

# Bringing the SAS Legacy into the R Environment

A Brief Talk for SAS and non-SAS  
Data Professionals



[www.directeffects.net](http://www.directeffects.net)

Georgette Asherman, 201 673-4301

# What is SAS?

- The company
  - SAS, Inc. positions itself now as ‘the leader in business analytics software and services, and the largest independent vendor in the business intelligence market.’ Broad range of enterprise products.
  - Sales in 2008 2.26 bil. Up from 1.13 bil. in 2001
  - 120 countries, 45,000 customer organizations, 92 of the Fortune 100 have a product of theirs.
- The core product-SAS<sup>®</sup> 9.2-An integrated language and analytics package that supports various add-ons in a common environment

# The Legacy



Jim Goodnight, a statistician at North Carolina State, wrote the 'Statistical Analysis System.' He and three others started The SAS Institute as a company in 1976. The company remains privately-held and is known for its full-service employee-friendly campus. They remained in core product but incorporated current technology as it becomes mainstream. This year they introduced an interface with R with their IML<sup>®</sup>/Studio Product.

# Data Structures

SAS Dataset

Obs. No.	Subject	City	Score
1			
2			
3			
4			
5			
6			
7			
8			

R Data Frame

Row Label	Subject	City	Score
1			
2			
3			
4			
5			
6			
7			
8			

# Comparison of R Data frame and SAS Dataset

## SAS Dataset

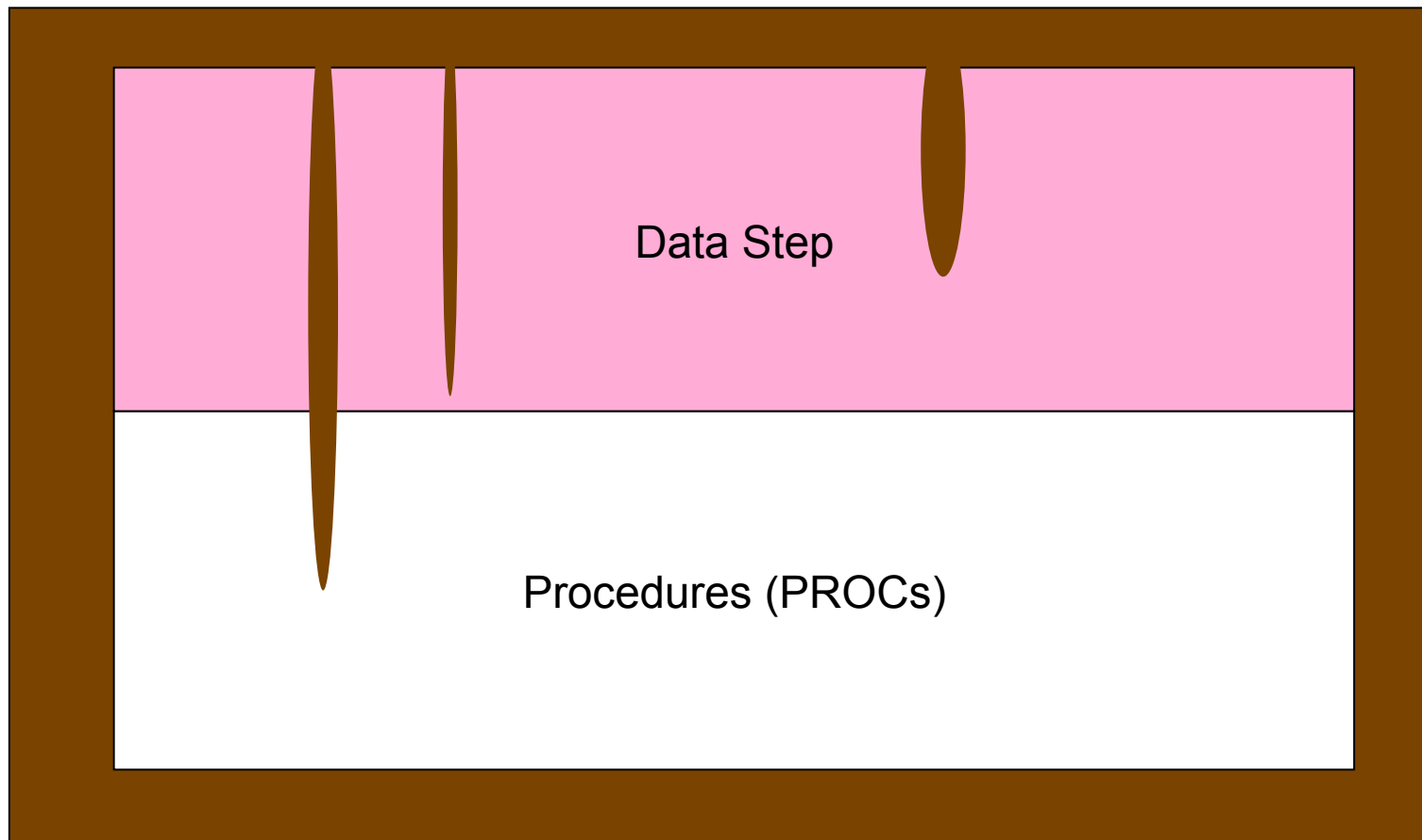
1. Each named column contains one type of data. Really a stack of rows
2. No row names but observation numbers, `_N_`
3. Subsetting via statements such as `IF`, `WHERE`, `KEEP` `DROP`
4. `INFILE` statement and `INPUT` statements work for flat files. `PROC IMPORT` reads in `EXCEL` files.

## R Data Frame

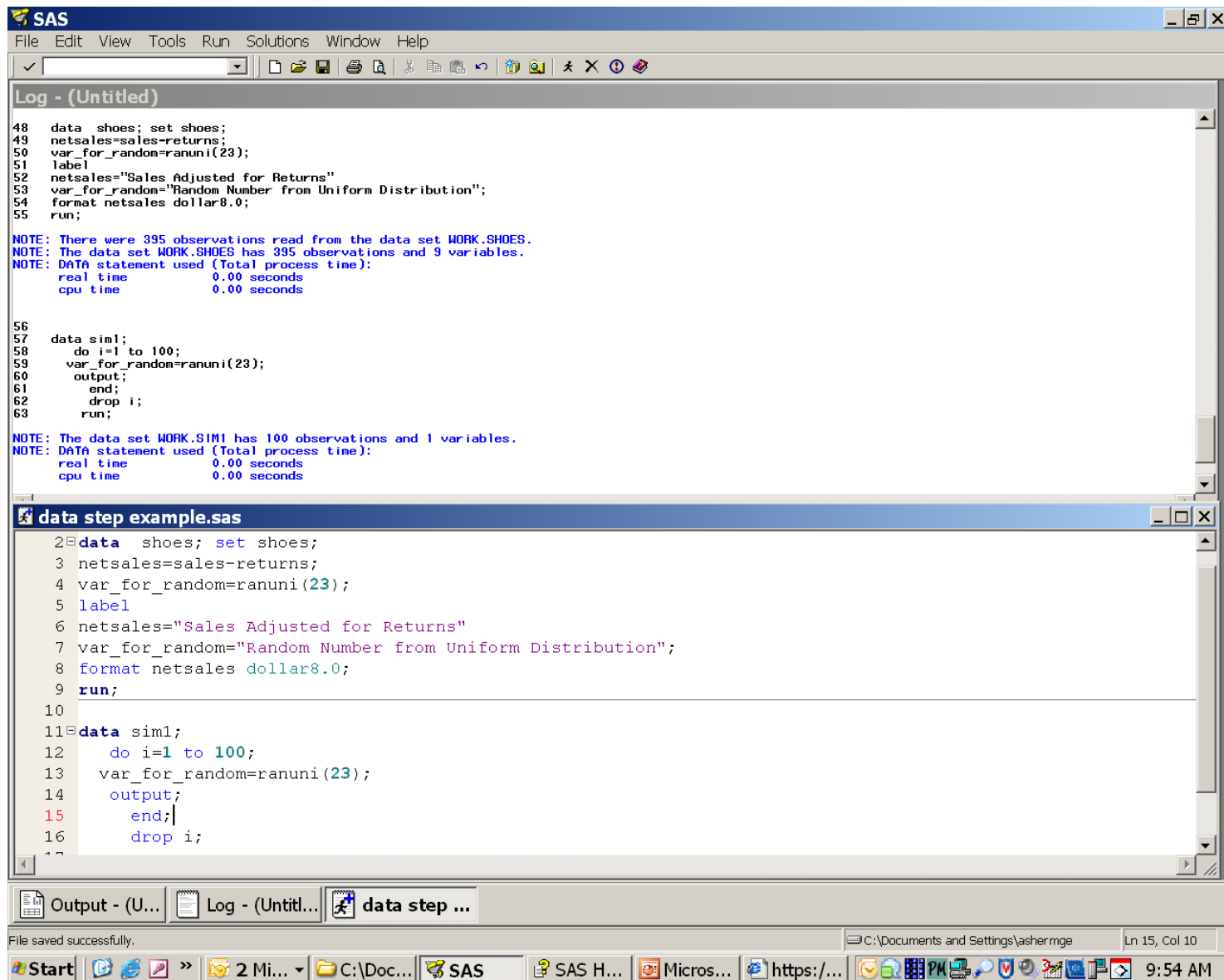
1. Each column contains one type of data. Really a list of named vectors
2. Row names-default to row numbers and retained
3. Subsetting via `[rows, cols]`
4. Extensive functions for reading in flat files and other proprietary data formats

# The SAS Computing Structure

Macro Language and Macro Variables



# A SAS Session in Windows



The screenshot displays the SAS software interface in a Windows environment. The main window is titled "SAS" and contains a "Log - (Untitled)" pane and a "data step example.sas" editor pane. The log pane shows the execution of two DATA steps. The first step, starting at line 48, reads data from the WORK.SHOES dataset, calculates 'netsales' (sales minus returns), and generates a random number for each observation. The second step, starting at line 56, creates a dataset named SIM1 with 100 observations, each containing a random number. The log output includes status messages and timing information for both steps.

```
48 data shoes; set shoes;
49 netsales=sales-returns;
50 var_for_random=ranuni(23);
51 label
52 netsales="Sales Adjusted for Returns"
53 var_for_random="Random Number from Uniform Distribution";
54 format netsales dollar8.0;
55 run;

NOTE: There were 395 observations read from the data set WORK.SHOES.
NOTE: The data set WORK.SHOES has 395 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

56
57 data sim1;
58   do i=1 to 100;
59     var_for_random=ranuni(23);
60     output;
61   end;
62   drop i;
63 run;

NOTE: The data set WORK.SIM1 has 100 observations and 1 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

```
2 data shoes; set shoes;
3 netsales=sales-returns;
4 var_for_random=ranuni(23);
5 label
6 netsales="Sales Adjusted for Returns"
7 var_for_random="Random Number from Uniform Distribution";
8 format netsales dollar8.0;
9 run;
10
11 data sim1;
12   do i=1 to 100;
13     var_for_random=ranuni(23);
14     output;
15     end;
16     drop i;
```

Output - (U... | Log - (Untitl... | data step ...

File saved successfully. C:\Documents and Settings\ashermge Ln 15, Col 10

Start | 2 Mi... | C:\Doc... | SAS | SAS H... | Micros... | https://... | 9:54 AM

# The Data Sets

The screenshot displays the SAS interface with two data tables. The primary table, 'VIEWTABLE: Work.Shoes', contains 44 rows of data with columns for Product, Subsidiary, Number of Stores, Total Sales, Total Inventory, Total Returns, Sales Adjusted for Returns, and Random Number from Uniform Distribution. The secondary table, 'VIEWTABLE: Work.Sim1', contains 44 rows of data with a column for 'var\_for\_random'.

	Product	Subsidiary	Number of Stores	Total Sales	Total Inventory	Total Returns	Sales Adjusted for Returns	Random Number from Uniform Distribution
1	Boot	Addis Ababa	12	\$29,761	\$191,821	\$769	\$28,992	0.2541391059
2	Men's Casual	Addis Ababa	4	\$67,242	\$118,036	\$2,284	\$64,958	0.31204046
3	Men's Dress	Addis Ababa	7	\$76,793	\$136,273	\$2,433	\$74,360	0.1959590401
4	Sandal	Addis Ababa	10	\$62,819	\$204,284	\$1,961	\$60,958	0.9661698437
5	Slipper	Addis Ababa	14	\$68,641	\$279,795	\$1,771	\$66,870	0.1968592909
6	Sport Shoe	Addis Ababa	4	\$1,690	\$16,634	\$79	\$1,611	0.2933790443
7	Women's Casual	Addis Ababa	2	\$51,541	\$98,641	\$940	\$50,601	0.4885209824
8	Women's Dress	Addis Ababa	12	\$108,942	\$311,017	\$3,233	\$105,709	0.2289096248
9	Boot	Algiers	21	\$21,297	\$73,737	\$710	\$20,587	0.1452626028
10	Men's Casual	Algiers	4	\$63,206	\$100,982	\$2,221	\$60,985	0.5310306868
11	Men's Dress	Algiers	13	\$123,743	\$428,575	\$3,621	\$120,122	0.8443271233
12	Sandal	Algiers	25	\$29,198	\$84,447	\$1,530	\$27,668	0.0490219938
13	Slipper	Algiers	17	\$64,891	\$248,198	\$1,823	\$63,068	0.6280691496
14	Sport Shoe	Algiers	9	\$2,617	\$9,372	\$168	\$2,449	0.5452452663
15	Women's Dress	Algiers	12	\$90,648	\$266,805	\$2,690	\$87,958	0.011548725
16	Boot	Cairo	20	\$4,846	\$18,965	\$229	\$4,617	0.8354611755
17	Men's Casual	Cairo	25	\$360,209	\$1,063,251	\$9,424	\$350,785	0.2700711485
18	Men's Dress	Cairo	5	\$4,051	\$45,962	\$97	\$3,954	0.8683815114
19	Sandal	Cairo	9	\$10,532	\$50,430	\$598	\$9,934	0.4655918355
20	Slipper	Cairo	9	\$13,732	\$54,117	\$1,216	\$12,516	0.2096729266
21	Sport Shoe	Cairo	3	\$2,259	\$20,815	\$44	\$2,215	0.8294406565
22	Women's Casual	Cairo	14	\$328,474	\$940,851	\$10,124	\$318,350	0.4935689254
23	Women's Dress	Cairo	3	\$14,095	\$51,145	\$745	\$13,350	0.8467445922
24	Boot	Johannesburg	14	\$8,365	\$33,011	\$483	\$7,882	0.6061890463
25	Sandal	Johannesburg	13	\$17,337	\$63,003	\$809	\$16,528	0.943154016
26	Slipper	Johannesburg	12	\$39,452	\$130,025	\$1,565	\$37,887	0.4347004706
27	Sport Shoe	Johannesburg	8	\$5,172	\$29,368	\$139	\$5,033	0.5940547244
28	Women's Dress	Johannesburg	4	\$42,682	\$120,127	\$966	\$41,716	0.578387393
29	Boot	Khartoum	24	\$19,282	\$105,370	\$700	\$18,582	0.3993751073
30	Men's Casual	Khartoum	1	\$9,244	\$16,230	\$478	\$8,766	0.671307491
31	Men's Dress	Khartoum	3	\$18,053	\$51,132	\$1,177	\$16,876	0.7523158778
32	Sandal	Khartoum	18	\$26,427	\$81,825	\$1,281	\$25,146	0.6518811773
33	Slipper	Khartoum	11	\$43,452	\$143,015	\$1,573	\$41,879	0.4179126622
34	Sport Shoe	Khartoum	7	\$2,521	\$27,041	\$84	\$2,437	0.3697611449
35	Women's Casual	Khartoum	1	\$19,582	\$30,727	\$384	\$19,198	0.562187725
36	Women's Dress	Khartoum	6	\$48,031	\$132,679	\$1,374	\$46,657	0.8361254655
37	Boot	Kinshasa	16	\$13,921	\$70,736	\$553	\$13,368	0.9939759681
38	Men's Dress	Kinshasa	5	\$57,691	\$95,663	\$1,699	\$55,992	0.8658710392
39	Sandal	Kinshasa	10	\$16,662	\$104,438	\$611	\$16,051	0.6290922089
40	Slipper	Kinshasa	11	\$52,807	\$183,937	\$1,440	\$51,367	0.8660320727
41	Sport Shoe	Kinshasa	10	\$4,888	\$27,998	\$162	\$4,726	0.8301003691
42	Women's Casual	Kinshasa	1	\$17,919	\$21,363	\$400	\$17,519	0.035412394
43	Women's Dress	Kinshasa	2	\$22,929	\$97,149	\$717	\$22,211	0.9771592906
44	Women's Dress	Kinshasa	2	\$22,929	\$97,149	\$717	\$22,211	0.9771592906

	var_for_random
1	0.2541391059
2	0.31204046
3	0.1959590401
4	0.9661698437
5	0.1968592909
6	0.2933790443
7	0.4885209824
8	0.2289096248
9	0.1452626028
10	0.5310306868
11	0.8443271233
12	0.0490219938
13	0.6280691496
14	0.5452452663
15	0.011548725
16	0.8354611755
17	0.2700711485
18	0.8683815114
19	0.4655918355
20	0.2096729266
21	0.8294406565
22	0.4935689254
23	0.8467445922
24	0.6061890463
25	0.943154016
26	0.4347004706
27	0.5940547244
28	0.578387393
29	0.3993751073
30	0.671307491
31	0.7523158778
32	0.6518811773
33	0.4179126622
34	0.3697611449
35	0.562187725
36	0.8361254655
37	0.9939759681
38	0.8658710392
39	0.6290922089
40	0.8660320727
41	0.8301003691
42	0.035412394
43	0.8721592906
44	0.4967605614

# Why people love SAS

```
proc contents data  
=sashelp.shoes;  
run;  
data work.shoes; set  
sashelp.shoes;  
    netsales=sales-returns;  
    var_for_random=ranuni(23);  
label  
    netsales="Sales Adjusted for  
Returns"  
    var_for_random="Random Number  
from Uniform Distribution";  
format netsales dollar8.0;  
run;
```

*Language not case sensitive!*

```
load(file = "filesfromSAS.RData")  
shoes$netsales<-shoes$sales-shoes$returns  
set.seed(23)  
shoes$var.for.random<- runif(nrow(shoes))  
#need Hmisc package to add variable  
    labels  
str(shoes)  
#Hmisc also has something more like  
    PROC CONTENTS
```

# Why people hate SAS

```
data sim1;  
  do i=1 to 100;  
var_for_random=ranuni(23);  
  output;  
  end;  
  drop i;  
run;
```

\*Data set with 1 variable and 100 rows.

```
set.seed(23)  
var.for.random<-runif(100)  
#Can do anything with the vector  
  now-rename, add to a list, bind to  
  a dataset, etc.
```

# The SAS Concept of Output

In Version 8.0, SAS introduced the Output Delivery System (ODS). This new language structure freed users from the default text-based list output. It allows the following:

- Send output to RTF, EXCEL, HTML and PDF formats.
- Modify output with template language.
- Modify default output and access requested output as SAS datasets.
- Retained OUTPUT statement for older procedures but new procedures now use ODS statements to make data sets for statistics, residuals and other analysis files.
- New ODS Graphics in Version 9.2.

# Descriptive Statistics

```
*default output from PROC  
PROC MEANS data=work.shoes;  
  RUN;  
*default output from PROC-can use  
  statements to change ;  
  
PROC MEANS data=work.shoes noprint;  
  OUTPUT out=work.meansdata;  
  RUN;  
*send to an OUTPUT data set;  
  
PROC PRINT data =meansdata;
```

```
summary(shoes)  
#default output-use other functions to  
  get different statistics  
output.for.shoes<-summary(shoes)  
#output with class 'table'. Similar to  
  list. Counts for factors and  
  percentiles for numeric variables
```

# Class Level Processing

- There is a BY Statement that requires data grouped by classification level or pre-sorted.
- There is a more flexible CLASS Statement

```
proc means data =shoes mean  
  noprint;  
  var sales;  
  output out=classdata  
  mean=mean ;  
  class region product;  
run;
```

```
proc print data=classdata noobs;  
run;
```

- by function similar to SAS BY statement but does accept a data frame in any ordering. Results in output of class table

```
by.output<-by(shoes[c("Sales")], shoes[c  
  ("Region","Product")],mean)
```

- aggregate results in data frame and tapply in matrix

```
ag.output<-aggregate(shoes$Sales,by=shoes[c  
  ("Region","Product")],mean)
```

```
tapply.output<-tapply(shoes$Sales,shoes[c  
  ("Region","Product")],mean)
```

**Need to write new function in order to get all the levels into one data object.**

# Simple Linear Regression

- Different PROCs available for regression depending on your focus-output a little different from each

Proc GLM;

**Model** Returns=Sales Inventory;

Run;

Proc REG;

**Model** Returns=sales inventory;

Run;

Different PROCs such as GENMOD for non-normal errors such as logistic regression

- lm function – Limited default output so assign to a list object with class “lm”

```
attach(shoes)
```

```
lmexample<-lm>Returns~Sales  
+Inventory)
```

```
summary(lmexample)
```

- Summary function for lm gives nearly the same output as PROC REG.

# Recommended Books

**R For SAS and SPSS Users**, Robert A. Muenchen,  
Springer Science & Business Media: New York, 2009

**Applied Statistics and the SAS Programming Language**, 5<sup>th</sup> ed. Ronald P. Cody and Jeffrey Smith,  
Prentice Hall: New York, 2005

**The Little SAS Book: A Primer**, 4<sup>th</sup> ed. Lora Delwiche and  
Susan Slaughter, SAS Institute: Cary 2008

Documentation and support available on SAS Website

[www.support.sas.com](http://www.support.sas.com)

SAS and IML are registered trademarks of SAS, Inc.

See txt and rtf files for output discussed in this talk.