

Utilize Dummy Variables/Datasets in Graph Generation

Amos Shu, AstraZeneca Pharmaceuticals, LP

ABSTRACT

The input SAS dataset does not always perfectly contain all necessary data that can be used for graph generation in the real world. Creating dummy variables or datasets is an effective way, sometimes the only way, to generate the desired graph. This paper discusses three different ways to utilize dummy variables/datasets in SAS visualization.

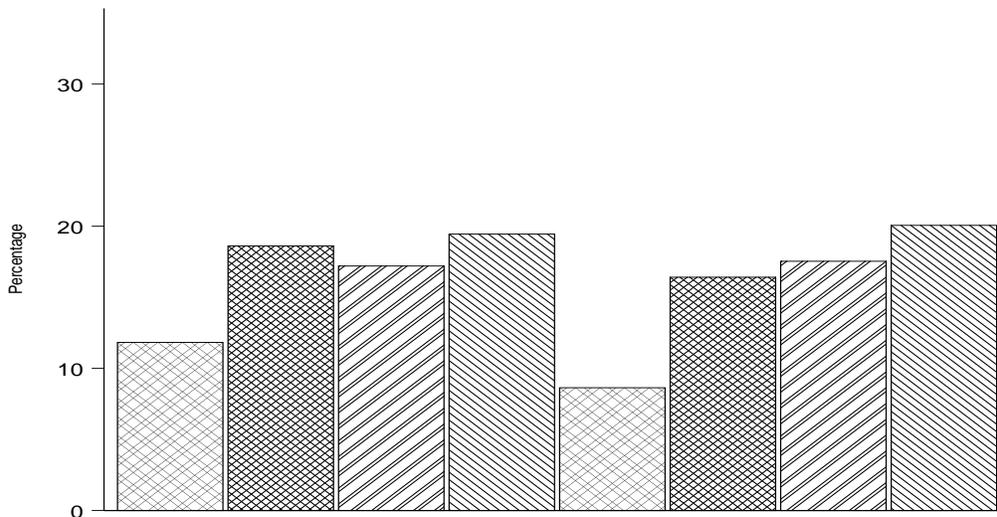
INTRODUCTION

To generate a desired graph in SAS, sometimes the input dataset alone is not enough. Adding dummy variables or dummy datasets is necessary to program graphs in some situations. There are three different ways to utilize dummy variables/datasets to generate the desired graph.

All programs presented in this paper were developed in Server SAS® 9.4 in the Windows environment.

1. BARCHART POSITION AND GROUPING

The following is a bar chart, along with the input dataset and the SAS program behind it. There are two subgroups in the input dataset, cat1=1 and cat1=2. The bar chart does not distinguish the two subgroups as intended because the vertical bars are evenly distributed along the X-axis.



cat1	cat2	x_value	y_value
1	1	2	12
1	2	3	19
1	3	4	17
1	4	5	19.5
2	1	8	9
2	2	9	16
2	3	10	17.5
2	4	11	20

```

LAYOUT OVERLAY / CYCLEATTRS=TRUE
  YAXISOPTS=(LABEL='Percentage' OFFSETMIN=0 TICKVALUEATTRS=(SIZE=12)
  LINEAROPTS=(TICKVALUELIST=(0 10 20 30) VIEWMIN=0 VIEWMAX=35))
  XAXISOPTS=(LABEL=' ' DISPLAY=(LINE))
  LINEAROPTS=(TICKVALUELIST=(1 2 3 4 5 8 9 10 11 15) VIEWMAX=16);

BARCHART X=x_value Y=EVAL(IFN(cat2=1, y_value, .))/GROUP=cat1
  GROUPDISPLAY=CLUSTER BARWIDTH=1 CLUSTERWIDTH=0.95
  OUTLINEATTRS=(COLOR=BLACK) FILLATTRS=(COLOR=WHITE)
  DISPLAY= ALL FILLPATTERNATTRS=(PATTERN=X5);

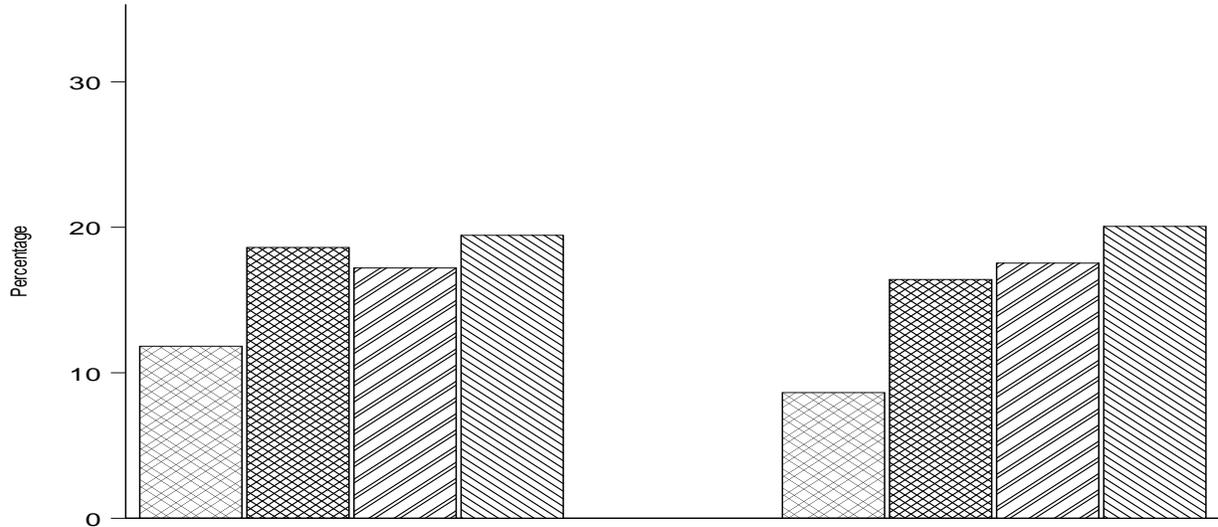
```

Even though there is the GROUPDISPLAY=CLUSTER option in the BARCHART statement and the values of variable "x_value" for the two cat1 subgroups are significantly different (2, 3, 4, 5 for cat1=1 and 8, 9, 10, 11 for cat1=2), surprisingly the SAS program fails to produce the visual of two clusters of bars with sufficient space between them for easy differentiation.

To resolve this issue, some dummy data can be added to the input dataset. The values of dummy data for X-axis should be between the two subgroups, and values of dummy data for Y-axis should be zero. Here is the updated input dataset containing dummy data.

cat1	cat2	x_value	y_value
1	1	2	12
1	2	3	19
1	3	4	17
1	4	5	19.5
		6	0
		7	0
2	1	8	9
2	2	9	16
2	3	10	17.5
2	4	11	20

Without changing the code, we get the figure below that has two clear clusters representing the two subgroups.



2. DISPLAY LEGEND

When using DISCRETEATTRMAP block to display a group of specified legends or make up a legend without plots, a dummy variable is needed to display the legends. For example, the variable TEMP below is a dummy variable with an assigned value of 0. The variable PLOTORDC also could be a dummy variable depending on your actual data and how many legends you would like to display. The second VECTORPLOT statement will draw the legends only while the first VECTORPLOT statement will draw the vector plots for each subject.

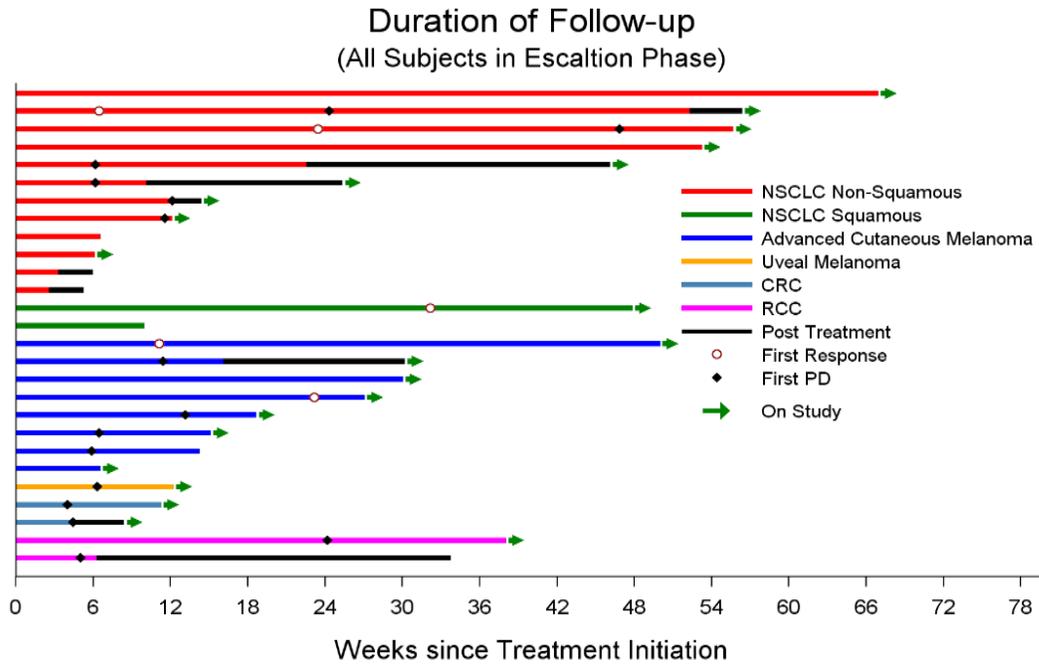
```
DISCRETEATTRMAP NAME="GLINE" / IGNORECASE=TRUE ;
    VALUE "NSCLC NON-SQUAMOUS" / LINEATTRS =(COLOR=RED2 ) ;
    VALUE "NSCLC SQUAMOUS" / LINEATTRS =(COLOR=GREEN) ;
    VALUE "ADVANCED CUTANEOUS MELANOMA" / LINEATTRS =(COLOR=BLUE ) ;
    VALUE "UVEAL MELANOMA" / LINEATTRS =(COLOR=ORANGE) ;
    VALUE "CRC" / LINEATTRS =(COLOR=STEELBLUE ) ;
    VALUE "RCC" / LINEATTRS =(COLOR=MAGENTA) ;
ENDDISCRETEATTRMAP ;
DISCRETEATTRVAR ATTRVAR= GLINES VAR= PLOTORDC ATTRMAP="GLINE" ;

temp= 0;

VECTORPLOT X=MAXWKS Y= Y_Value XORIGIN=MINWKS YORIGIN= Y_Value
    /GROUP=SUBJID ARROWHEADS= FALSE
    LINEATTRS =(THICKNESS=4PX PATTERN=SOLID COLOR=BLACK);

VECTORPLOT X=MAXWKS Y=TEMP XORIGIN=MINWKS YORIGIN=TEMP / GROUP=GLINES ARROWHEADS=
FALSE LINEATTRS =(THICKNESS=4PX PATTERN=SOLID) SCALE=0.0001 NAME='GTYPE' ;

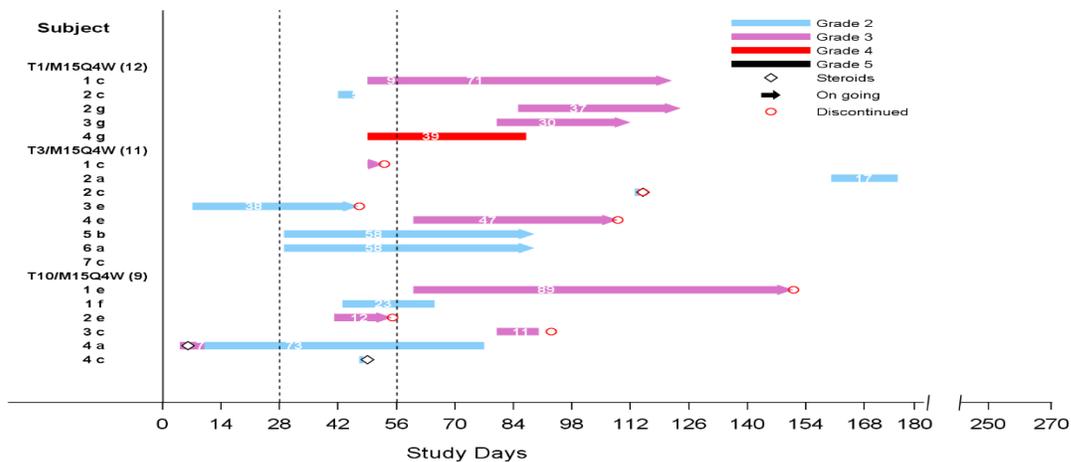
DISCRETELEGEND 'GTYPE' / VALUEATTRS=(SIZE=10) LOCATION=INSIDE BORDER=FALSE
HALIGN= RIGHT ACROSS=1 PAD=( BOTTOM=30PX);
```



3. DETERMINE POSITIONS IN AXIS

Sometimes an axis has no tick values but still needs to be displayed. The following example displays adverse events on Y-axis although there is no tick values for the Y-axis. In this case, you need to create a dummy variable that contains distinguished tick values to display each adverse event clearly on the Y-axis. One method is to use the system variable `_N_`, which contains observation numbers from 1 through n, to derive a value (e.g. `_N_/10`) for each adverse event within each subject.

```
VECTORPLOT X=AEENDY Y=EVAL(IFN(AESEVN=4, Y_VALUE, .))
           XORIGIN=AESTDY YORIGIN= Y_VALUE /GROUP=AHQCAT
ARROWHEADS= FALSE LINEATTRS =(THICKNESS=8PX PATTERN=SOLID COLOR = RED2);
```



^a Dermatitis, ^b Endocrinopathy, ^c Enterocolitis/Colitis, ^d Hepatitis/Hepatic Toxicity,
^e ILD/ Pneumonitis, ^f Neuromuscular Toxicity/ Neuropathy, ^g Pancreatitis, ^h Renal Events.

CONCLUSION

Creating a dummy variable or dataset is easy and simple. But understanding when and how to create a variable or dummy dataset in your programming process is important in graph generation.

CONTACT INFORMATION

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

Amos Shu

AstraZeneca Pharmaceuticals, LP

101 Orchard Ridge Dr.

Gaithersburg, MD 20878

Email: amos.shu@astrazeneca.com

TRADEMARK INFORMATION

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.