Please login…

- Take a seat at one of the work stations
- Login with your HawkID
- Locate SAS 9.3 in the Start Menu
  - Start / All Programs / SAS / SAS 9.3 (64 bit)
- Make SAS "go"
  - Raise your hand if you need assistance

Day 2 of SAS® Summer Institute 2015

will begin soon…

Procedures for Data Insight

University of Iowa SAS User Group

SAS Summer Institute

August 19, 2015

Presented by: Fred Ullrich, ASG
Department of Health Management and Policy
College of Public Health

2nd day morning session (8:07 – 11:34)

overview

“Procedures for Data Insight”

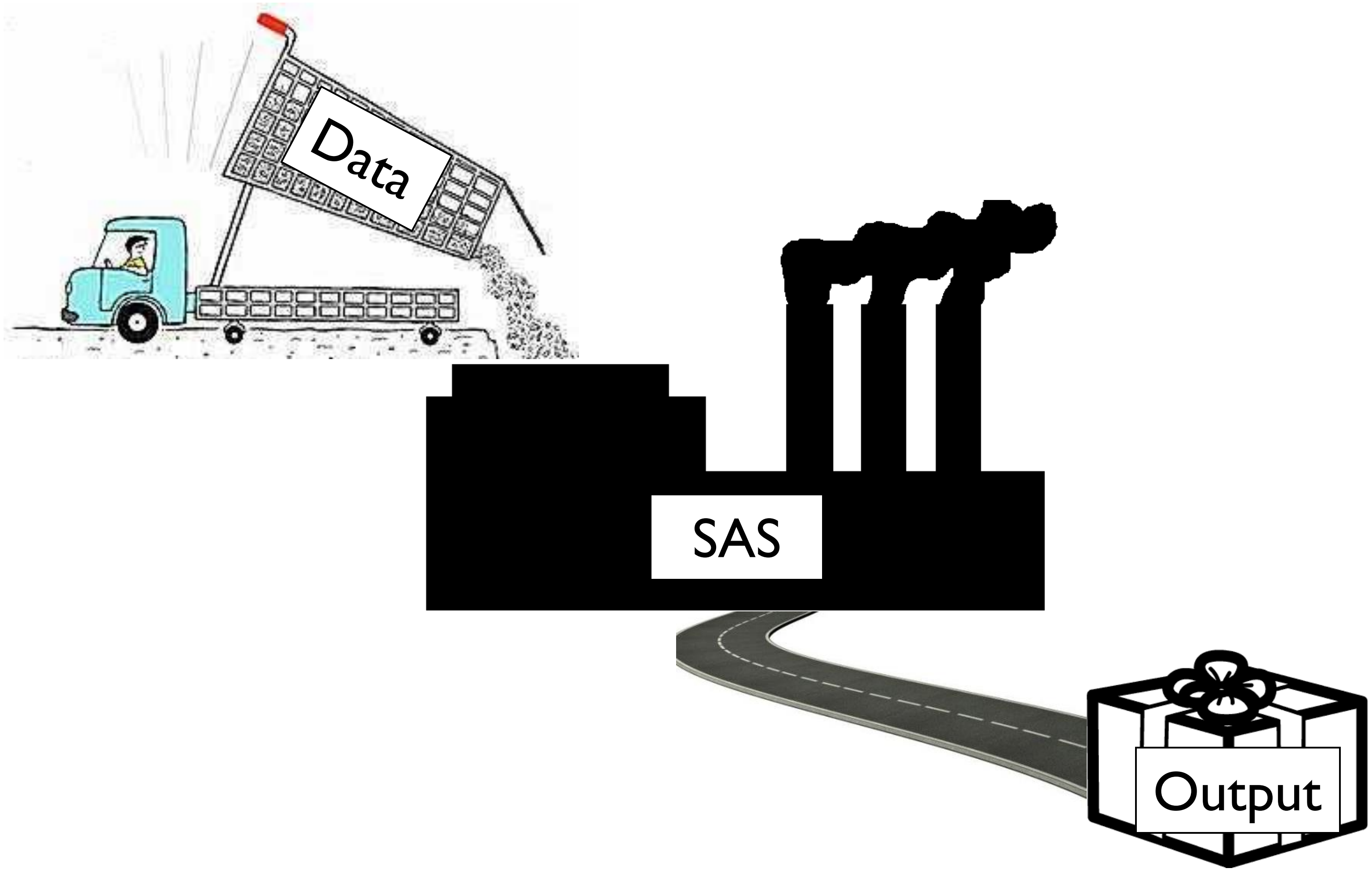
Part 1:

- SAS program and data review
- PROC Overview
- Some PROC specifics

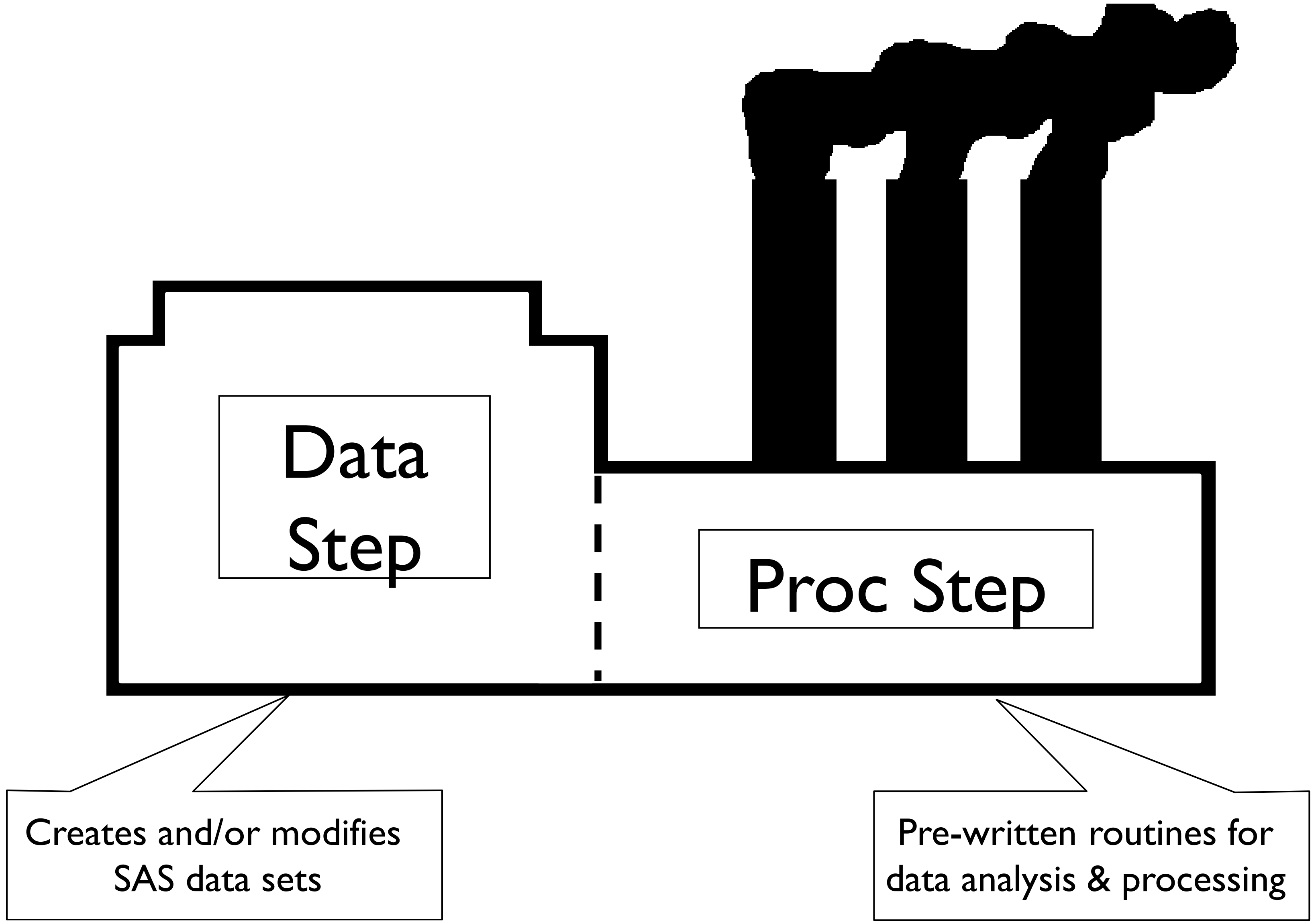
Part 2:

- Elements of style
- More PROC specifics
- Intro to data import/export

SAS Program and Data Review, (1)



SAS Program and Data Review, (2)


Data Step
Proc Step
Creates and/or modifies SAS data sets
Pre-written routines for data analysis & processing

SAS Program and Data Review, (3)

Sample Data Steps

“inline” data

```
DATA wages;
INPUT id  days wages;
totwage=days*wages;
DATALINES;
101    55   165.10
156    35   132.56
204   125   115.89
245    78   155.25
397    32   112.90
456    44   118.21
678    67   156.20
875    95   134.00
941    88   122.45
;
run;
```

“external” data

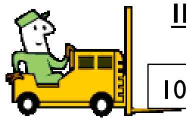
```
data trial1;
infile 'C:\wagedata.txt';
input id days wages;
totwage=days*wages;
run;
```

c:\wagedata.txt

```
101    55   165.10
156    35   132.56
204   125   115.89
245    78   155.25
397    32   112.90
456    44   118.21
678    67   156.20
875    95   134.00
941    88   122.45
```

SAS Program and Data Review, (4)

Sample Data Steps

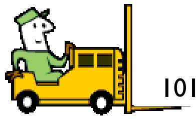


input data

101	55	165.10
156	35	132.56
204	125	115.89
245	78	155.25
397	32	112.90
456	44	118.21
678	67	156.20
875	95	134.00
941	88	122.45

SAS Program

```
data trial1;
infile 'C:\wagedata.txt';
input id days wages;
totwage=days*wages;
run;
```



output data

```
101    55    165.10    9080.5
```

SAS Program and Data Review, (5)

SAS Log Window

```
1     data wages;
2     infile 'C:\wagedata.txt';
3     input id days wages;
4     totwage=days*wages;
5     run;

NOTE:       9 records were read from the infile 'C:\junk\wagedata.txt'.
            The minimum record length was 18.
            The maximum record length was 18.
NOTE:       The data set WORK.WAGES has 9 observations and 4 variables.
NOTE:       DATA statement used (Total process time):
            real time           0.00 seconds
            cpu time            0.00 seconds
```

- “WORK” is the name of a library
- Libraries can contain many datasets
- Datasets in the “WORK” library are deleted at the end of your SAS session (i.e. they are temporary)
- To make your datasets “permanent” use a different library name
- Before you can do that, the library (name) must be associated with a directory (folder) accessible from your computer

SAS Program and Data Review, (6)

$$\underbrace{\text{LIBNAME}}_{\textit{LIBNAME statement}}\ \overbrace{\text{mylib}}^{\textit{name you give the library}}\ \underbrace{\text{‘C:\textbackslash Research\textbackslash Data\textbackslash’}}_{\textit{location of the library}};$$

Sample Program:
<u>Create permanent dataset</u>

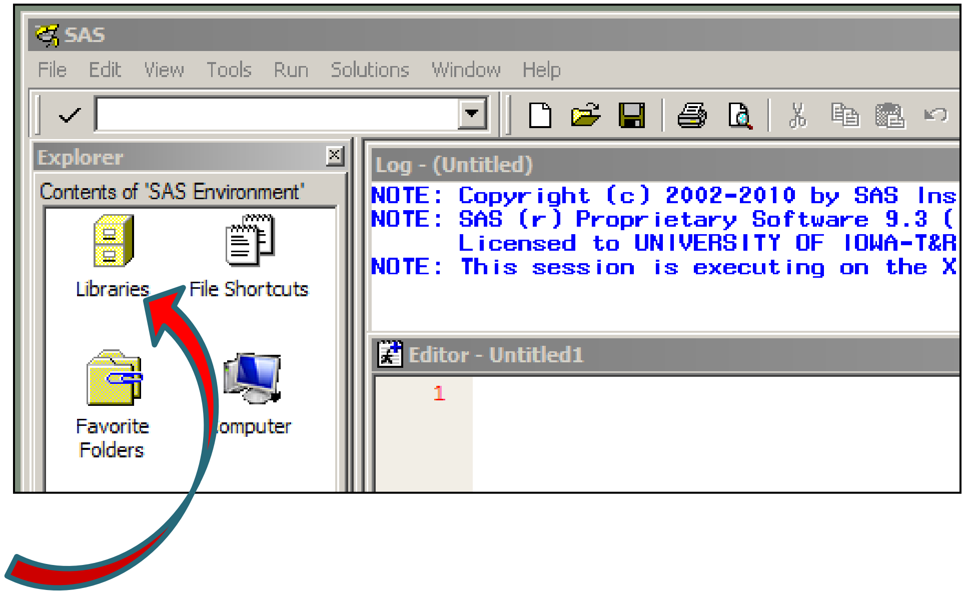
```
libname mylib ‘c:\research\data\’;
data mylib.trial1;
infile 'C:\wagedata.txt';
input id days wages;
run;
```

Sample Program:
<u>Use permanent dataset</u>

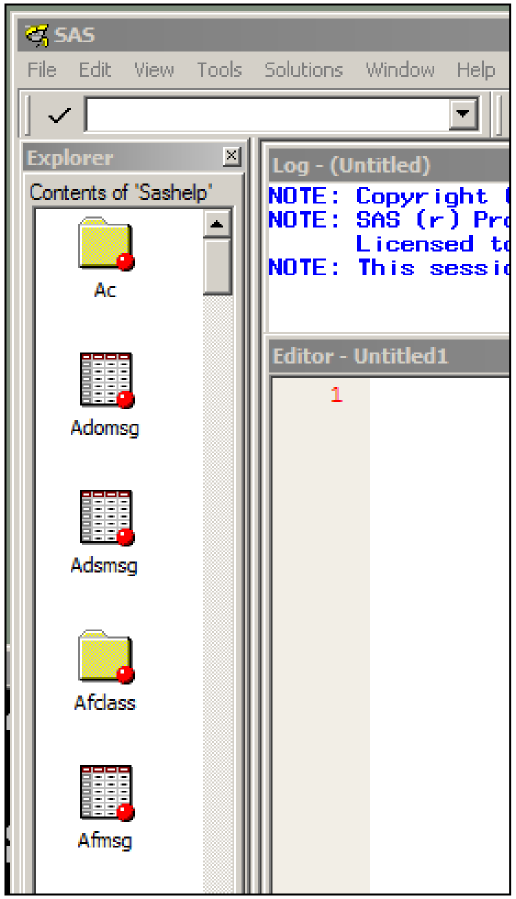
```
libname mylib ‘c:\research\data\’;
proc print data=mylib.trial1;
run;
```

SAS Program and Data Review, (7)

SAS comes with data libraries



```
NOTE: Copyright (c) 2002-2010 by SAS Ins
NOTE: SAS (r) Proprietary Software 9.3 (
      Licensed to UNIVERSITY OF IOWA-T&R
NOTE: This session is executing on the X
```



```
NOTE: Copyright (
NOTE: SAS (r) Pro
      Licensed to
NOTE: This sessio
```

SAS Procedures*

"Base" software

- APPEND
- BMDP
- CALENDAR
- CATALOG
- CDISC
- CHART
- CIMPORT
- COMPARE
- CONTENTS
- CONVERT
- COPY
- CORR
- CPORT
- DATASETS
- DBCSTAB
- DISPLAY
- DOCUMENT
- EXPLODE
- EXPORT
- FCMP
- FORMS
- FREQ
- HTTP
- IMPORT
- INFOMAPS
- ITEMS
- JAVAINFO
- MEANS
- METADATA
- METALIB
- METAOPERATE
- MIGRATE
- OPTIONS
- OPTLOAD
- OPTSAVE
- PDS
- PDSCOPY
- PLOT
- PMENU
- PRINT
- PRINTTO
- PROTO
- PRTDEF
- PRTEXP
- PWENCODE
- RANK
- REGISTRY
- RELEASE
- REPORT
- SCAPROC
- SOAP
- SORT
- SOURCE
- SQL
- STANDARD
- SUMMARY
- TABULATE
- TAPECOPY
- TAPELABEL
- TEMPLATE
- TIMEPLOT
- TRANSPOSE
- TRANTAB
- UNIVARIATE
- VAXTOINTEG
- XSL (Preproduction)

SAS/Stat software

- ACECLUS
- ANOVA
- BOXPLOT
- CALIS
- CANCORR
- CANDISC
- CATMOD
- CLUSTER
- CORRESP
- DISCRIM
- DISTANCE
- FACTOR
- FASTCLUS
- FREQ
- GAM
- GENMOD
- GLIMMIX
- GLM
- GLMMOD
- GLMPOWER
- GLMSELECT
- HPMIXED
- INBREED
- KDE
- KRIGE2D
- LATTICE
- LIFEREG
- LIFETEST
- LOESS
- LOGISTIC
- MCMC
- MDS
- MI
- MIANALYZE
- MIXED
- MODECLUS
- MULTTEST
- NESTED
- NLIN
- NLMIXED
- NPAR1WAY
- ORTHOREG
- PHREG
- PLAN
- PLM
- PLS
- POWER
- PRINCOMP
- PRINQUAL
- PROBIT
- QUANTREG
- REG
- ROBUSTREG
- RSREG
- SCORE
- SEQDESIGN
- SEQTEST
- SIM2D
- SIMNORMAL
- STDIZE
- STEPDISC
- SURVEYFREQ
- SURVEYLOGISTIC
- SURVEYMEANS
- SURVEYPHREG
- SURVEYREG
- SURVEYSELECT
- TPSPLINE
- TRANSREG
- TREE
- TTEST
- VARCLUS
- VARCOMP
- VARIOGRAM

*in two of 20 SAS product lines

Basic PROC Syntax

$$\overbrace{\text{PROC}}^{\textit{“PROC”}}\ \underbrace{\text{Print}}_{\textit{Procedure name}}\ \overbrace{\text{data=sashelp.heart;}}^{\textit{Options (may be many)}}$$

$$\begin{array}{l}\textit{Additional} \\ \textit{procedure} \\ \textit{statements}\end{array}\left\{\ \text{var sex weight;}\right.$$

run;

PROC PRINT

- The PRINT procedure is used to organize and display data in the ‘output’ window as lists or case reports.
- PRINT has many options for controlling the appearance of data in reports.
- PRINT lists data, but has some selection, grouping and summary capabilities.

PROC CONTENTS

- The CONTENTS procedure does much more than generate text output.
  - Review: default text output order is to list variable names alphabetically
    - The VARNUM option sets the list to order by column position (as you would see it in the display manager view of the dataset).
- NOPRINT: Making use of ‘metadata’
  - Simple lists (lists of datasets to process)
  - Simple lines (lines of repetitive code to create and submit)

PROC SORT

- The SORT procedure is used to organize datasets typically in preparation for merging related data, or putting the data into a sequence that will match the 'BY' processing used in a future procedure.
- Sorts can be ASCENDING or DESCENDING and can include one, or all variables in a dataset.
- Sorts can create new datasets
- Sorts can be used to eliminate duplication (but this can be tricky)

PROC FREQ

- The FREQ or FREQUENCY procedure is useful for examining categorical variables. It simply reports frequency counts and calculates some simple sums and percentages.
- Tables can be crossed: variables a X b
- Any table request can include a BY grouping request provided the data are pre-sorted.

PROC MEANS

- Use for descriptive statistics
- Stats for All records is default.
- Can use BY or CLASS statements to report by categories in other variables (ie, Gender, Year, City, etc…).

PROC UNIVARIATE

- Use for descriptive statistics (again)
- Stats for All records is default.
- Can use BY or CLASS statements to report by categories in other variables (ie, Gender, Year, City, etc…).

--break--

Time for a 10-minute break….

Make sure you have “saved” your program

http://uisug.org.uiowa.edu/

Elements of style

sample program

```
data trial1;
infile 'C:\wagedata.txt';
input id days wages;
wage_rate=wages/days;
if wage_rate>20 then lvl='hi';
  else lvl='lo';
run;
```

Elements of style
(don't do this)

```
data trial1;infile 'C:\wagedata.txt'; input id days wages;wage_rate
=wages/days;if wage_rate>20 then lvl='hi';else lvl='lo';run;
```

```
data trial1;  Infile 'C:\wagedata.txt';
Input                                  id
days                                wages;
wage_rate                               =
wages/                               days;
if                           wage_rate>20
then lvl='hi';   else   lvl='lo';   run;
```

```
data
trial1;
Infile
'C:\wagedata.txt
';
input
id
days
wages;
wage_rate
=
wages/
days;
if
wage_rate>20
then
lvl=
'hi';
else
lvl='lo';
run;
```

Elements of style
(do this)

```
/* this is a sample program used to
demonstrate some of the basic
elements of programming style  */

data trial1;
infile 'C:\wagedata.txt';
input id days wages;
wage_rate=wages/days;


* “20” is industry standard for hi;
if wage_rate>20 then lvl='hi';
  else lvl='lo';
run;
```

Large block comment at beginning describing program and purpose

One statement per line

Blank lines to separate sections of the program

Short comment to explain code

Indenting subordinate statements

PROC FORMAT *(build)*

- Lets you define your own informats and formats for variables.

$$\overbrace{\text{PROC Format}}^{\text{“PROC”}}\overset{\text{There could be options here}}{\swarrow};$$

$$\text{Value Statement}\ \left\{\ \text{value }\underbrace{\text{<\$>}}_{\text{char format?}}\ \underbrace{\text{formatname}}_{\text{SAS “normal” name rules}};\right.$$

$$1=\text{‘male’}$$

$$\underbrace{2}_{\text{data value}}=\underbrace{\text{‘female’}}_{\text{“displayed” value}};$$

run;

Merging data

- The MERGE statement is used within a datastep to combine two or more SAS datasets.
- One or more datasets can be merged by a ‘key’ variable, or group of variables that creates a unique key
  - SAS will let you merge with repeats on the key, and it will note this in the log.

Merging data

Patient Data

patno	lname	sex
11	Jones	M
66	Smith	M
33	Brown	F
55	Harris	F
44	Anderson	F
22	Collins	M

Visit Data

patno	visit #	wt
11	1	137
11	2	135
33	1	186
33	2	182
33	3	176
66	1	157

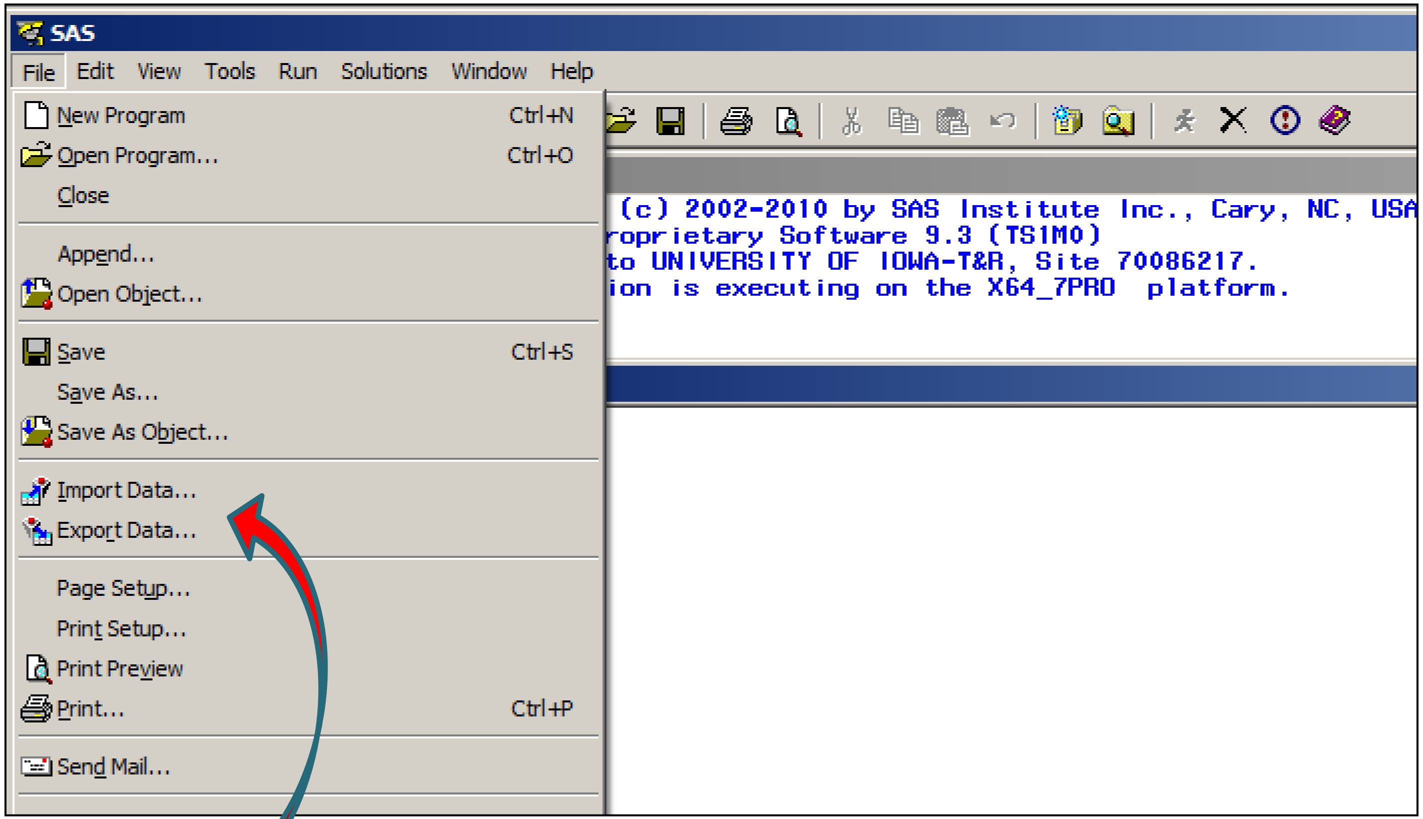
“Merged” Data

patno	lname	sex	visit #	wt
11	Jones	M	1	137
11	Jones	M	2	135
22	Collins	M	.	.
33	Brown	F	1	186
33	Brown	F	2	182
33	Brown	F	3	176
44	Anderson	F	.	.
55	Harris	F	.	.
66	Smith	M	1	157

Import/Export Data

- SAS can import data from, and export data to, many different formats
  - MS-Excel
  - MS-Access
  - .csv
  - SPSS
  - Stata
  - many others
- A variety of methods for importing/exporting
- Best approach depends on variety of factors
  - Operating system (Linux, Windows, 32/64-bit)
  - SAS version (9.2, 9.3, 32/64-bit)
  - Originating/destination software (Excel, .csv, SPSS)
- Use the Wizard
  - Be careful, pay attention

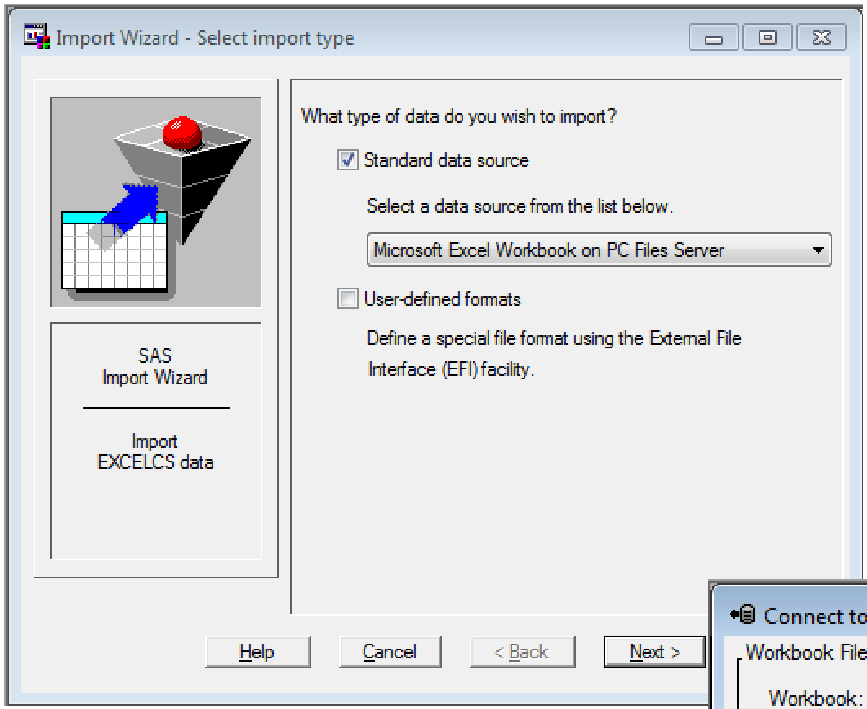
Import/Export Data (2)



```
 (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA
roprietary Software 9.3 (TS1M0)
to UNIVERSITY OF IOWA-T&R, Site 70086217.
ion is executing on the X64_7PRO  platform.
```

Wizards!

Import/Export Data (3)



Import Wizard - Select import type

What type of data do you wish to import?

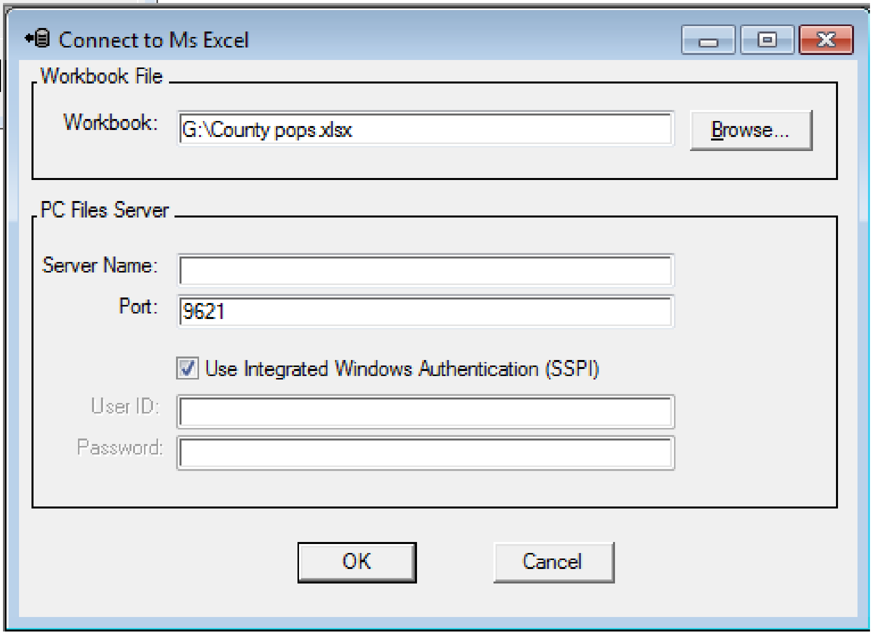
[✓] Standard data source

Select a data source from the list below.

Microsoft Excel Workbook on PC Files Server

[ ] User-defined formats

Define a special file format using the External File Interface (EFI) facility.



Connect to Ms Excel

Workbook File

Workbook: G:\County pops.xlsx

PC Files Server

Server Name:

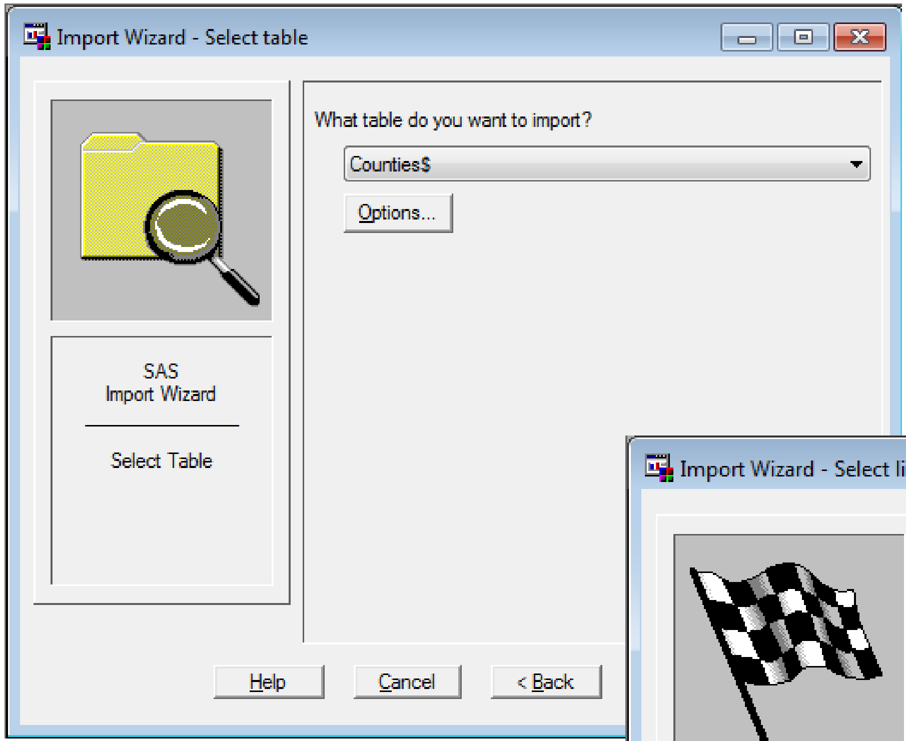
Port: 9621

[✓] Use Integrated Windows Authentication (SSPI)

User ID:

Password:

Import/Export Data (4)

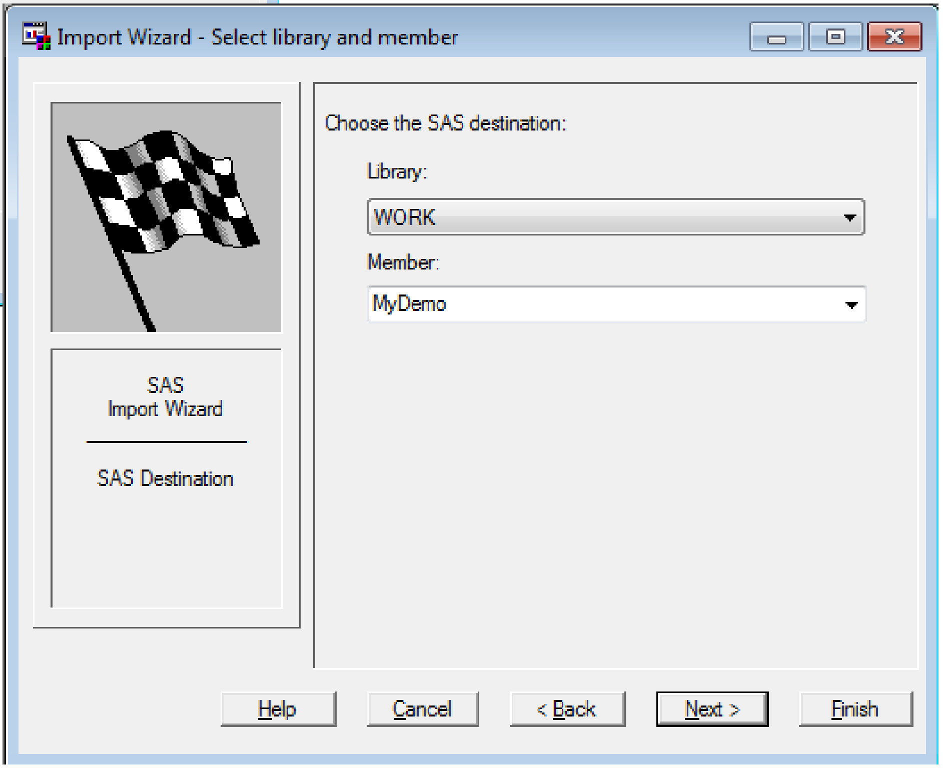

Import Wizard - Select table

SAS Import Wizard

Select Table

What table do you want to import?

Counties$


Import Wizard - Select library and member

SAS Import Wizard

SAS Destination

Choose the SAS destination:

Library:

WORK

Member:

MyDemo

PROC SQL

Structured Query Language (SQL)

A language used for managing data in many different computer applications (primarily database applications). It has been available in SAS since the late 1980's and can be used for a wide variety of purposes including nearly everything we have done this morning.

Syntax:

```
PROC SQL;
   SQL statements;
quit;
```

PROC TRANSPOSE

- The TRANSPOSE procedure is what we use to flip data on it's side.
- It is recommended to do this in small chunks so that it is easy for you to understand and explain to others.
- With experience you can transpose multiple variables simultaneously.

PROC TRANSPOSE

Raw Data

Variable	Bob	Tim	Kim	Ann	Pat
age	23	25	21	26	24
height	62	61	66	71	69
weight	120	125	160	220	205
score	88	93	100	75	98

"Transposed" Data

Name	Age	Height	Weight	Score
Bob	23	62	120	88
Tim	25	61	125	93
Kim	21	66	160	100
Ann	26	71	220	75
Pat	24	69	205	98

Review and questions

THANKS to our SAS experts for their onsite assistance

-- break for lunch --

- **Thank you** for participating in our session.
- We hope these materials will be helpful.

- Enjoy a lunch break and return for the next session featuring SAS Interactive Graphics.