

## **Labels: What They Are; How To Use Them**

David D. Chapman, Chapman Analytics LLC, Alexandria, Virginia

### **ABSTRACT**

Labels for data sets, variables, and values of variables are useful and valuable to a SAS® programmer. This paper discusses how to assign labels both using code and interactively with SAS explorer. In addition to showing how to assign labels to data sets and to each variable in a data set, the paper discusses why it is useful to do it. It also discusses how to assign variables to the values of variables in the data set. Labels make management of the data set easier for the programmer and use of the data set easier for the user.

### **INTRODUCTION**

All SAS data sets have data set and variable names and values for variables. Some SAS data sets also have labels associated with those names and values. These labels add value to the user and make them more efficient and flexible to program. This note discusses how to assign labels to data sets, to variables in the data sets, and to the values of the variables. It also shows how to identify the value of labels associated with a data set.

### **NAMES**

The names of the data set and the variables that make up the data set are attributes that are unique to a SAS data set. The variables that make up the SAS data set usually have different values for each variable for each record or observation in the SAS data set.

### **DATA SET AND VARIABLE NAMES**

SAS data set and variables names can be:

- 1 to 32 characters long
- Contain numbers, letters, or underscores
- Begin with a letter or underscore

### **USER-DEFINED FORMATS**

User-defined formats are created using PROC FORMAT and can be created for both numeric and character variables. A user-defined format name can be up to 31 characteristics in length. Formats associated with a character variable must be preceded by a "\$". A label created and assigned using PROC FORMAT may be up to 32,767 characters in length. However some procedures use only the first 8 or 16 characters.

### **VALUES FOR VARIABLES**

For every observation or record in a SAS data set, each variable will have a value. Each value of a variable may have a label associated with it. The label is assigned using a format. This format may either be SAS a format or a user-defined format. A user-defined format assigns labels to individual variable values – either numeric or character – either individually or in groups. Example code to illustrate the assignment of numeric and character formats is given below. Knowing the name of the format associated with a variable is equivalent to knowing the values of the labels associated with a variable.

Figure 1: Assigning Labels to Values of Variables

```

248 PROC FORMAT ;
249     VALUE BMI
250     LOW-18.5='Underweight'
251     18.5-24.9='Normal'
252     25.0-29.9='Overweight'
253     30.0 - high="Obese";
NOTE: Format BMI has been output.

        VALUE $GENDER
255     'M'  ='Male '
256     'F'  ='Female'
257     OTHER ='Unk  '
258     ;
NOTE: Format $GENDER has been output.
259 run;
NOTE: PROCEDURE FORMAT used (Total process time):
      real time          0.01 seconds
      cpu time           0.00 seconds

```

## DETERMINING SAS DATA SET NAMES AND LABELS

The use of labels with a SAS data set is optional. Not all SAS data sets have labels. Labels associated new or different information with a SAS data set or variable name, or the values a variable takes. The current value of the labels associated with a SAS data set can be identified with PROC CONTENTS. Interactive determination of labels use SAS explorer will be discussed later. Figure 2 shows an example of PROC CONTENTS output that identify the value of a data set's labels. The "Format" and "Label" columns give the format and label associated with the variable in the variable column. The label column at the top show the label associated with the data set name at the top.

Figure 2: Output of PROC CONTENTS To Determine Value of Labels In SAS Data Set

The CONTENTS Procedure					
Data Set Name	WORK.ADMIT	Observations	21		
Member Type	DATA	Variables	10		
Engine	V9	Indexes	0		
Created	Thursday, August 02, 2012 03:15:00 PM	Observation Length	72		
Last Modified	Thursday, August 02, 2012 03:15:00 PM	Deleted Observations	0		
Protection		Compressed	NO		
Data Set Type		Sorted	NO		
Label	SESUG LABEL				
Data Representation	WINDOWS_64				
Encoding	wlatin1 Western (Windows)				
Engine/Host Dependent Information					
Data Set Page Size	8192				
Number of Data Set Pages	1				
First Data Page	1				
Max Obs per Page	113				
Obs in First Data Page	21				
Number of Data Set Repairs	0				
Filename	C:\Users\AS Temporary Files\TD9560\admit.sas7bdat				
Release Created	9.0202M3				
Host Created	X64_VSPRO				
Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
1	ID	Char	4		
2	Name	Char	14		
3	Sex	Char	1	\$GENDER.	GENDER
4	Age	Num	8		AGE2012
5	Date	Num	8	DATE9.	Day of Month
6	Height	Num	8		HT
7	Weight	Num	8		WT
8	ActLevel	Char	4		
10	BMI	Num	8	BMI.	
9	Fee	Num	8	7.2	

## CREATING LABELS

SAS data set and variable names and the values for variables may have labels. Labels are created or changed either programmatically using procedures and data steps or interactively using SAS explorer. Labels for a data set name or a variable may be up to 256 characters. Labels for variables created with a user-defined format can be up to 32,767 characters. For the purpose of this note, a label is assigned to a value of a variable when a user-defined format is associated with the variable.

All SAS data sets have data set and variable names and values for variables. Some SAS data sets have labels associated with those names and values. These labels add value to the user by making them more efficient and flexible. This note will discuss how to assign labels to data sets, the variables in the data sets, and the values of the variables.

## ASSIGNING LABELS PROGRAMATICALLY

Data set and variables labels and formats can be assigned programmatically using in either a data step or a procedure. Figure 3 below shows how labels and formats are assigned in the data step. This codes used three different statements:

1. LABEL option in the DATA statement to assign the data set label
2. LABEL statement in the data step to assign each variable label
3. FORMAT statement to assign the variable value label using a user-define format.

**Figure 3. Assigning Labels in the Data Step**

```
data WORK.ADMIT ( LABEL='SESUG LABEL' ) ;  
set SASUSER.ADMIT;  
format BMI bmi. SEX $GENDER. DATE DATE9. ;  
BMI=( WEIGHT/ ( HEIGHT*HEIGHT ) ) *703;  
label  
sex      = 'GENDER'  
age      = 'AGE2012'  
DATE     = 'Birthday'  
height   = 'HT'  
weight   = 'WT' ;  
run;
```

This same set of labels label could have been assigned using the PROC DATASETS procedure.

## ASSIGNING LABELS INTERACTIVELY USING SAS EXPLORER

SAS Explorer can be used to both determine the current setting and to change to a new setting for the data set name and variable names and formats. Figure 4 below shows SAS explorer displaying the define SAS libraries and the files in the Work library.

A double left mouse click on the file name “Admit” displays the entire file as seen in Figure 5. A single left mouse click (Figure 6) displaying available options and the selection of view columns shows all variables in the SAS data set along with its current label and format. Highlighting a variable, selecting modify displays a dialog box (Figure 7) that shows the current label and format for that variable. This dialog box can also be used to change the label and format associated with the variable.

Figure 4. Display of SAS Data set Using SAS Explorer

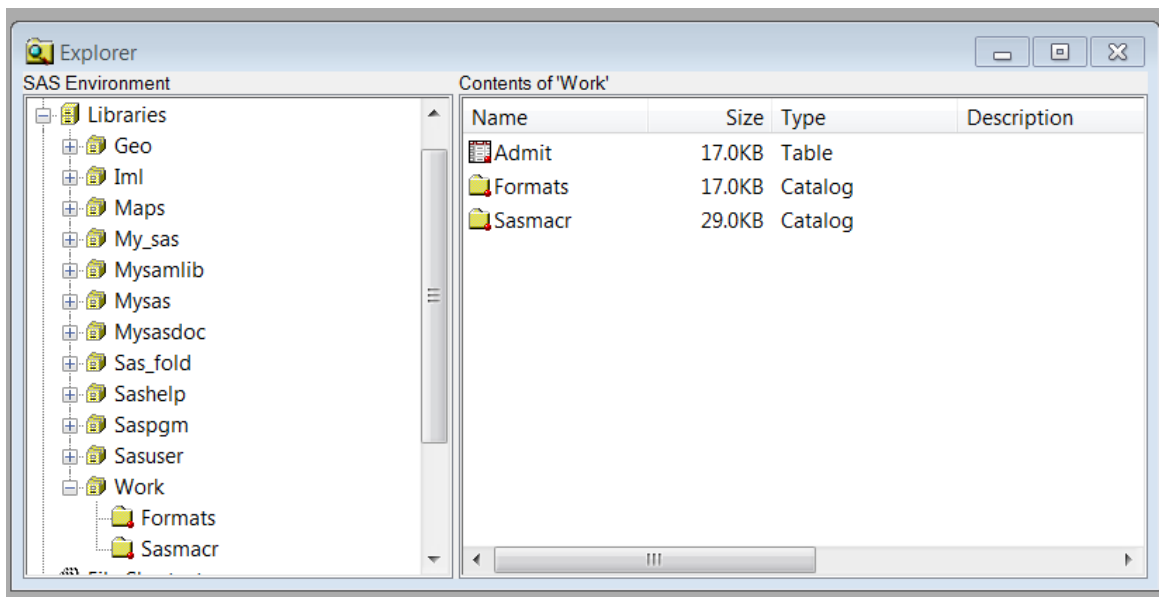


Figure 5. Display of SAS Data Set Using SAS Explorer

The screenshot shows the SAS Explorer window displaying a data table. The table has the following columns: ID, Name, GENDER, AGE2012, Day of Month, HT, WT, ActLevel, Fee, and BMI. The data is as follows:

ID	Name	GENDER	AGE2012	Day of Month	HT	WT	ActLevel	Fee	BMI
1	2458 Murray, W	Male	27	02JAN1960	72	168	HIGH	85.20	Normal
2	2462 Almers, C	Female	34	04JAN1960	66	152	HIGH	124.80	Normal
3	2501 Bonaventure, T	Female	31	18JAN1960	61	123	LOW	149.75	Normal
4	2523 Johnson, R	Female	43	01FEB1960	63	137	MOD	149.75	Normal
5	2539 LaMance, K	Male	51	05JAN1960	71	158	LOW	124.80	Normal
6	2544 Jones, M	Male	29	07JAN1960	76	193	HIGH	124.80	Normal
7	2552 Reberson, P	Female	32	10JAN1960	67	151	MOD	149.75	Normal
8	2555 King, E	Male	35	14JAN1960	70	173	MOD	149.75	Normal
9	2563 Pitts, D	Male	34	23JAN1960	73	154	LOW	124.80	Normal
10	2568 Eberhardt, S	Female	49	28JAN1960	64	172	LOW	124.80	Overweight
11	2571 Nunnally, A	Female	44	20JAN1960	66	140	HIGH	149.75	Normal
12	2572 Oberon, M	Female	28	18JAN1960	62	118	LOW	85.20	Normal
13	2574 Peterson, V	Male	30	07JAN1960	69	147	MOD	149.75	Normal
14	2575 Quigley, M	Female	40	09JAN1960	69	163	HIGH	124.80	Normal
15	2578 Cameron, L	Male	47	06JAN1960	72	173	MOD	124.80	Normal
16	2579 Underwood, K	Male	60	23JAN1960	71	191	LOW	149.75	Overweight
17	2584 Takahashi, Y	Female	43	30JAN1960	65	123	MOD	124.80	Normal
18	2586 Derber, B	Male	25	24JAN1960	75	188	HIGH	85.20	Normal
19	2588 Ivan, H	Female	22	21JAN1960	63	139	LOW	85.20	Normal
20	2589 Wilcox, E	Female	41	17JAN1960	67	141	HIGH	149.75	Normal
21	2595 Warren, C	Male	54	08JAN1960	71	183	MOD	149.75	Overweight

At the bottom of the window, a status bar indicates: "NOTE: Table has been opened in browse mode." and the path "C:\SAS\_Program" is shown.

Figure 6. Identifying SAS Variables in SAS Data Set in SAS Explorer

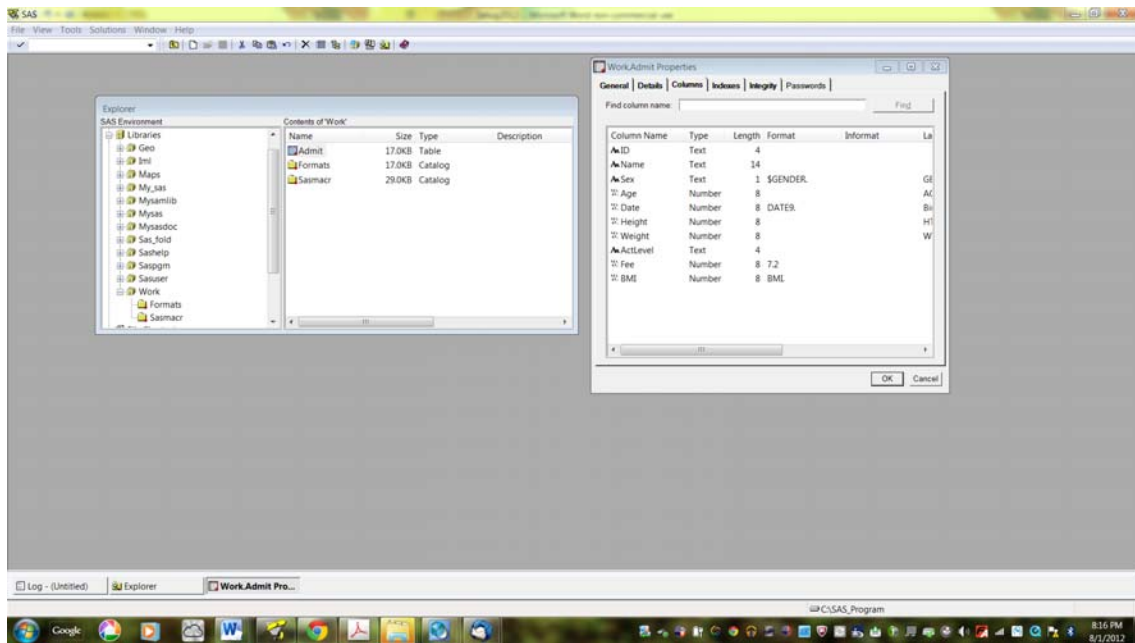
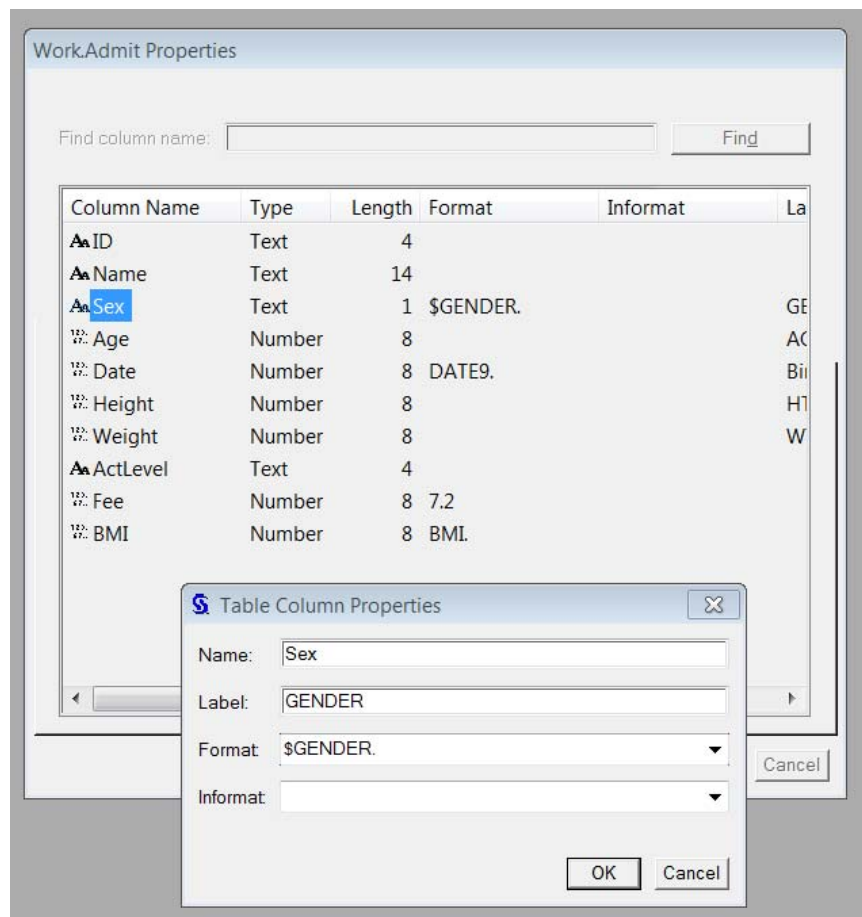


Figure 7. Assigning SAS Variable Labels and User Defined Forms in SAS Explorer



## HOW TO USE LABELS

Labels can be used to improve both the efficiency and quality of the SAS output.

### DATA SET LABELS

A data set label can have 256 characters. This is enough to provide notes on date and type of changes. It can also be used to document the contents of the data set. Documentation can consist of the name of a pdf or html file containing the information. It can tell when, why, or by whom the SAS data set was changed.

### VARIABLE NAME LABELS

Variable labels provide an alternate name that can be changed based on the audience without changing the structure or content of the SAS data set. Changes to variable labels only affect the descriptor portion of the SAS data set. Some procedures such as PROC REPORT and PROC PRINT can, with the use a special option, substitute a variable label for the variable name in their output. Variable names may be generic; but, labels may be customized to English, Spanish, or French audiences.

### VARIABLE VALUE LABELS

Labels associated with the value of a variable are based on formats. Formats are assigned by associating a format with a variable in the descriptor portion of the SAS data set. Formats can be used to rename a variable value or to group multiple values of a variable together.

## CONCLUSION

Labels for a data set name, variable names, and the value of variables can be used to improve both the efficiency and usefulness of the SAS data set. These labels can be created and assigned either in the data step or interactively using SAS explorer.

## RECOMMENDED READING

- SAS 9.3 Language Reference Concepts, Chapter 16, Introduction to SAS Windowing Environment
- SAS 9.3 Language Reference Concepts, Chapter 17, Managing Your Data in the SAS Windowing Environment
- SAS 9.2 BASE SAS PROCEDURES GUIDE, Chapter 25, The FORMAT Procedure

## CONTACT INFORMATION

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

Name: David D. Chapman  
Enterprise: Chapman Analytics LLC  
City, State: Alexandria, Virginia  
Work Phone: (703) 498-5782  
E-mail: chapman.analytics.va@gmail.com

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.

\*\*\*\*\*