Introduction to SAS

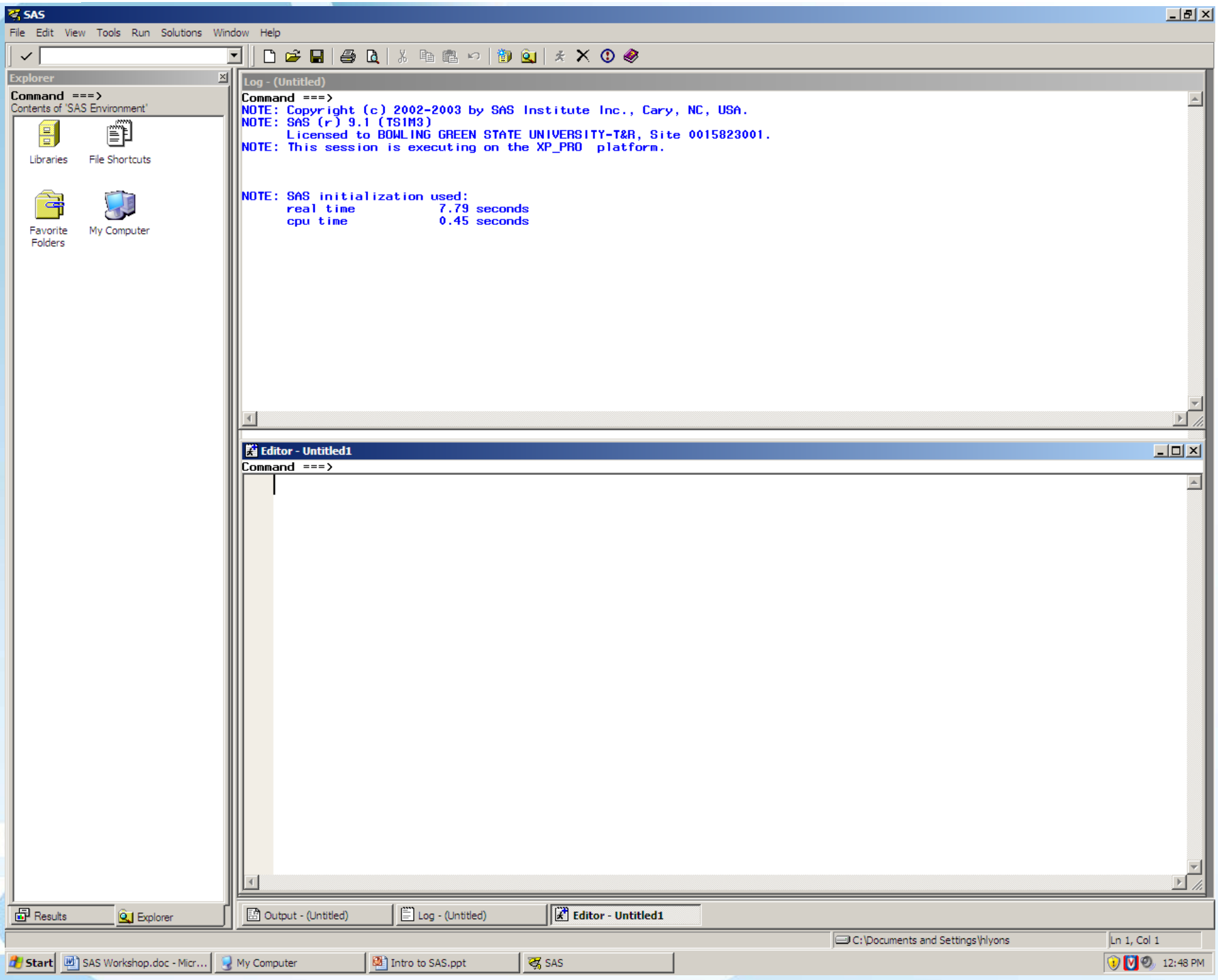
Heidi Lyons

October 5, 2009

BGSU
Center for Family and Demographic Research

Why SAS?

- Often better with very large datasets and memory.
- Can deal with multiple datasets at the same time.
- Better for data manipulation.
- Better on the Job Market.


SAS
File Edit View Tools Run Solutions Window Help
Explorer
Command ===>
Contents of 'SAS Environment'
Libraries
File Shortcuts
Favorite Folders
My Computer
Results
Explorer
Log - (Untitled)
Command ===>
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
Licensed to BOWLING GREEN STATE UNIVERSITY-T&R, Site 0015823001.
NOTE: This session is executing on the XP_PRO platform.
NOTE: SAS initialization used:
real time 7.79 seconds
cpu time 0.45 seconds
Editor - Untitled1
Command ===>
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1
C:\Documents and Settings\hlyons
Ln 1, Col 1
Start
SAS Workshop.doc - Micr...
My Computer
Intro to SAS.ppt
SAS
12:48 PM

Editor - Untitled1 *

Command ===>

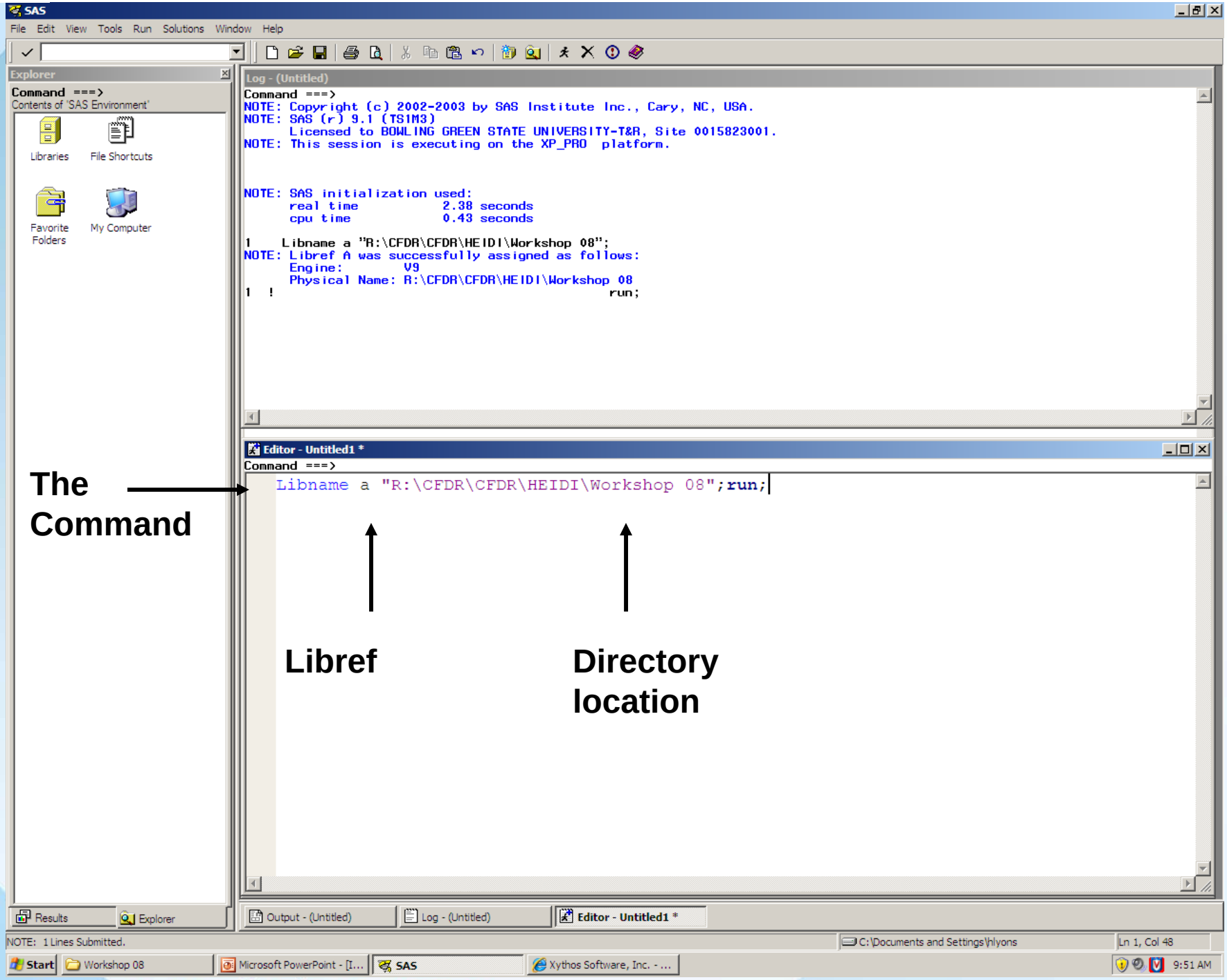
```
This is ; wrong; Help!;

data girls; set fun;
proc freq;tables sex;run;

/*coment*/
```

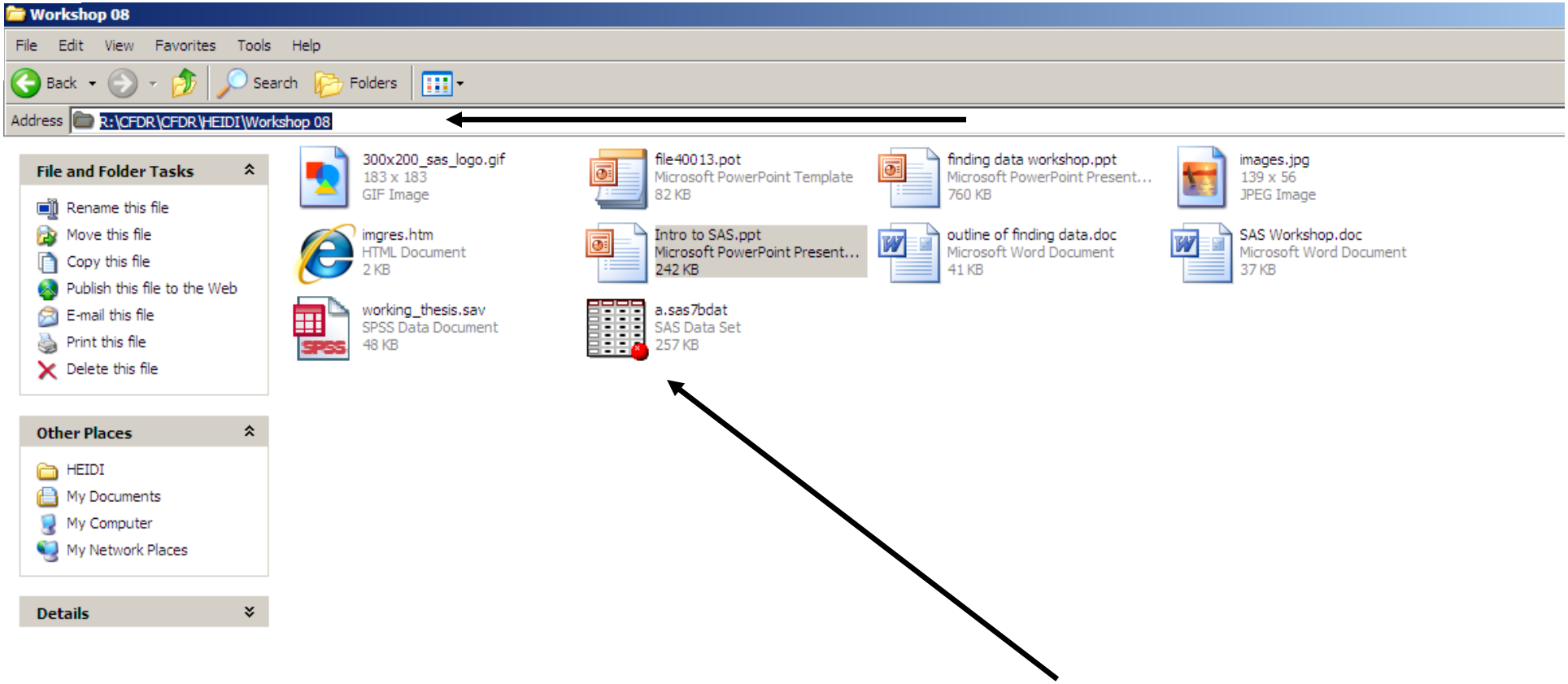
I have data on the….how do I get it into SAS

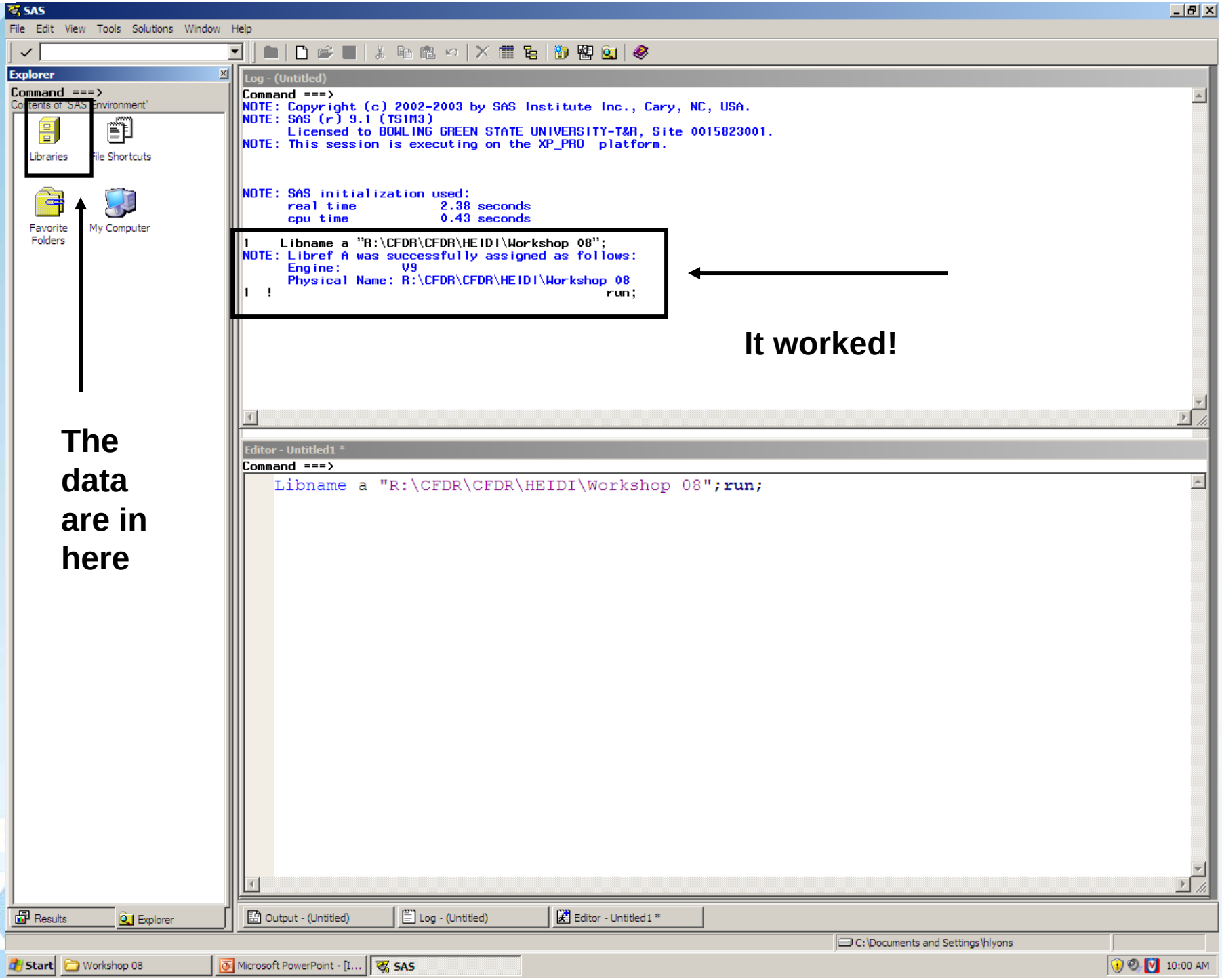
- The infamous libname statement
  - Makes datasets permanent
  - Points to and creates directories
  - Can have and often will have two or more libname statements

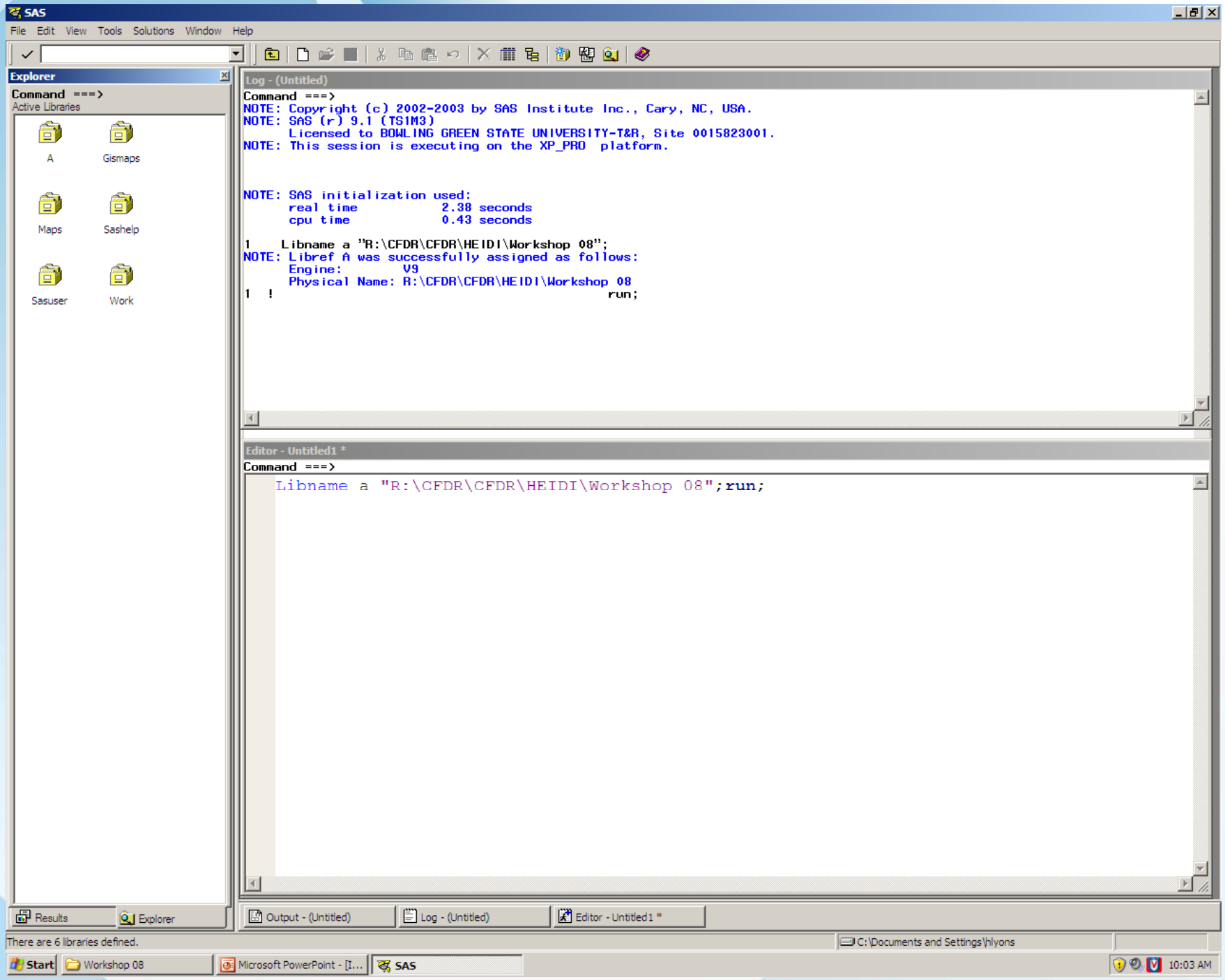
Let’s break it down.

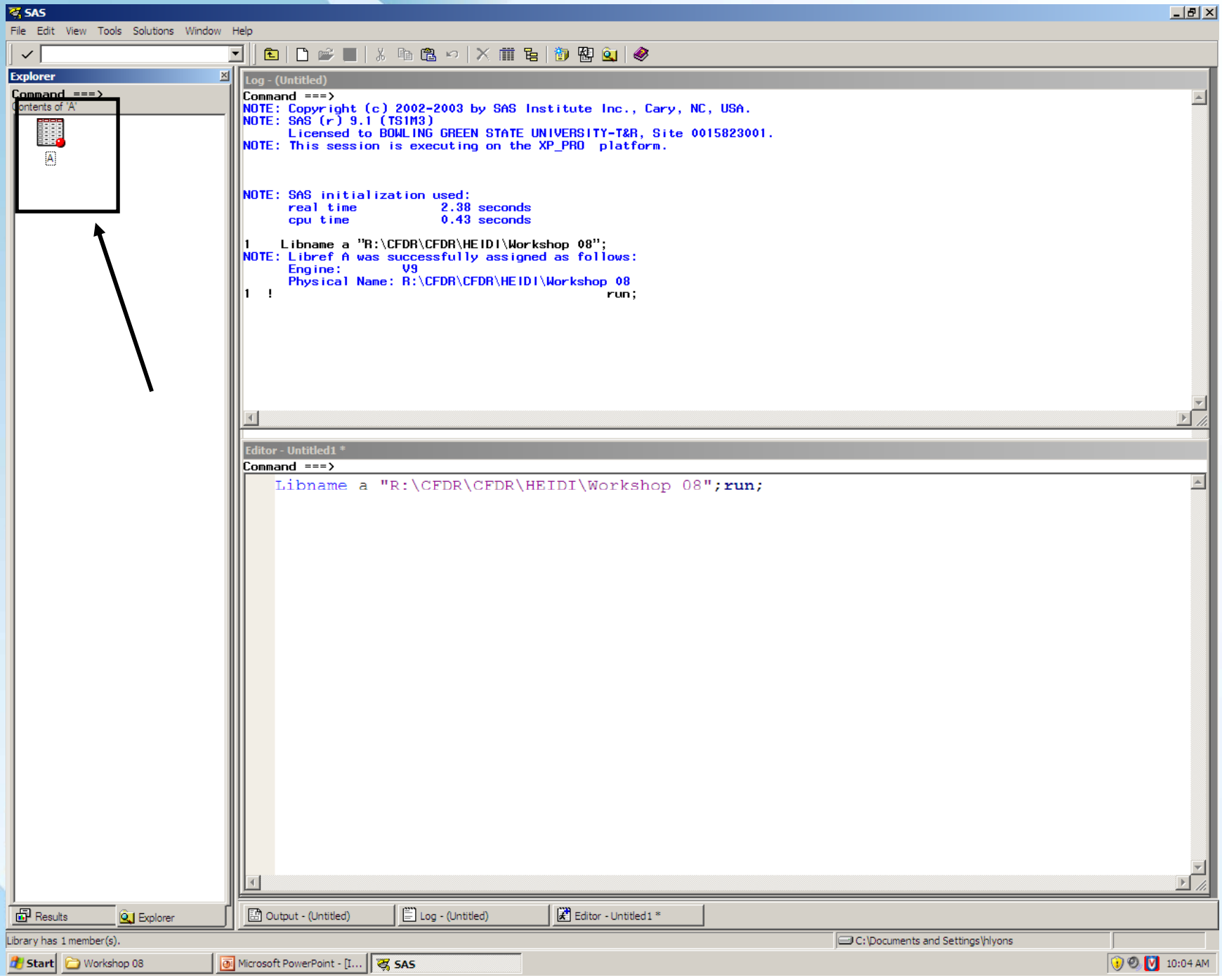
- Libname-just the command.
- Libref-what you are going to call it?
- Directory location-where is your dataset or where do you want to store it?


SAS
File Edit View Tools Run Solutions Window Help
Explorer
Command ===>
Contents of 'SAS Environment'
Libraries
File Shortcuts
Favorite Folders
My Computer
Log - (Untitled)
Command ===>
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
Licensed to BOWLING GREEN STATE UNIVERSITY-T&R, Site 0015823001.
NOTE: This session is executing on the XP_PRO platform.
NOTE: SAS initialization used:
real time 2.38 seconds
cpu time 0.43 seconds
1 Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";
NOTE: Libref A was successfully assigned as follows:
Engine: V9
Physical Name: R:\CFDR\CFDR\HEIDI\Workshop 08
1 ! run;
Editor - Untitled1 *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
The Command
Libref
Directory location
Results
Explorer
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 *
NOTE: 1 Lines Submitted.
C:\Documents and Settings\hlyons
Ln 1, Col 48
Start
Workshop 08
Microsoft PowerPoint - [I...
SAS
Xythos Software, Inc. - ...
9:51 AM


Workshop 08
File Edit View Favorites Tools Help
Back Search Folders
Address R:\CFDR\CFDR\HEIDI\Workshop 08
File and Folder Tasks
Rename this file
Move this file
Copy this file
Publish this file to the Web
E-mail this file
Print this file
Delete this file
Other Places
HEIDI
My Documents
My Computer
My Network Places
Details
300x200_sas_logo.gif
183 x 183
GIF Image
file40013.pot
Microsoft PowerPoint Template
82 KB
finding data workshop.ppt
Microsoft PowerPoint Present...
760 KB
images.jpg
139 x 56
JPEG Image
imgres.htm
HTML Document
2 KB
Intro to SAS.ppt
Microsoft PowerPoint Present...
242 KB
outline of finding data.doc
Microsoft Word Document
41 KB
SAS Workshop.doc
Microsoft Word Document
37 KB
working_thesis.sav
SPSS Data Document
48 KB
a.sas7bdat
SAS Data Set
257 KB


SAS
File Edit View Tools Solutions Window Help
Explorer
Command ===>
Libraries
File Shortcuts
Favorite Folders
My Computer
Log - (Untitled)
Command ===>
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TSIM3)
Licensed to BOWLING GREEN STATE UNIVERSITY-T&R, Site 0015823001.
NOTE: This session is executing on the XP_PRO platform.
NOTE: SAS initialization used:
real time 2.38 seconds
cpu time 0.43 seconds
1 Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";
NOTE: Libref A was successfully assigned as follows:
Engine: V9
Physical Name: R:\CFDR\CFDR\HEIDI\Workshop 08
1 ! run;
It worked!
Editor - Untitled1 *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
The data are in here
Results
Explorer
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 *
C:\Documents and Settings\hlyons
Start
Workshop 08
Microsoft PowerPoint - [I...
SAS
10:00 AM


SAS
File Edit View Tools Solutions Window Help
Explorer
Command ===>
Active Libraries
A
Gismaps
Maps
Sashelp
Sasuser
Work
Results
Explorer
Log - (Untitled)
Command ===>
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
Licensed to BOWLING GREEN STATE UNIVERSITY-T&R, Site 0015823001.
NOTE: This session is executing on the XP_PRO platform.
NOTE: SAS initialization used:
real time 2.38 seconds
cpu time 0.43 seconds
1 Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";
NOTE: Libref A was successfully assigned as follows:
Engine: V9
Physical Name: R:\CFDR\CFDR\HEIDI\Workshop 08
1 ! run;
Editor - Untitled1 *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 *
There are 6 libraries defined.
C:\Documents and Settings\hlyons
Start
Workshop 08
Microsoft PowerPoint - [I...
SAS
10:03 AM


SAS
File Edit View Tools Solutions Window Help
Explorer
Command ===>
Contents of 'A'
A
Log - (Untitled)
Command ===>
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
Licensed to BOWLING GREEN STATE UNIVERSITY-T&R, Site 0015823001.
NOTE: This session is executing on the XP_PRO platform.
NOTE: SAS initialization used:
real time 2.38 seconds
cpu time 0.43 seconds
1 Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";
NOTE: Libref A was successfully assigned as follows:
Engine: V9
Physical Name: R:\CFDR\CFDR\HEIDI\Workshop 08
1 ! run;
Editor - Untitled1 *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
Results
Explorer
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 *
Library has 1 member(s).
C:\Documents and Settings\hlyons
Start
Workshop 08
Microsoft PowerPoint - [I...
SAS
10:04 AM

Sometimes you may need two (or more) libnames

- Why?

Things to remember

- Libnames just create a directory. It is not creating datasets.
- Once you create datasets this is how you make them permanent.
- You do need to run you libname statement every time you open SAS.

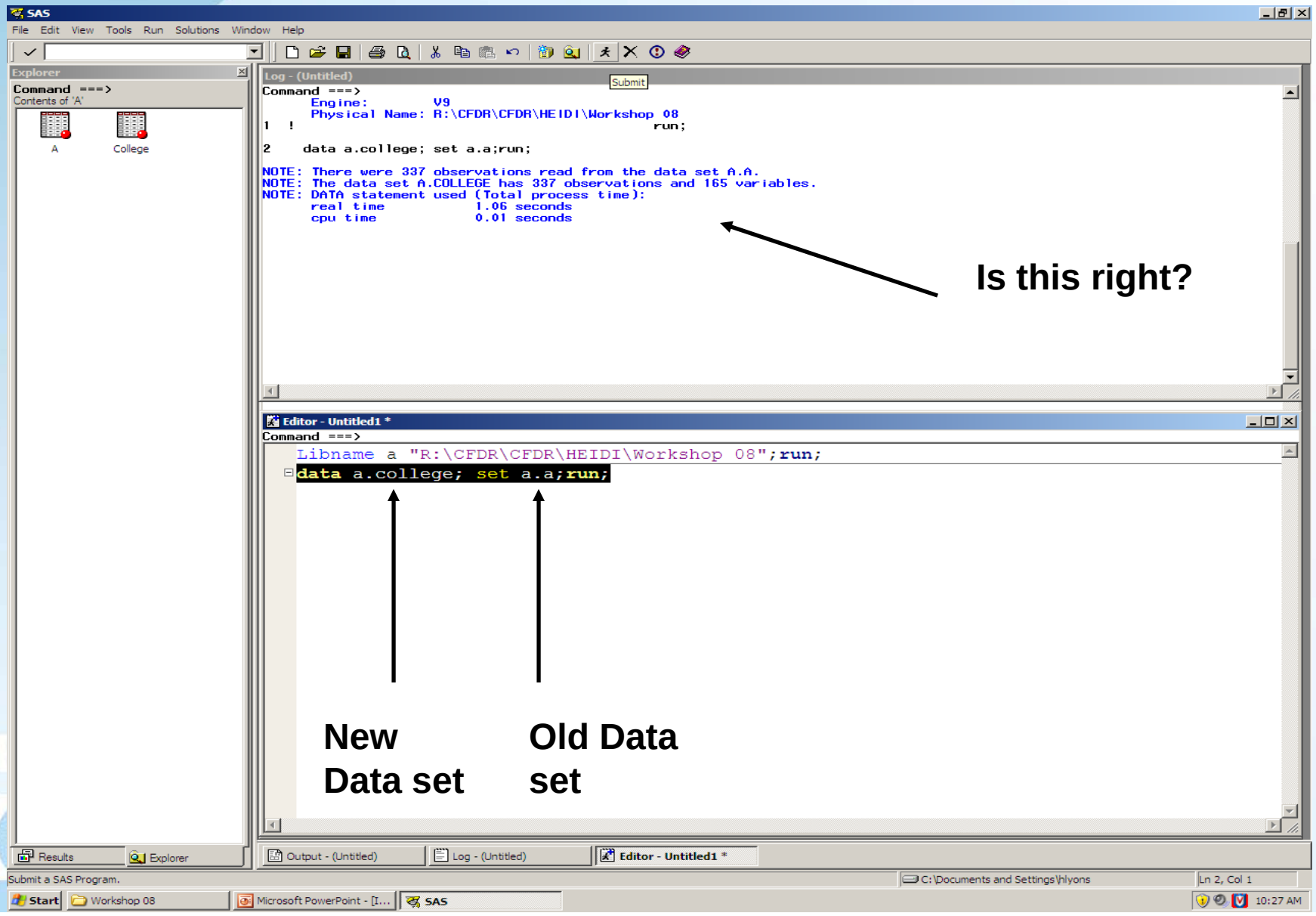
Data and Proc Steps

- Beside the libname statement Data and Proc Steps are the only other thing you need to know how to do in basic SAS!

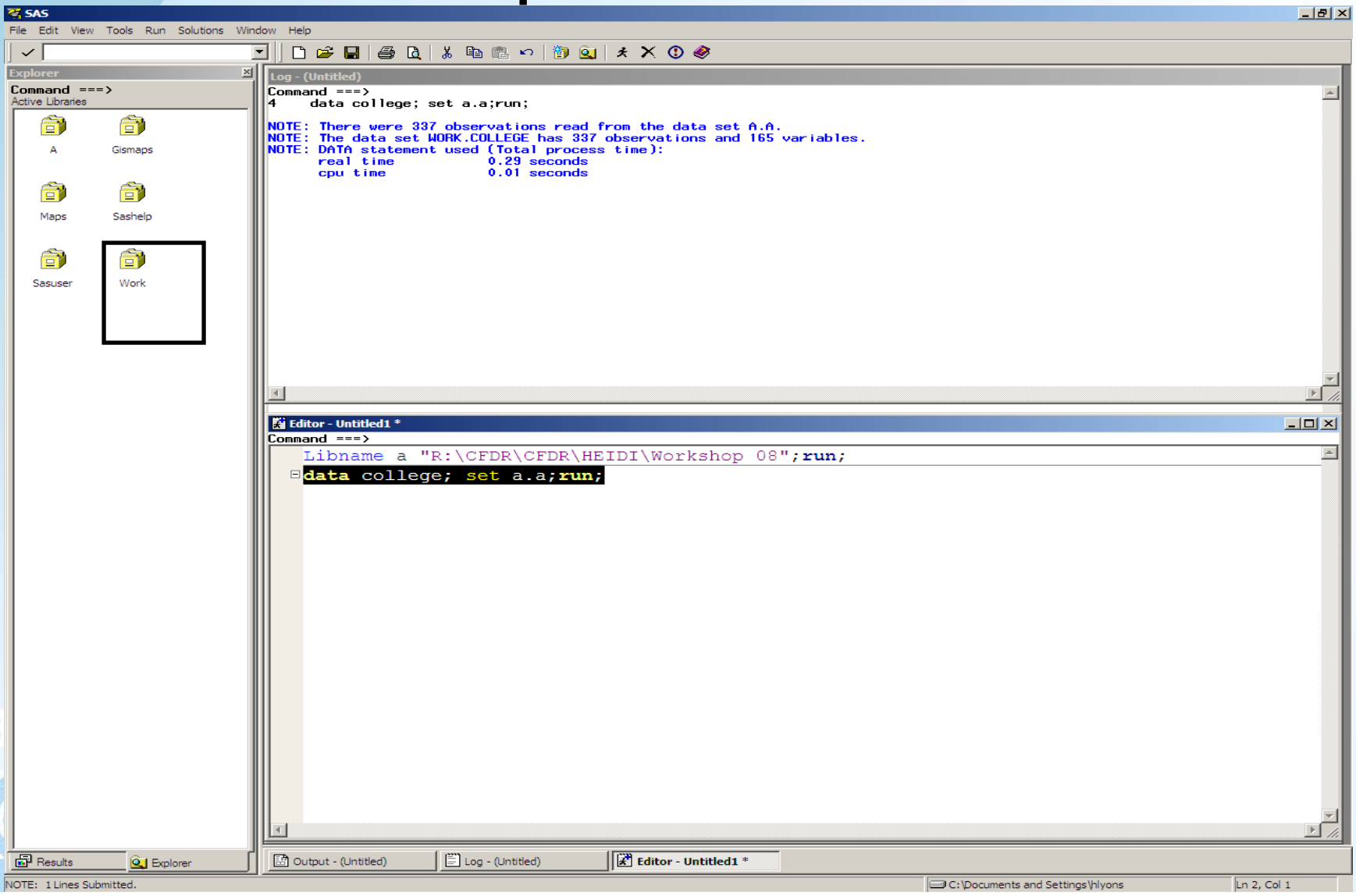
Data Step

- Prepares data for analysis.
- Creates new datasets.
- Modifies and manipulates old datasets.
  - This is one of SAS's strengths.

Remember Data *new* Set *old*

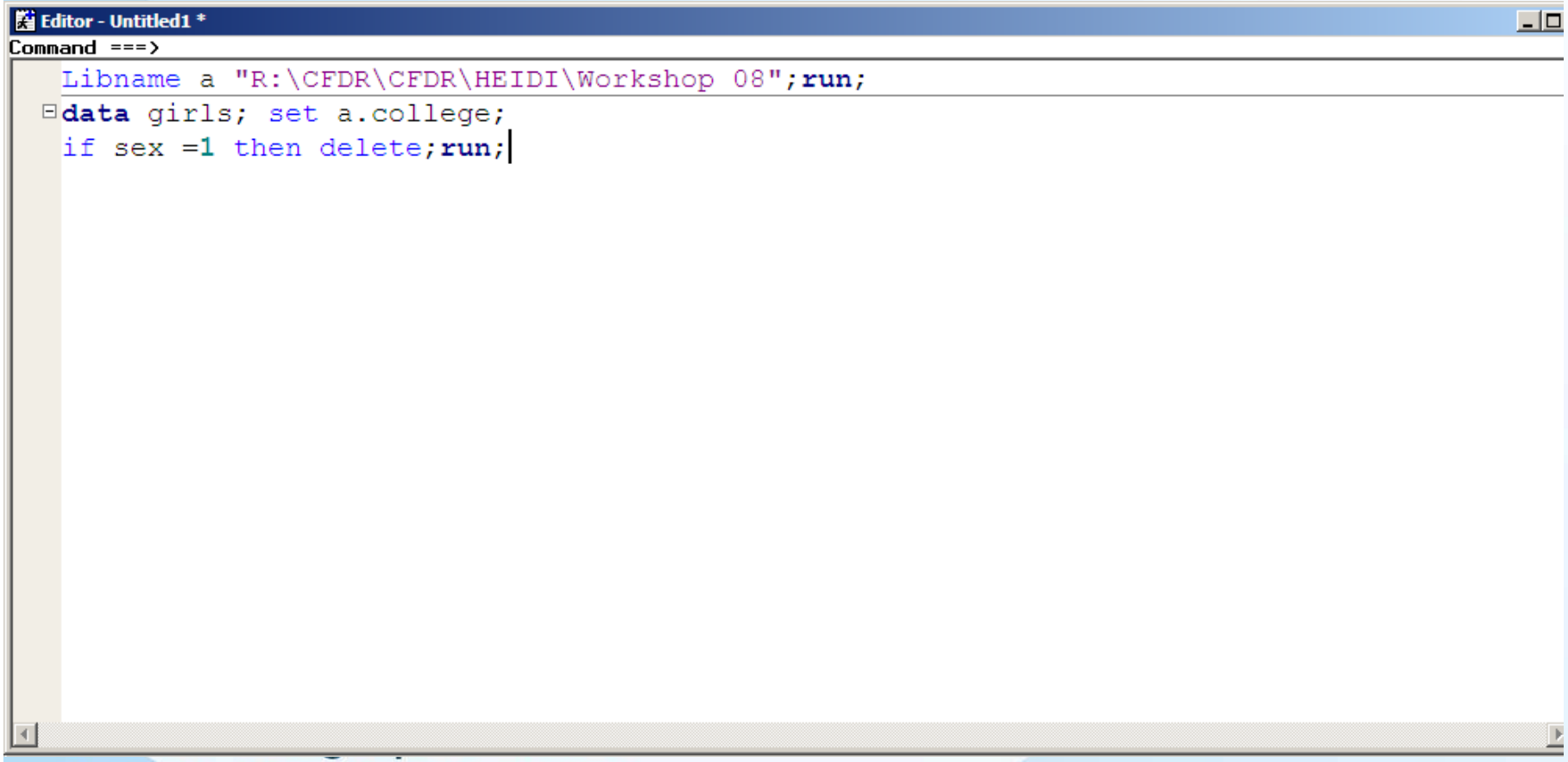


What if I don't want my dataset to be permanent?



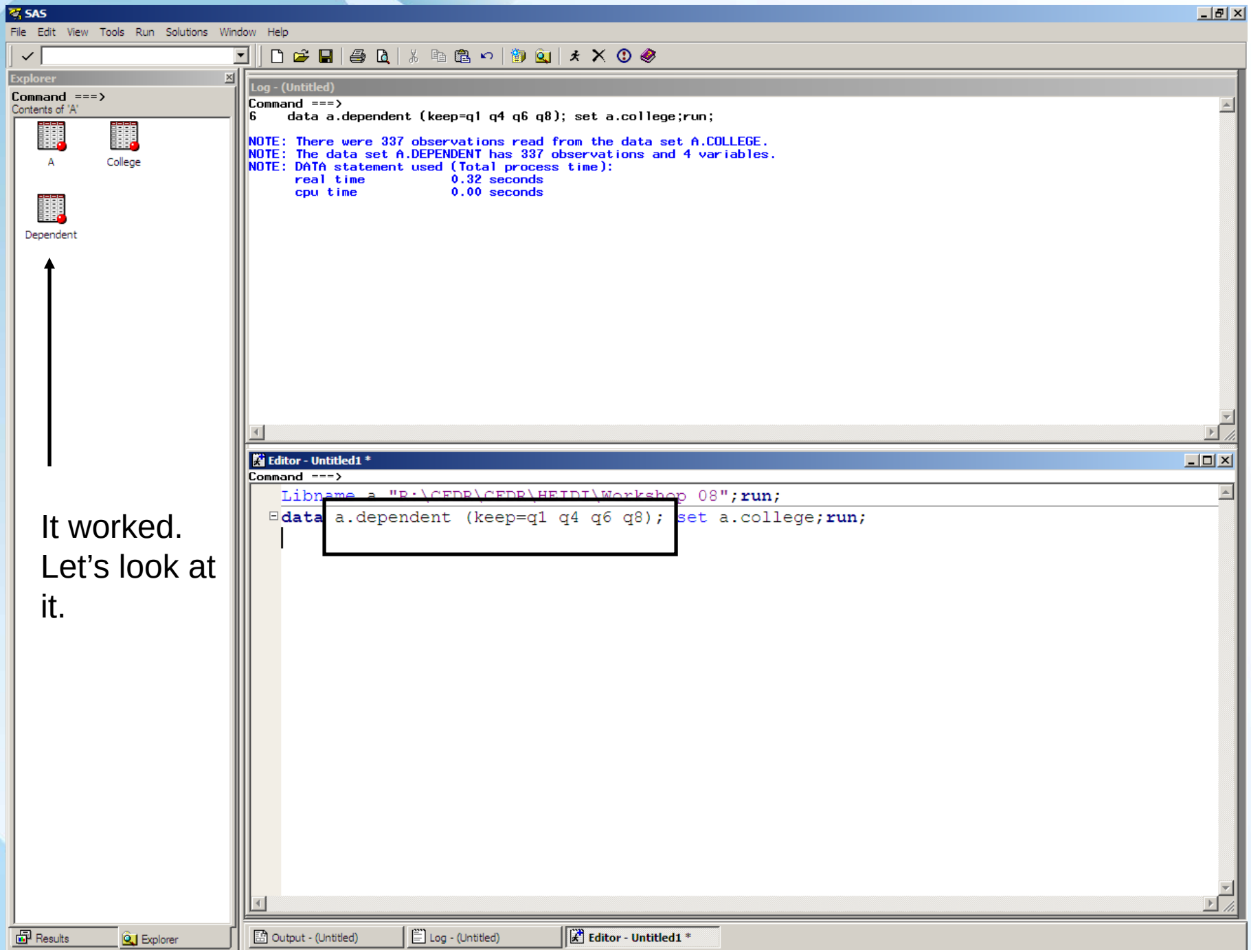
How do I make a work dataset permanent?

What is the next line of code?

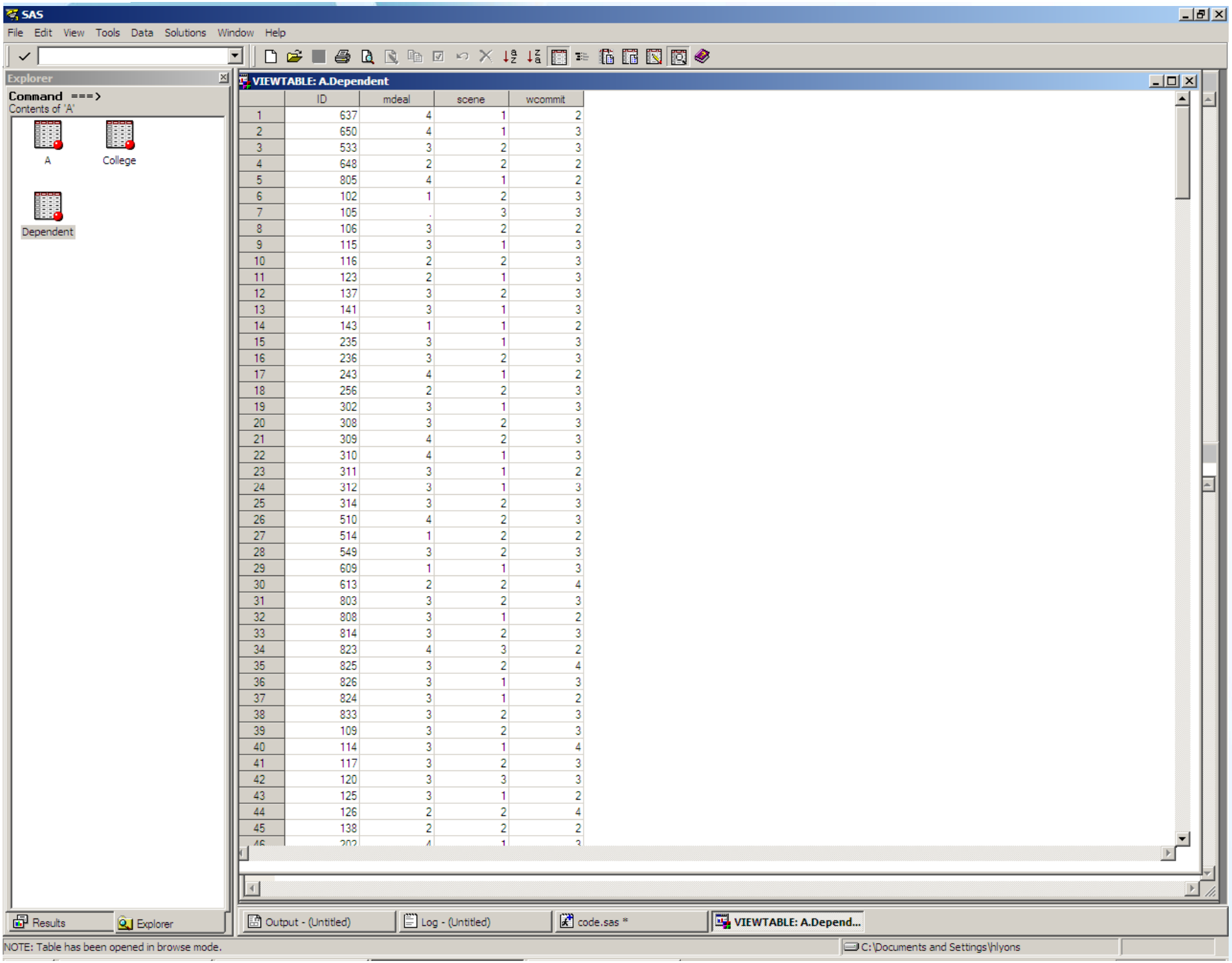


Keep/Drop Statements

- Makes your dataset smaller.



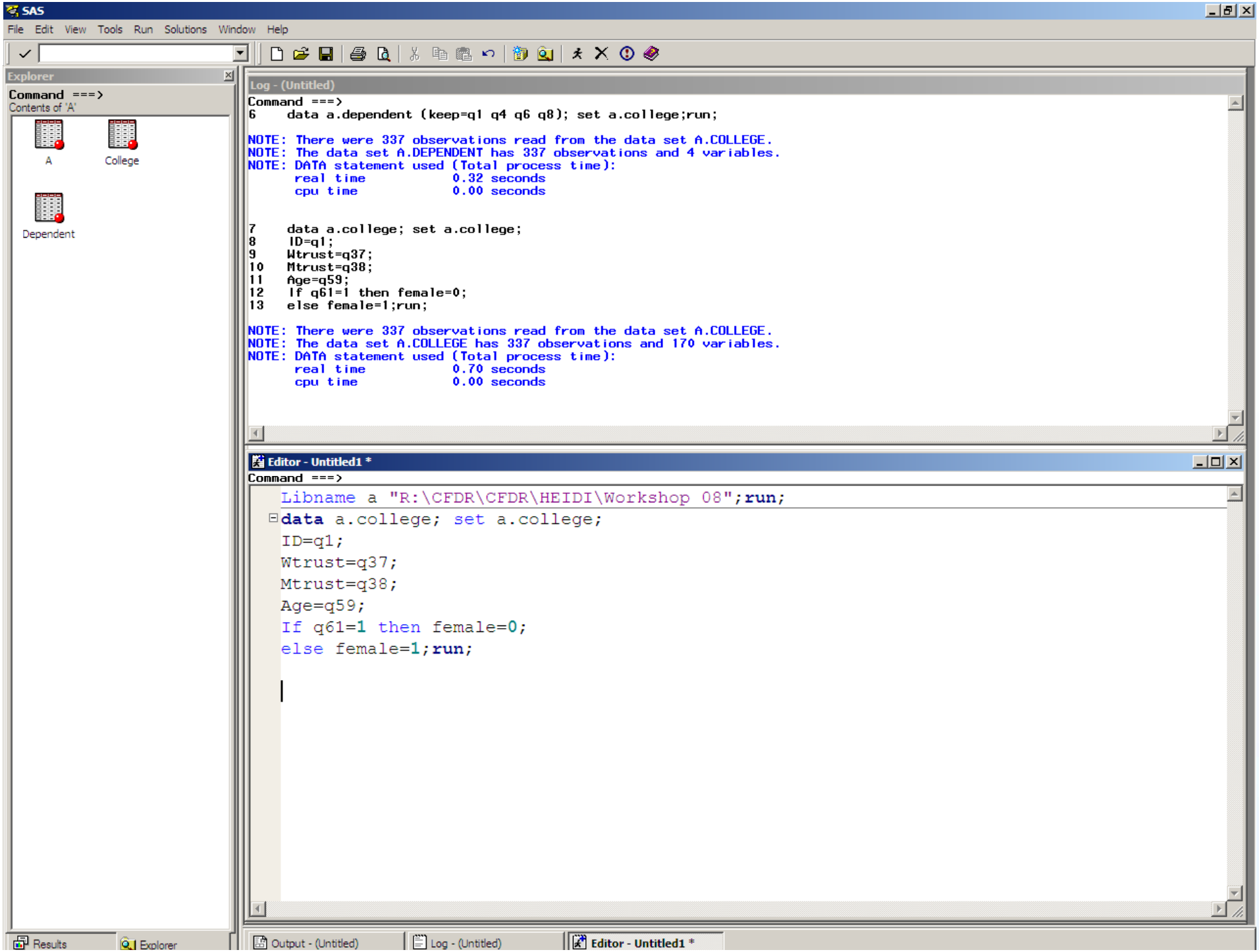
It worked. Let's look at it.



	ID	mdeal	scene	wcommit
1	637	4	1	2
2	650	4	1	3
3	533	3	2	3
4	648	2	2	2
5	805	4	1	2
6	102	1	2	3
7	105	.	3	3
8	106	3	2	2
9	115	3	1	3
10	116	2	2	3
11	123	2	1	3
12	137	3	2	3
13	141	3	1	3
14	143	1	1	2
15	235	3	1	3
16	236	3	2	3
17	243	4	1	2
18	256	2	2	3
19	302	3	1	3
20	308	3	2	3
21	309	4	2	3
22	310	4	1	3
23	311	3	1	2
24	312	3	1	3
25	314	3	2	3
26	510	4	2	3
27	514	1	2	2
28	549	3	2	3
29	609	1	1	3
30	613	2	2	4
31	803	3	2	3
32	808	3	1	2
33	814	3	2	3
34	823	4	3	2
35	825	3	2	4
36	826	3	1	3
37	824	3	1	2
38	833	3	2	3
39	109	3	2	3
40	114	3	1	4
41	117	3	2	3
42	120	3	3	3
43	125	3	1	2
44	126	2	2	4
45	138	2	2	2
46	202	4	1	3

What else can the data step do for me?

- Create variables
  - SAS uses + - / and *
  - If-then-else statements


SAS
File Edit View Tools Run Solutions Window Help
Explorer
Command ===>
Contents of 'A'
A
College
Dependent
Log - (Untitled)
Command ===>
6 data a.dependent (keep=q1 q4 q6 q8); set a.college;run;
NOTE: There were 337 observations read from the data set A.COLLEGE.
NOTE: The data set A.DEPENDENT has 337 observations and 4 variables.
NOTE: DATA statement used (Total process time):
real time 0.32 seconds
cpu time 0.00 seconds
7 data a.college; set a.college;
8 ID=q1;
9 Wtrust=q37;
10 Mtrust=q38;
11 Age=q59;
12 If q61=1 then female=0;
13 else female=1;run;
NOTE: There were 337 observations read from the data set A.COLLEGE.
NOTE: The data set A.COLLEGE has 337 observations and 170 variables.
NOTE: DATA statement used (Total process time):
real time 0.70 seconds
cpu time 0.00 seconds
Editor - Untitled1 *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
data a.college; set a.college;
ID=q1;
Wtrust=q37;
Mtrust=q38;
Age=q59;
If q61=1 then female=0;
else female=1;run;
Results
Explorer
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 *

SAS Operations

- EQ = *equal to*
- NE ~= *never equal*
- LT < *less than*
- *GT > greater than*
- GE >= *greater than or equal to*
- LE <= *less than or equal to*

Review so Far

- You need a libname statement every time you open SAS.
- You need a libname statement to tell SAS where to store the data.
- SAS can create both work and permanent datasets.
- Data new Set old
- SAS will use the most recent dataset.
- Remember the semicolon.

The Proc step

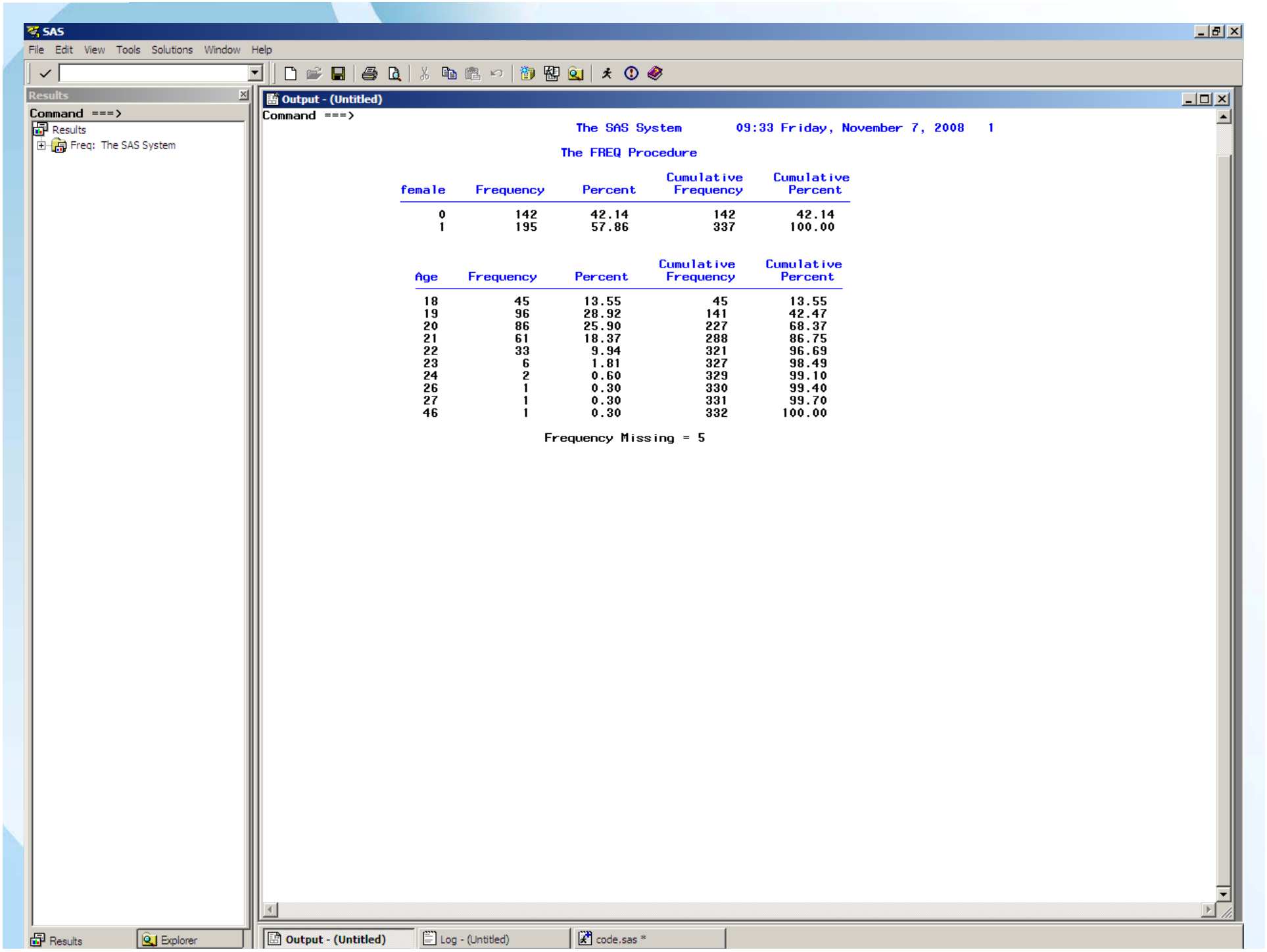
- This is how you analyzed your data
- The Proc Step can
  - Specify the dataset you want to use
  - Name the procedure
  - Any other details of the analysis

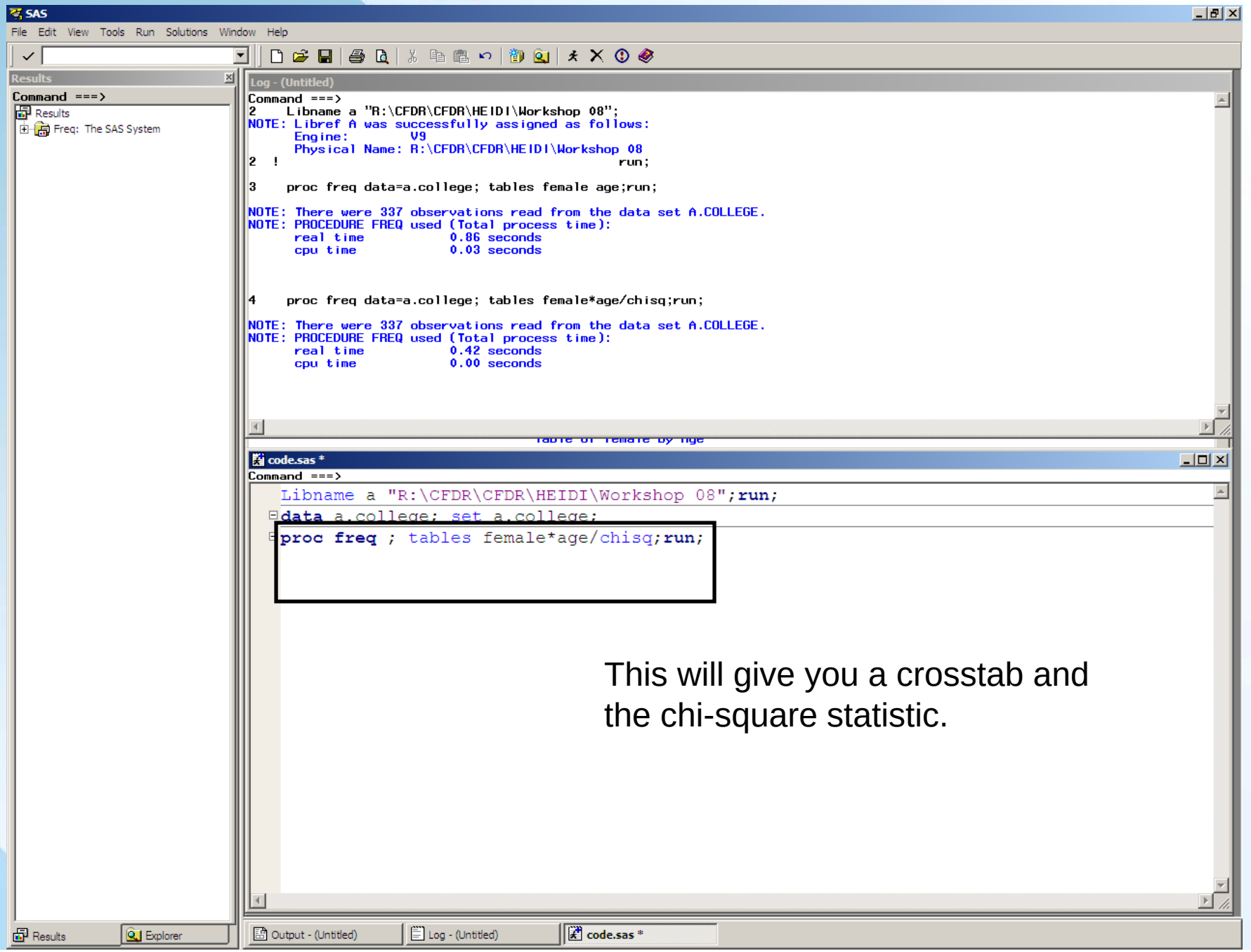
Specifying the Dataset

code.sas *

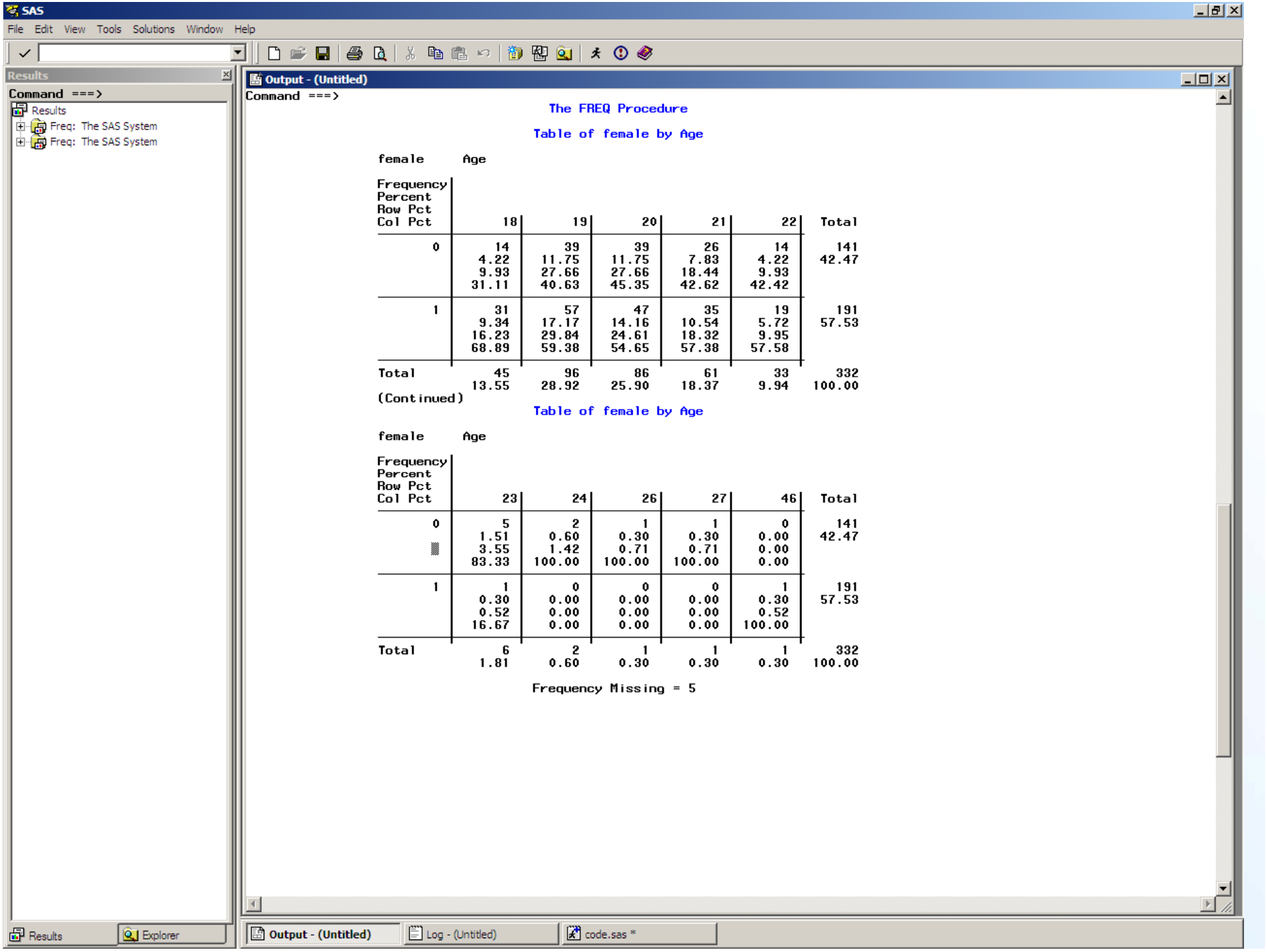
Command ===>

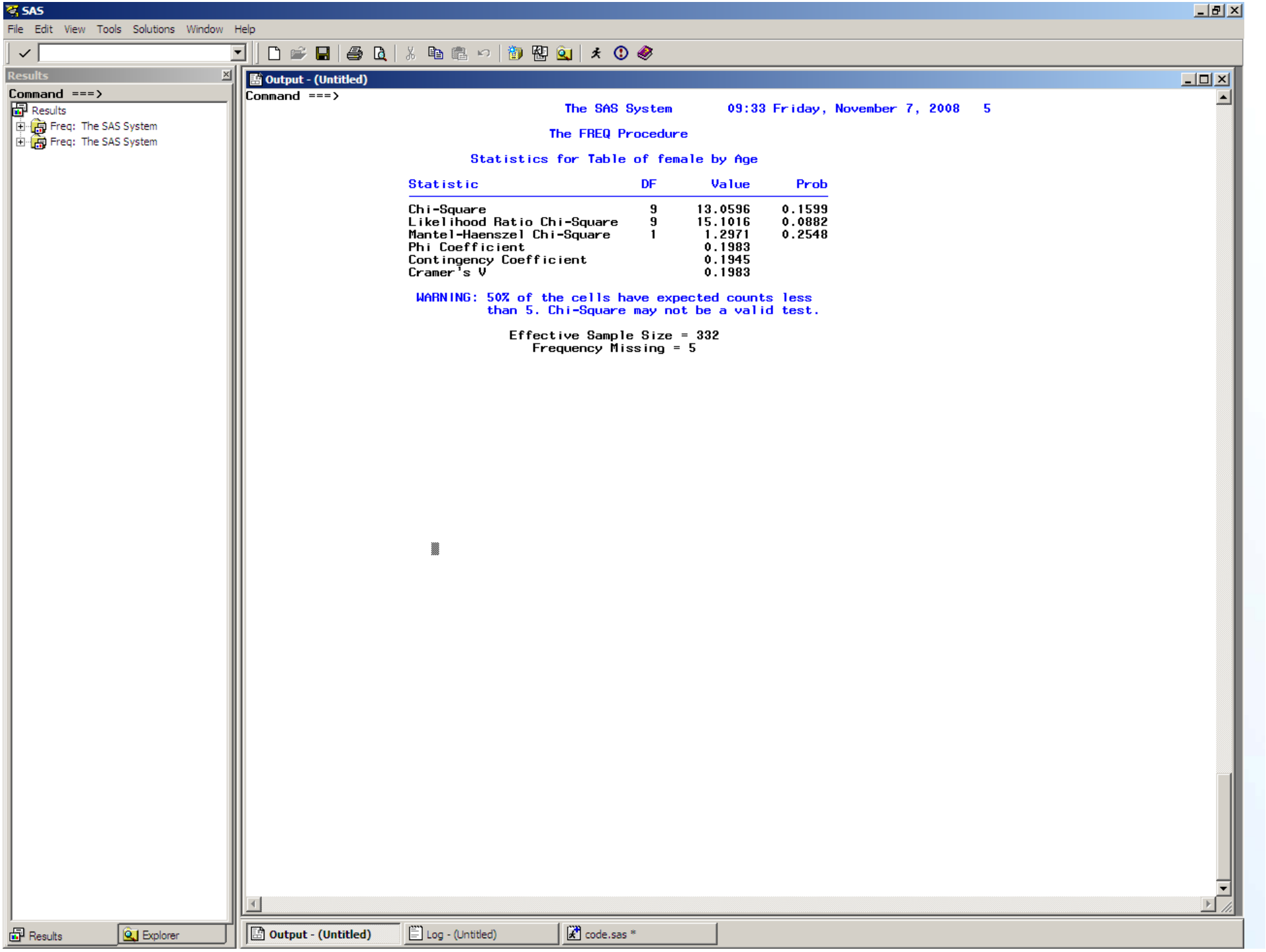
```
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
data a.college; set a.college;
proc freq data=a.college; tables female age;run;
```


SAS
File Edit View Tools Solutions Window Help
Results
Command ===>
Results
Freq: The SAS System
Output - (Untitled)
Command ===>
The SAS System 09:33 Friday, November 7, 2008 1
The FREQ Procedure
female Frequency Percent Cumulative Frequency Cumulative Percent
0 142 42.14 142 42.14
1 195 57.86 337 100.00
Age Frequency Percent Cumulative Frequency Cumulative Percent
18 45 13.55 45 13.55
19 96 28.92 141 42.47
20 86 25.90 227 68.37
21 61 18.37 288 86.75
22 33 9.94 321 96.69
23 6 1.81 327 98.49
24 2 0.60 329 99.10
26 1 0.30 330 99.40
27 1 0.30 331 99.70
46 1 0.30 332 100.00
Frequency Missing = 5
Results
Explorer
Output - (Untitled)
Log - (Untitled)
code.sas *



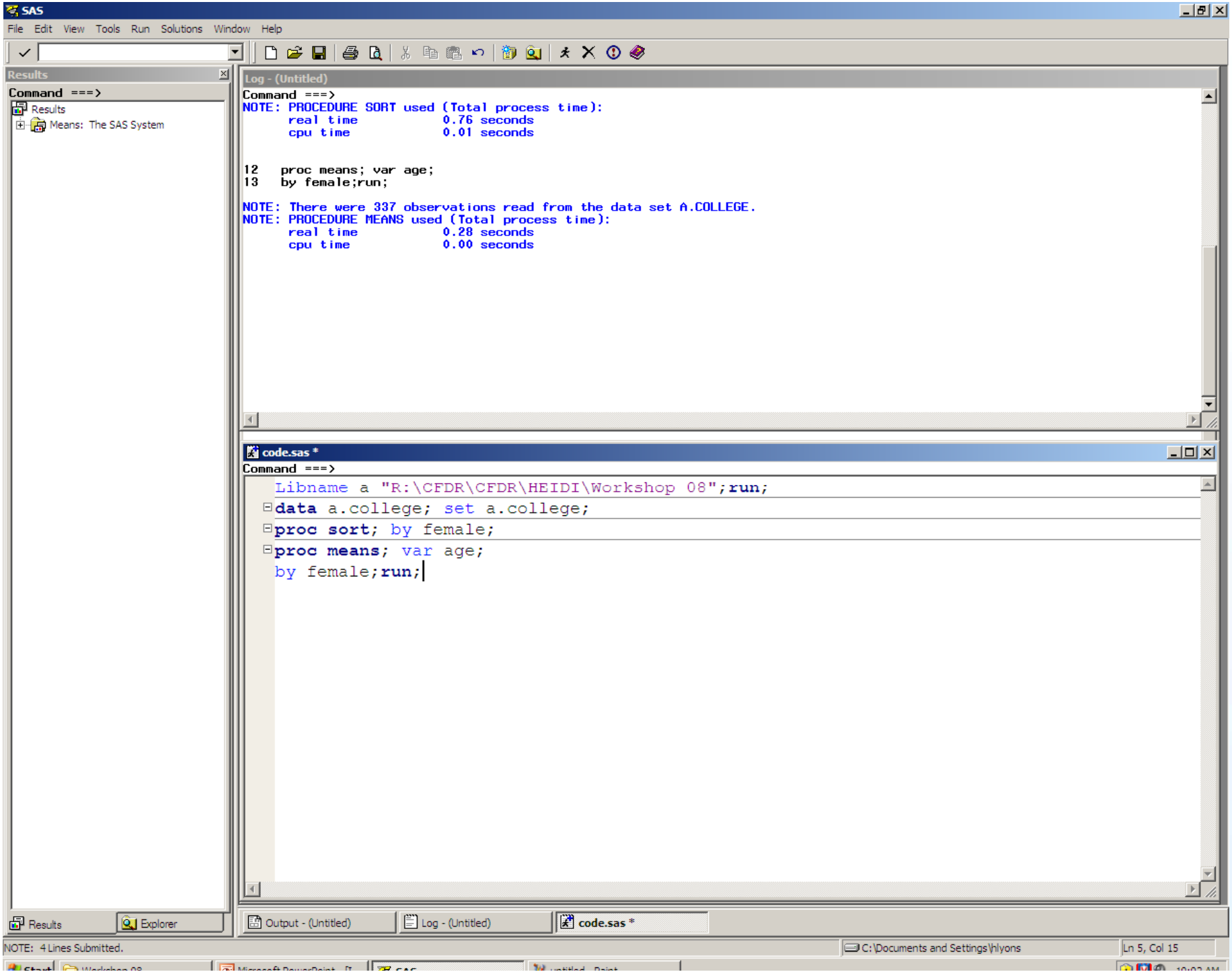
This will give you a crosstab and the chi-square statistic.

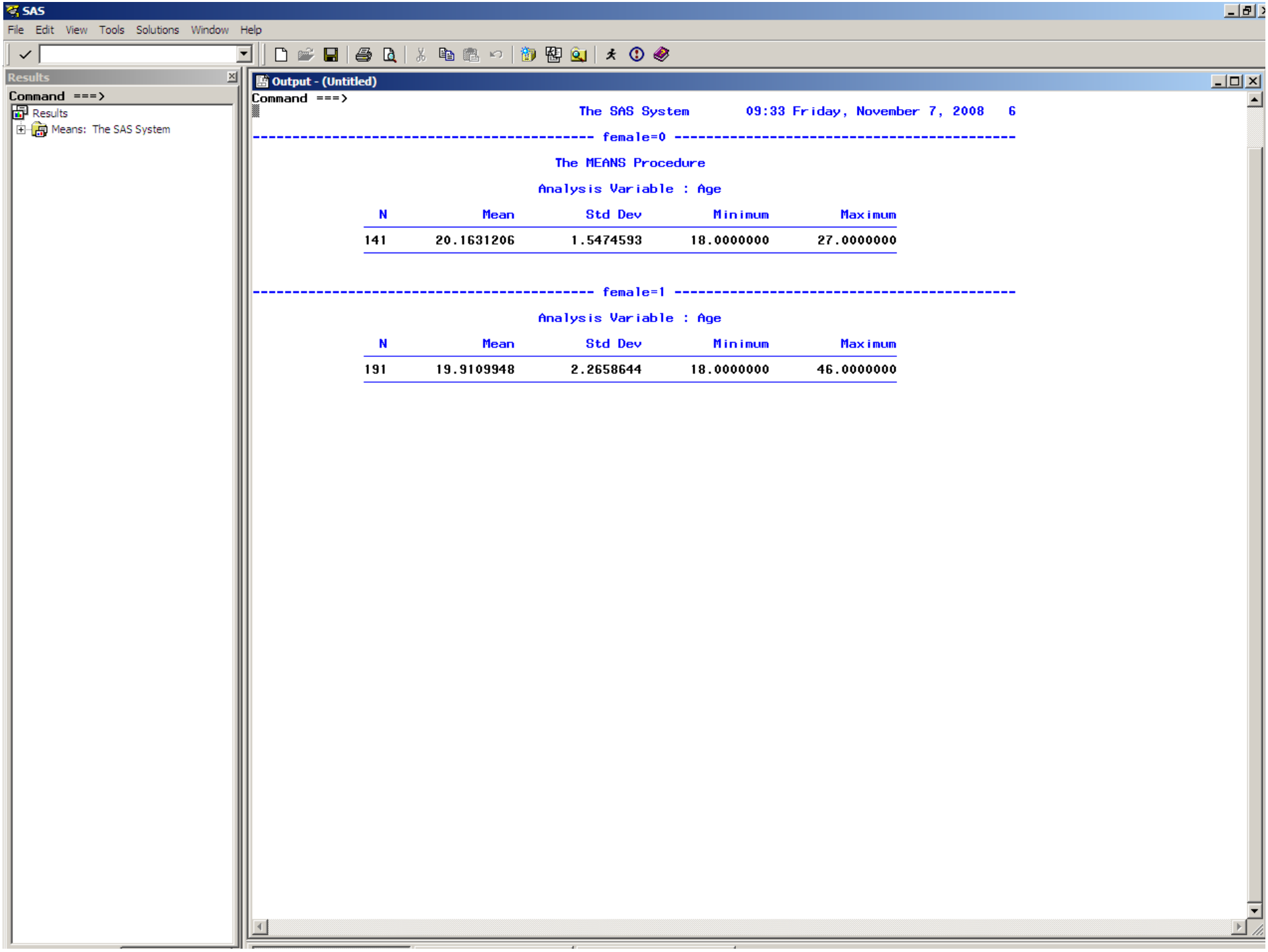

SAS
File Edit View Tools Solutions Window Help
Results
Command ===>
Results
Freq: The SAS System
Freq: The SAS System
Output - (Untitled)
Command ===>
The FREQ Procedure
Table of female by Age
female Age
Frequency
Percent
Row Pct
Col Pct
18 19 20 21 22 Total
0 14 39 39 26 14 141
4.22 11.75 11.75 7.83 4.22 42.47
9.93 27.66 27.66 18.44 9.93
31.11 40.63 45.35 42.62 42.42
1 31 57 47 35 19 191
9.34 17.17 14.16 10.54 5.72 57.53
16.23 29.84 24.61 18.32 9.95
68.89 59.38 54.65 57.38 57.58
Total 45 96 86 61 33 332
13.55 28.92 25.90 18.37 9.94 100.00
(Continued)
Table of female by Age
female Age
Frequency
Percent
Row Pct
Col Pct
23 24 26 27 46 Total
0 5 2 1 1 0 141
1.51 0.60 0.30 0.30 0.00 42.47
3.55 1.42 0.71 0.71 0.00
83.33 100.00 100.00 100.00 0.00
1 1 0 0 0 1 191
0.30 0.00 0.00 0.00 0.30 57.53
0.52 0.00 0.00 0.00 0.52
16.67 0.00 0.00 0.00 100.00
Total 6 2 1 1 1 332
1.81 0.60 0.30 0.30 0.30 100.00
Frequency Missing = 5
Results
Explorer
Output - (Untitled)
Log - (Untitled)
code.sas *


SAS
File Edit View Tools Solutions Window Help
Results
Command ===>
Results
Freq: The SAS System
Freq: The SAS System
Output - (Untitled)
Command ===>
The SAS System 09:33 Friday, November 7, 2008 5
The FREQ Procedure
Statistics for Table of female by Age
Statistic DF Value Prob
Chi-Square 9 13.0596 0.1599
Likelihood Ratio Chi-Square 9 15.1016 0.0882
Mantel-Haenszel Chi-Square 1 1.2971 0.2548
Phi Coefficient 0.1983
Contingency Coefficient 0.1945
Cramer's V 0.1983
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.
Effective Sample Size = 332
Frequency Missing = 5
Results
Explorer
Output - (Untitled)
Log - (Untitled)
code.sas *

Where/By

- “Where” is used when using logical conditions.
- “By” is used to classify groups
  - Sort
- What would be an example?


SAS
File Edit View Tools Run Solutions Window Help
Results
Command ===>
Results
Means: The SAS System
Log - (Untitled)
Command ===>
NOTE: PROCEDURE SORT used (Total process time):
real time 0.76 seconds
cpu time 0.01 seconds
12 proc means; var age;
13 by female;run;
NOTE: There were 337 observations read from the data set A.COLLEGE.
NOTE: PROCEDURE MEANS used (Total process time):
real time 0.28 seconds
cpu time 0.00 seconds
code.sas *
Command ===>
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
data a.college; set a.college;
proc sort; by female;
proc means; var age;
by female;run;
Results
Explorer
Output - (Untitled)
Log - (Untitled)
code.sas *
NOTE: 4 Lines Submitted.
C:\Documents and Settings\hlyons
Ln 5, Col 15


SAS
File Edit View Tools Solutions Window Help
Results
Command ===>
Results
Means: The SAS System
Output - (Untitled)
Command ===>
The SAS System 09:33 Friday, November 7, 2008 6
female=0
The MEANS Procedure
Analysis Variable : Age
N Mean Std Dev Minimum Maximum
141 20.1631206 1.5474593 18.0000000 27.0000000
female=1
Analysis Variable : Age
N Mean Std Dev Minimum Maximum
191 19.9109948 2.2658644 18.0000000 46.0000000

Proc Freq

- Other options
  - Expected (independence test)
  - Agree (McNemar test)
  - Exact (Fisher's Test)

Proc Means

- If you just use proc means you will get the mean, SD, Max, and Min.
- You can also get
  - Missing, N, Nmiss, Mean, STD, Min, Max, Range, Sum, Var, STDERR, T, PRT.

Other Helpful Code

- Comments

code.sas *

Command ===>

```
Libname a "R:\CFDR\CFDR\HEIDI\Workshop 08";run;
*this is a comment*
/*so is this*/
```

Other Helpful Code

- Variable label

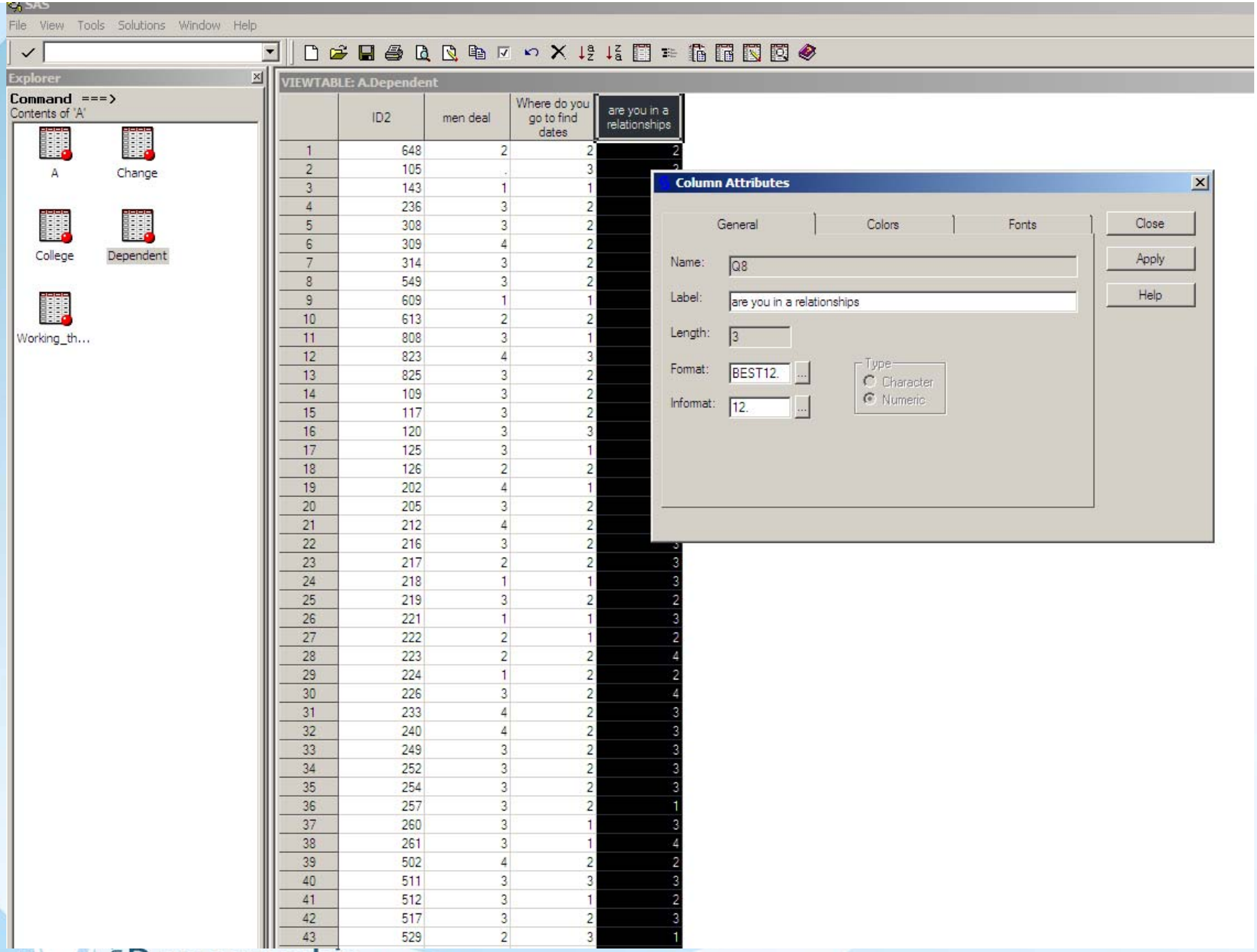
```
data a.dependent ; set a.dependent;
label q1="ID2"
      q4="men deal"
      q6="Where do you go to find dates"
      q8="are you in a relationships";run;
```

- Value label

```
proc format;
    value relationship 1="single"
                       2="married"
                       3="cohab"
                       4="divorced";
    value  attitudes   1="Strongly Agree"
                       2="Agree"
                       3="Disagree"
                       4="Strongly Disagree";run;
```

```
proc freq;
    format q4 attitudes.
           q8 relationship.;
table q4 q8;run;
```

BGSU Center for Family and Demographic Research



	ID2	men deal	Where do you go to find dates	are you in a relationships
1	648	2	2	2
2	105	.	3	
3	143	1	1	
4	236	3	2	
5	308	3	2	
6	309	4	2	
7	314	3	2	
8	549	3	2	
9	609	1	1	
10	613	2	2	
11	808	3	1	
12	823	4	3	
13	825	3	2	
14	109	3	2	
15	117	3	2	
16	120	3	3	
17	125	3	1	
18	126	2	2	
19	202	4	1	
20	205	3	2	
21	212	4	2	
22	216	3	2	
23	217	2	2	3
24	218	1	1	3
25	219	3	2	2
26	221	1	1	3
27	222	2	1	2
28	223	2	2	4
29	224	1	2	2
30	226	3	2	4
31	233	4	2	3
32	240	4	2	3
33	249	3	2	3
34	252	3	2	3
35	254	3	2	3
36	257	3	2	1
37	260	3	1	3
38	261	3	1	4
39	502	4	2	2
40	511	3	3	3
41	512	3	1	2
42	517	3	2	3
43	529	2	3	1

The SAS System 10:27 Wednesday, September 30, 20(

The FREQ Procedure

men deal

Q4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Strongly Agree	24	7.23	24	7.23
Agree	86	25.90	110	33.13
Disagree	197	59.34	307	92.47
Strongly Disagree	25	7.53	332	100.00

Frequency Missing = 5

are you in a relationships

Q8	Frequency	Percent	Cumulative Frequency	Cumulative Percent
single	17	5.09	17	5.09
married	81	24.25	98	29.34
cohab	210	62.87	308	92.22
divorced	26	7.78	334	100.00

Frequency Missing = 3

Other Helpful Code

- Titles
- Footnotes

```
title "attitudes for marrieds";
footnote "both males and females are in this sample";
proc freq; where q8=2; table q4;run;
```

Output - (Untitled)

Command ===>

attitudes for marrieds 13

10:27 Wednesday, September 30, 2009

The FREQ Procedure

men deal

Q4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Strongly Agree	8	9.88	8	9.88
Agree	20	24.69	28	34.57
Disagree	46	56.79	74	91.36
Strongly Disagree	7	8.64	81	100.00

both males and females are in this sample

My Program Will Not Run!

- End each step with a run statement.
- Do you have the full path name?
- Check for a missing semicolon.
- Are all the quotation marks closed?
- Did you close all your comments?
- Check the log for warning and error messages.
- Make sure the number of observations and variables are correct.
- Come ask me!