 'Source data file' (Local File System), 'Location' (C:\Documents and Settings\Mira\Desktop\EG EXCEL 2012\Movies char), 'File path' (C:\Documents and Settings\Mira\Desktop\EG EXCEL 2012\Movies char), 'Data type' (Excel Workbook), 'Output SAS data set' (Local), 'SAS server' (Local), 'Library' (WORK), and 'Data set' (Movies_character_in_numeric_fie1). The second screenshot shows the 'Select Data Source' step (2 of 4). It has options to 'Use a worksheet' (selected, showing Sheet1), 'Use a specific range of cells within the worksheet' (with top-left and lower-right cell fields), and 'Use a predefined named range'. Checkboxes for 'First row of range contains field names' and 'Rename columns to comply with SAS naming conventions' are present. Navigation buttons at the bottom include <Back, Next>, Finish, Cancel, and Help.

If we had clicked the **Finish** button after selecting the data source, we would have obtained the result below. Notice that the length in line 11 contains an "x", which results in the **Length** field being defined as character. Rather than trying to manipulate the data after the import to correct this problem, we will attempt to correct it during the import by either continuing on to step 3 or, if we had already imported the data, by modifying the **Import Task**.

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

Import Data (Movies character in numeric field.xls[Sheet1]) ▾

Code Log Output Data

Modify Task Filter and Sort Query Builder Data Describe Graph And

	Title	Length	Category	Year	Studio	Rating
1	Brave Heart	177	Action Adventure	1995	Paramount Pictu...	R
2	Casablanca	103	Drama	1942	MGM / UA	PG
3	Christmas Vacati...	97	Comedy	1989	Warner Brothers	PG-13
4	Coming to Ameri...	116	Comedy	1988	Paramount Pictu...	R
5	Dracula	130	Horror	1993	Columbia TriStar	R
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
7	Forrest Gump	142	Drama	1994	Paramount Pictu...	PG-13
8	Ghost	127	Drama Romance	1990	Paramount Pictu...	PG-13
9	Jaws	125	Action Adventure	1975	Universal Studios	PG
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
11	Lethal Weapon	x23	Action Cops & R...	1987	Warner Brothers	R
12	Michael	106	Drama	1997	Warner Brothers	PG-13
13	National Lampoo...	98	Comedy	1983	Warner Brothers	PG-13
14	Pollergeist	115	Horror	1982	MGM / UA	PG

On step 3 of the **Import Task** the user is given the opportunity to view and/or modify the **Field Attributes**. As a result of the "x" in the **Length** column on line 11, the Type for **Length** is "string". Ideally we would like this column to be "numeric" and the value in the field to be coded as missing (.). In the Type designation for Length, we use the pull-down menu and change the type to Number.

Import Data from Movies character in numeric field.... X

3 of 4 Define Field Attributes

Select columns and define attributes:

Inc	Source Name	Name	Label	Type	Source Informat	Len.	Output Format	Output Informat
<input checked="" type="checkbox"/>	Title	Title	Title	String	\$CHAR27.	27	\$CHAR27.	\$CHAR27.
<input checked="" type="checkbox"/>	Length	Length	Length	String	\$CHAR3.	3	\$CHAR3.	\$CHAR3.
<input checked="" type="checkbox"/>	Category	Category	Category	String	\$CHAR20.	20	\$CHAR20.	\$CHAR20.
<input checked="" type="checkbox"/>	Year	Year	Year	Number	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	Studio	Studio	Studio	Currency	\$CHAR18.	18	\$CHAR18.	\$CHAR18.
<input checked="" type="checkbox"/>	Rating	Rating	Rating	Date	\$CHAR5.	5	\$CHAR5.	\$CHAR5.

Select All Clear All Modify...

<Back Next> Finish Cancel Help

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

Once the type is changed to Number, the user then has the choice to change the Informat and Formats. In this example we will use the defaults of BEST3. and BEST12.

3 of 4 Define Field Attributes

Select columns and define attributes:

Inc	Source Name	Name	Label	Type	Source Informat	Len.	Output Format	Output Informat
<input checked="" type="checkbox"/>	Title	Title	Title	String	\$CHAR27.	27	\$CHAR27.	\$CHAR27.
<input checked="" type="checkbox"/>	Length	Length	Length	Number	BEST3.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	Category	Category	Category	String	\$CHAR20.	20	\$CHAR20.	\$CHAR20.
<input checked="" type="checkbox"/>	Year	Year	Year	Number	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	Studio	Studio	Studio	String	\$CHAR18.	18	\$CHAR18.	\$CHAR18.
<input checked="" type="checkbox"/>	Rating	Rating	Rating	String	\$CHAR5.	5	\$CHAR5.	\$CHAR5.

Select All Clear All Modify...

< Back Next> Finish Cancel Help

Now when the file is imported, the **Length** field is numeric, and the value in the 11th observation is set to missing.

Code Log Output Data							
Modify Task Filter and Sort Query Builder Data Describe Graph Ar							
	Title	Length	Category	Year	Studio	Rating	
1	Brave Heart	177	Action Adventure	1995	Paramount Pictu...	R	
2	Casablanca	103	Drama	1942	MGM / UA	PG	
3	Christmas Vacati...	97	Comedy	1989	Warner Brothers	PG-13	
4	Coming to Ameri...	116	Comedy	1988	Paramount Pictu...	R	
5	Dracula	130	Horror	1993	Columbia TriStar	R	
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R	
7	Forrest Gump	142	Drama	1994	Paramount Pictu...	PG-13	
8	Ghost	127	Drama Romance	1990	Paramount Pictu...	PG-13	
9	Jaws	125	Action Adventure	1975	Universal Studios	PG	
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13	
11	Lethal Weapon	.	Action Cops & R...	1987	Warner Brothers	R	
12	Michael	106	Drama	1997	Warner Brothers	PG-13	
13	National Lampoo...	98	Comedy	1983	Warner Brothers	PG-13	
14	Poltergeist	115	Horror	1982	MGM / UA	PG	
15	Rocky	120	Action Adventure	1976	MGM / UA	PG	
16	Scarface	170	Action Cops & R...	1983	Universal Studios	R	

Using SAS® Enterprise Guide® to Coax Your Excel Data In To SAS®, continued

The code generated by Enterprise Guide to read in the Movies data set with the character in a numeric field is shown below. The original generated code prior to the change to the field attributes for **Length** is shown with the lines that Enterprise Guide **changed highlighted** in yellow and the modifications shown to the right.

```
DATA WORK.Movies_character_in_numeric_fiel;
  LENGTH
    Title           $ 27
    Length          $ 3 -----> Length 8
    Category        $ 20
    Year            8
    Studio          $ 18
    Rating          $ 5 ;
  FORMAT
    Title           $CHAR27.
    Length          $CHAR3. -----> Length BEST12.
    Category        $CHAR20.
    Year            BEST12.
    Studio          $CHAR18.
    Rating          $CHAR5. ;
  INFORMAT
    Title           $CHAR27.
    Length          $CHAR3. -----> Length BEST12.
    Category        $CHAR20.
    Year            BEST12.
    Studio          $CHAR18.
    Rating          $CHAR5. ;
  INFILE 'C:\Documents and Settings\Mira\Local Settings\Temp\
SEG4020\Movies character in numeric field-03c43d06f1ba4bf6aa0f081a96be5d38.txt'
  LRECL=72
  ENCODING="WLATIN1"
  TERMSTR=CRLF
  DLM='7F'x
  MISSOVER
  DSD ;
  INPUT
    Title           : $CHAR27.
    Length          : $CHAR3. -----> Length : BEST3.
    Category        : $CHAR20.
    Year            : BEST32.
    Studio          : $CHAR18.
    Rating          : $CHAR5. ;
RUN;
```

CONCLUSION

In this discussion we have seen how Enterprise Guide helps the user overcome some common problems associated with importing Excel spreadsheets. By providing wizards and the ability to manipulate the field characteristics, Enterprise Guide provides the user with the tools to create analytic data sets without writing SAS code from scratch.

TRADEMARK CITATIONS

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.

AUTHOR BIOS

Mira Shapiro has been a SAS user since 1979 and is currently serving as SAS User Liaison for DC - SUG, a Washington-DC SAS Users Group. She has used SAS throughout her career as a Capacity Planner, Consultant and Biostatistician. She holds a BA in Statistics / Computer Science and an MS in Public Health / Biostatistics and works on analytics and pre-sales projects across multiple industries.

Kirk Paul Lafler is consultant and founder of Software Intelligence Corporation and has been using SAS since 1979. He is a SAS Certified Professional, provider of IT consulting services, trainer to SAS users around the world, and sasCommunity.org Advisory Board member. As the author of four books including PROC SQL: Beyond the Basics Using SAS, Kirk has written more than four hundred peer-reviewed papers, been an Invited speaker and trainer at more than four hundred SAS International, regional, local, and special-interest user group conferences and meetings throughout North America, and is the recipient of 17 "Best" contributed paper awards. His popular SAS Tips column, "Kirk's Korner of Quick and Simple Tips", appears regularly in several SAS User Group newsletters and Web sites, and his fun-filled SASword Puzzles is featured in SAScommunity.org.

REFERENCES

- Delwiche, Lora D. and Susan J. Slaughter (2006), "*Producing Summary Tables in SAS® Enterprise Guide®*," Proceedings of the 2006 NorthEast SAS Users Group (NESUG) Conference
- Hemedinger, Chris (2005), "*Boost Your Programming Productivity with SAS® Enterprise Guide®*," Proceedings of the Thirtieth SAS Users Group International (SUGI) Conference, SAS Institute Inc., Cary, NC, USA.
- Hettinger, Patricia (2009), "*Tips for Moving SAS® Enterprise Guide® on Unix*," Proceedings of the 2009 SouthEast SAS Users Group (SESUG) Conference.
- Lafler, Kirk Paul (2004), "*Creating HTML Output with Output Delivery System*," Proceedings of the 2004 Western Users of SAS Software (WUSS) Conference, Software Intelligence Corporation, Spring Valley, CA, USA.
- SAS® Software Essentials Using SAS Enterprise Guide® Course Notes, First Edition (2009)*. Software Intelligence Corporation, Spring Valley, CA, USA.
- Delwiche, Lora D. and Susan J. Slaughter (2006), "*Producing Summary Tables in SAS® Enterprise Guide®*," Proceedings of the 2006 North East SAS Users Group (NESUG) Conference.
- Shapiro, Mira and Kirk Paul Lafler (2010), "Point-and-Click Programming Using *SAS® Enterprise Guide®*," Proceedings of the 2010 MidWest SAS Users Group (MWSUG) Conference.
- Shapiro, Mira and Kirk Paul Lafler (2010), "Point-and-Click Programming Using *SAS® Enterprise Guide®*," Proceedings of the 2010 Western Users of SAS Software (WUSS) Conference.
- Shapiro, Mira and Kirk Paul Lafler (2010), "Point-and-Click Programming Using *SAS® Enterprise Guide®*," Proceedings of the 2010 South East SAS Users Group (SESUG) Conference.
- Todd, Michael (2008), "*Transitioning to SAS® Enterprise Guide®*," Proceedings of the 2008 NorthEast SAS Users Group (NESUG) Conference.

CONTACT INFORMATION

Mira Shapiro

E-mail: Mira.Shapiro@analyticdesigners.com

Kirk Paul Lafler

E-mail: KirkLafler@cs.com