# Understanding and Applying Multilabel Formats

PDXSUG

March 20,2002


*Presented By:*

**Andrew H. Karp**

SAS Certified Professional Version 8

**Sierra Information Services, Inc.**

**19229 Sonoma Highway #264**

**Sonoma, California 95476 USA**

**707 996 7380**

**SierraInfo** @ AOL.COM

http://www.**SierraInformation.**com






# Using MULTILABEL Formats

- New to Version 8, MULTILABEL formats allow you to specify both a
  - Primary label
  - Secondary label
    - To values of variables upon which the format is to be applied
  - Not all SAS PROCs support multilabel formats
    - Only PROCs MEANS, SUMMARY, TABULATE and REPORT support MLF's
    - Those that do not (e.g., PRINT, FREQ) utilize only the primary label

# A Simple Example

- Using MLFs in PROC MEANS to analyze the ages of persons on a data set
  - Children
  - Adolescents
  - Adults




# A Simple Example

```
data ages;
length name $ 25;
  input name $ age @@;
  datalines;
  amy 5   fred 10   susan 11   tina 9   john 15   stephanie 16
  thomas 7   ingrid 19   hannah 16   irving 20   dianne 21   mike 18
  mitch 27   maryanne 24   mort 22   debbie 8   peter 2   pamela 30
  ;
  run;

options nonumber nodate;
proc print data=ages;
title 'Using Multilabel Formats';
title2 'Data Set';
run;
```

## Using Multilabel Formats

### Data Set

Data Set

| Obs | name | age |
|---|---|---|
| 1 | amy | 5 |
| 2 | fred | 10 |
| 3 | susan | 11 |
| 4 | tina | 9 |
| 5 | john | 15 |
| 6 | stephanie | 16 |
| 7 | thomas | 7 |
| 8 | ingrid | 19 |
| 9 | hannah | 16 |
| 10 | irving | 20 |
| 11 | dianne | 21 |
| 12 | mike | 18 |
| 13 | mitch | 27 |
| 14 | maryanne | 24 |
| 15 | mort | 22 |
| 16 | debbie | 8 |
| 17 | peter | 2 |
| 18 | pamela | 30 |





# A Simple Example

```
proc format;
value agefmt
low-19 = '0 to 19'
20-high = '20 and older';
run;
```

Creating a "regular" Format

```
proc means n mean median data=ages;
class age;
format age agefmt.;
var age;
title2 'Proc Means with AGEFMT Applied';
run;
```

Associating a "regular" (single label) Format to a Variable

# A Simple Example

**Using Multilabel Formats**

**Proc Means with AGEFMT Applied**

***The MEANS Procedure***

**Analysis Variable : age**

| age | N Obs | N | Mean | Median |
|---|---|---|---|---|
| **0 to 19** | 12 | 12 | 11.3333333 | 10.5000000 |
| 20 and older | 6 | 6 | 24.0000000 | 23.0000000 |





# A Simple Example

```
proc format;
value age2fmt (multilabel)
  0-12 = '1) Child'
  13-19 = '2) Adolescents'
  0 - 19 = '3) Children & Adolescents'
  20 - high = '4) Adults';
  run;
proc means n mean median data=ages;
class age/mlf; * <== apply the mlf;
format age age2fmt.;
var age;
title2 'Proc Means with Multilabel Format AGE2FMT Applied';
run;
```

Creating a Multilabel Format

Applying the Multilabel Format in the CLASS Statement

# A Simple Example

*The MEANS Procedure*

| Analysis Variable : age | | | | |
|---|---|---|---|---|
| age | N Obs | N | Mean | Median |
| 1) Child | 7 | 7 | 7.4285714 | 8.0000000 |
| 2) Adolescents | 5 | 5 | 16.8000000 | 16.0000000 |
| 3) Children & Adolescents | 12 | 12 | 11.3333333 | 10.5000000 |
| 4) Adults | 6 | 6 | 24.0000000 | 23.0000000 |




# A Simple Example

```
proc format;
value age3fmt (multilabel)
  0-12 = 'Child'
  13-19 = 'Adolescents'
  0 - 19 = 'Children & Adolescents'
  20 - high = 'Adults';
  run;
proc means n mean median data=ages;
class age/mlf; * <== apply the mlf;
format age age3fmt.;
var age;
title2 'Proc Means with Multilabel Format AGE3FMT Applied';
run;
```

# A Simple Example

**Using Multilabel Formats**

**Proc Means with Multilabel Format AGE3FMT Applied**

***The MEANS Procedure***

**Analysis Variable : age**

| age | N Obs | N | Mean | Median |
|---|---|---|---|---|
| **Adolescents** | 5 | 5 | 16.8000000 | 16.0000000 |
| Adults | 6 | 6 | 24.0000000 | 23.0000000 |
| Child | 7 | 7 | 7.4285714 | 8.0000000 |
| Children & Adolescents | 12 | 12 | 11.3333333 | 10.5000000 |




# A Work Around

- The next section shows a fairly simple “work around” that can assist in applying MLFs

# Using MULTILABEL Formats

- Example:
  - create a MLF (Multilabel Format) for the two digit variable in ADVREPT.TRANSACTIONS representing credit card type.
  - This Data Set contains 336,000 observations, each representing a single credit card transaction.
    - Used in the “Advanced Reporting with the Output Delivery System” seminar




# Using MULTILABEL Formats

```
proc format library=advrept;
   value $cardf (multilabel)
   'VC' = 'Visa Classic'
   'MC' = 'Master Card Classic'
   'VP' = 'Visa Platinum'
   'MP' = 'Master Card Platinum'
   'MG' = 'Master Card Gold'
   'VG' = 'Visa Gold'
   'VC','VP','VG' = 'Total, Visa'
   'MC','MG','MP' = 'Total, Master Card';
   run;
```

Primary Label

Secondary Label

# Using MULTILABEL Formats

```
options fmtsearch=(advrept) nonumber nodate nocenter;
proc freq data=advrept.transactions;
tables cardtype;
format cardtype $cardf.;
title 'Advanced Reporting';
title2 'PROC FREQ OUTPUT';
run;
```

Since PROC FREQ does not support the use of multilabel formats, only the primary label is used when format $CARDF. is applied in this task.




# Using MULTILABEL Formats

Advanced Reporting
PROC FREQ OUTPUT

The FREQ Procedure

Result of using a multilabel format with a PROC FREQ, which does not support this capability. Only the primary labels are utilized.

Credit Card Type

| cardtype | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| Master Card Classic | 52798 | 20.32 | 52798 | 20.32 |
| Master Card Gold | 86611 | 33.34 | 139409 | 53.67 |
| Master Card Platinum | 34746 | 13.38 | 174155 | 67.04 |
| Visa Classic | 52733 | 20.30 | 226888 | 87.34 |
| Visa Gold | 14665 | 5.65 | 241553 | 92.99 |
| Visa Platinum | 18221 | 7.01 | 259774 | 100.00 |

# Using MULTILABEL Formats

```
proc means data=advrept.transactions
    maxdec = 2 mean sum p50 nway;
format cardtype $cardf. charge_amount dollar16.2;
class cardtype/mlf;
var charge_amount;
output out=x mean= sum= p50=/autoname;
run;

proc print data=x;
title2 'Multilabel Formats: Data Set Created by PROC MEANS';
run;
```

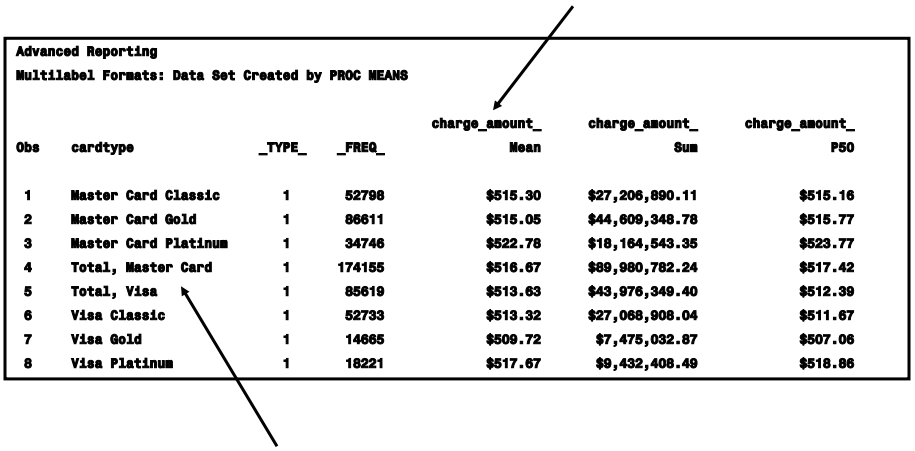



# Using MULTILABEL Formats

Advanced Reporting
Multilabel Formats: Data Set Created by PROC MEANS
The MEANS Procedure
Analysis Variable : charge_amount Transaction Charge Amount

| Credit Card Type | N Obs | Mean | Sum | 50th Pctl |
|---|---|---|---|---|
| Master Card Classic | 52798 | 515.30 | 27206890.11 | 515.16 |
| Master Card Gold | 86611 | 515.05 | 44609348.78 | 515.77 |
| Master Card Platinum | 34746 | 522.78 | 18164543.35 | 523.77 |
| Total, Master Card | 174155 | 516.67 | 89980782.24 | 517.42 |
| Total, Visa | 85619 | 513.63 | 43976349.40 | 512.39 |
| Visa Classic | 52733 | 513.32 | 27068908.04 | 511.67 |
| Visa Gold | 14665 | 509.72 | 7475032.87 | 507.06 |
| Visa Platinum | 18221 | 517.67 | 9432408.49 | 518.86 |

# Using MULTILABEL Formats



Advanced Reporting
Multilabel Formats: Data Set Created by PROC MEANS

| Obs | cardtype | _TYPE_ | _FREQ_ | charge_amount_ Mean | charge_amount_ Sum | charge_amount_ P50 |
|---|---|---|---|---|---|---|
| 1 | Master Card Classic | 1 | 52798 | $515.30 | $27,206,890.11 | $515.16 |
| 2 | Master Card Gold | 1 | 86611 | $515.05 | $44,609,348.78 | $515.77 |
| 3 | Master Card Platinum | 1 | 34746 | $522.78 | $18,164,543.35 | $523.77 |
| 4 | Total, Master Card | 1 | 174155 | $516.67 | $89,980,782.24 | $517.42 |
| 5 | Total, Visa | 1 | 85619 | $513.63 | $43,976,349.40 | $512.39 |
| 6 | Visa Classic | 1 | 52733 | $513.32 | $27,068,908.04 | $511.67 |
| 7 | Visa Gold | 1 | 14665 | $509.72 | $7,475,032.87 | $507.06 |
| 8 | Visa Platinum | 1 | 18221 | $517.67 | $9,432,408.49 | $518.86 |




# Using MULTILABEL Formats

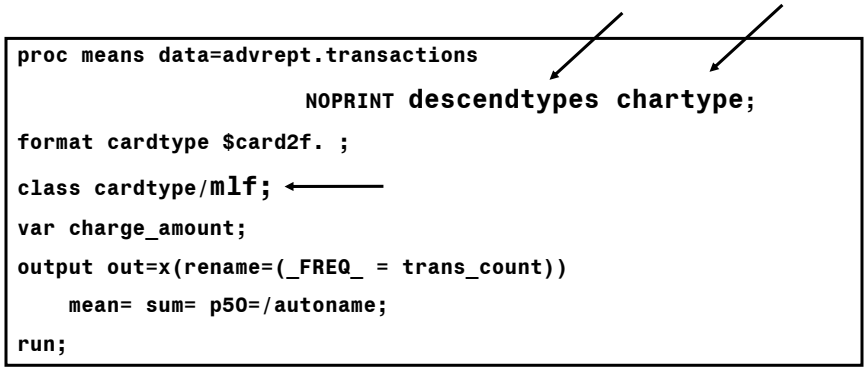
- A Work-Around if the Formatted Values are not portrayed correctly
  - A new format is created. Each value label has a number in the first column, so that the they are portrayed in the desired order.
  - Two Version 8 enhancements are used in the PROC MEANS task
    - **DESCENDTYPES**: orders the observations in the output data set by descending value of _TYPE_, so that _TYPE_ = 0 is the last observation
    - **CHARTYPE**: converts the default numeric variable _TYPE_ to a character variable.

# Using MULTILABEL Formats

```
proc format library=advrept;
   value $card2f (multilabel)
   'VC' = '5Visa Classic'
   'MC' = '1Master Card Classic'
   'VP' = '7Visa Platinum'
   'MP' = '3Master Card Platinum'
   'MG' = '2Master Card Gold'
   'VG' = '6Visa Gold'
   'VC','VP','VG' = '8Subtotal, Visa'
   'MC','MG','MP' = '4Subtotal, Master Card';
   run;
```




# Using MULTILABEL Formats



# Using MULTILABEL Formats

Advanced Reporting
Multilabel Formats: Data Set Created by PROC MEANS
A Work-Around for Problems with MLFs

| Obs | cardtype | _TYPE_ | trans_ count | charge_ amount_ Mean | charge_ amount_Sum | charge_ amount_ P50 |
|---|---|---|---|---|---|---|
| 1 | 1Master Card Classic | 1 | 52798 | 515.302 | 27206890.11 | 515.160 |
| 2 | 2Master Card Gold | 1 | 86611 | 515.054 | 44609348.78 | 515.770 |
| 3 | 3Master Card Platinum | 1 | 34746 | 522.781 | 18164543.35 | 523.765 |
| 4 | 4Subtotal, Master Card | 1 | 174155 | 516.671 | 89980782.24 | 517.420 |
| 5 | 5Visa Classic | 1 | 52733 | 513.320 | 27068908.04 | 511.670 |
| 6 | 6Visa Gold | 1 | 14665 | 509.719 | 7475032.87 | 507.060 |
| 7 | 7Visa Platinum | 1 | 18221 | 517.667 | 9432408.49 | 518.860 |
| 8 | 8Subtotal, Visa | 1 | 85619 | 513.628 | 43976349.40 | 512.390 |
| 9 | | 0 | 259774 | 515.668 | 133957131.64 | 515.730 |



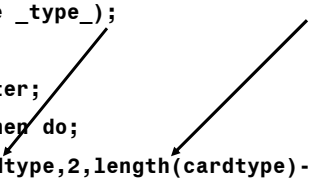

# Using MULTILABEL Formats

- Now that the observations in the output data set are ordered correctly, all we need to do is
  - Strip off the first column of the variable CARDTYPE
  - Replace the blank space where _TYPE_ = 0 with <<< GRAND TOTAL >>>
  - These are both accomplished in the Data Step on the next page.

# Using MULTILABEL Formats

The LENGTH programming language function returns the byte-length of a character variable. In the Assignment Statement creating CARD, the SUBSTR function starts at the second byte of CARDTYPE and the length of the substring is the byte length of CARDTYPE minus 1.

```
* clean up card type;
data x2(drop=cardtype _type_);
  set x;
  * _TYPE_ is character;
   if _type_ = '1' then do;
   card = substr(cardtype,2,length(cardtype)-1) ;
     end;
   else do;
    card = '<<< Grand Total >>>';
           end;
  run;
```







# Using MULTILABEL Formats

```
proc print data=x2;
format trans_count comma12. charge_amount_mean
       charge_amount_sum charge_amount_p50 dollar16.2;
id card;
title3 'Result of Work-Around for MLFs';
run;
```

# Using MULTILABEL Formats

Advanced Reporting
Multilabel Formats: Data Set Created by PROC MEANS
Result of Work-Around for MLFs

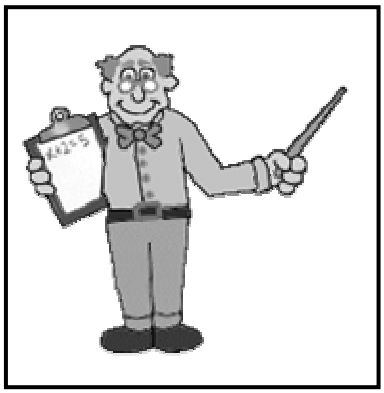
| card | trans_count | charge_amount_ Mean | charge_amount_ Sum | charge_amount_ P50 |
|---|---|---|---|---|
| Master Card Classic | 52,798 | $515.30 | $27,206,890.11 | $515.16 |
| Master Card Gold | 86,611 | $515.05 | $44,609,348.78 | $515.77 |
| Master Card Platinum | 34,746 | $522.78 | $18,164,543.35 | $523.77 |
| Subtotal, Master Card | 174,155 | $516.67 | $89,980,782.24 | $517.42 |
| Visa Classic | 52,733 | $513.32 | $27,068,908.04 | $511.67 |
| Visa Gold | 14,665 | $509.72 | $7,475,032.87 | $507.06 |
| Visa Platinum | 18,221 | $517.67 | $9,432,408.49 | $518.86 |
| Subtotal, Visa | 85,619 | $513.63 | $43,976,349.40 | $512.39 |
| <<< Grand Total >>> | 259,774 | $515.67 | $133,957,131.64 | $515.73 |




# Comments and Conclusions

- MLF's are an important enhancement to Version 8 of the SAS System, but they have several limitations
  - The NOSORTED option in PROC FORMAT does NOT work with MLF's
    - This may be changed in Version 9
  - Output data sets created by MLF-supported PROCs do not store the observations in the "right order"
    - The work around shown in this presentation helps
    - This may also be "fixed" in Version 9.

# Thanks for Attending!



**Questions?**

**Comments?**