CREATING DOT DIAGRAMS
Karen G. Malley, ARC Professional Services Group, Inc.

ABSTRACT
The Dot Diagram macro is used to create a type of frequency plot called a dot diagram, similar in concept to a histogram. Each dot on the picture represents an observation of a numeric response variable, thus showing the general shape of the distribution. The horizontal axis of the diagram represents discrete groups into which each observation is categorized (see Figure 1).

Dot diagrams are suited to a small number of groups. The macro will print one dot per observation with non-missing group and response values. Dots are centered over the group label. If you have a large number of groups (more than ten), or a large number of observations (more than 1000), a histogram is recommended, although this macro will still work. PROC CHART or PROC GCHART will produce histograms.

Two pictures may be produced: one from PROC GPLOT suitable for use in publications, and one from PROC PLOT to give the user a rough idea of the distribution by group.

GROUPING VARIABLE
The input data set should contain a numeric or character grouping variable. Groups will be ordered by the formatted values of the grouping variable. No sorting is necessary prior to calling the macro.

Example
The %DOTDIAG statement is used once for every picture to be displayed. At a minimum, the user must specify the YVAR parameter to identify the response variable. Also, the LABELS parameter must be specified if GPLOT output is being produced. All other parameters have appropriate defaults.

%DOTDIAG(DATAIN=DATA, YVAR=Y, LABELS='DC' 'MD' 'VA', GPLOT=YES, HLENGTH=2.6 IN., HVALUE=F NONE) OPTIONS CHARTYPE=1 DEVICE=HPLJSP2; AXIS2 LENGTH=1.6 IN LABEL=None MINOR=None ORIGIN=(4 PCT,75 PCT); TITLE J=L F NONE H=1 'No. Children in Registered'; TITLE2 J=L 'Private-Home Day Care'; TITLE3 J=L 'Survey Results by State'; TITLE4 J=L '==' Invented Data '==' RUN;

Figure 2 Example: using the macro
The user should supply a RUN statement after the macro call. This statement was purposely omitted from the macro definition to allow the user to customize the macro’s GPLOT step by adding other SAS/GRAPH statements (see Figure 2).

PARAMETERS

All parameters are keyword parameters, so they may be specified in any order, separated by commas. The parameters that may be used on the %DOTDIA statement are listed below in alphabetical order:

DATAIN = data set name
Example: DATAIN=SUBSET

Optional. Specifies the name of the SAS data set to be used by the macro. The name may be the two-level name of a permanent SAS data set (such as IN.SMOKERS). If the DATAIN parameter is omitted, the %DOTDIA macro uses the most recently created data set (_LAST_).

DOTSIZE = LARGE | SMALL

Optional. Specifies the use of small- or large-sized dots in publication-quality pictures. If the user’s data set has few observations, large size dots are preferable, but particularly with laser printer graphics, the CPU time increases with the number of observations. The default is LARGE. For PROC PLOT pictures, this parameter is ignored.

The LARGE sized plotting symbol is a heavy dot, which shows up well and cannot be mistaken for a smear from a photocopier. However, the dots are fairly wide (see Figure 1), so if a lot of observations are to be plotted they can run together to form bars, rather than discrete dots. Figure 3 shows the same picture with small dots.

Users of the macro may wish to modify the SYMBOL statement to a particular device driver.

GPLOT = NO | YES

Optional. Specifies whether publication-quality pictures are to be produced. If this parameter is not specified, or GPLOT=NO is specified, only PROC PLOT output will be produced.

GROUP = grouping variable name
Example: GROUP=AGEGRP

Optional. Specifies the name of a single categorical variable to appear on the horizontal axis. This variable may be either character or numeric. See “Grouping Variable” above. The default group variable name is GROUP.

GRPSPACE = number
Example: GRPSPACE=30

Optional. Specifies the distance between groups on the horizontal axis. The number is given in units internal to the macro, with no relationship to the actual group or response values. The default is 10. Increase the number if the frequency counts are high enough to cause the dots to run into another group’s dots (try 30). For PROC PLOT pictures, this parameter is ignored.

HVALUE = value
Example: HVALUE=F=COMPLEX H=1.2

Optional. Specifies an alternate font and height for the text on the horizontal axis. The default font is the hardware font, and the default height is 0.7. If you change the font, the height will automatically change to 1.0, so it is recommended that both font and height be specified. The example shown above changes these values to font COMPLEX, height 1.2. See the SAS/GRAPH User’s Guide under the AXIS statement VALUE option for other features of the axis values that may be changed. For PROC PLOT pictures, this parameter is ignored.

LABELS = 'string1' 'string2' 'string3' ...
Example: LABELS = "CHRONIC FATIGUE" "CARCINOMA" "AM BURKITT'S" "AF BURKITT'S" "ROSAI DORFMAN"

No. Children in Registered Private-Home Day Care Survey Results by State

Figure 3 Dot Diagram, small dots
Not needed for PROC PLOT pictures, although it will not cause an error if LABELS = is specified, but the specifications will be ignored. Required only if publication-quality plots are being requested.

Specifies one label for each group, to appear on the horizontal axis. These specifications may continue onto multiple lines. There are no default labels, although a picture will still be printed: if this parameter is not specified, the graphics output will show the values of a variable _X (used in the macro to represent the position of the group code on the horizontal axis), and an error message is issued.

IT IS THE USER'S RESPONSIBILITY TO BE AWARE OF THE ORDER OF THE GROUPS. When in doubt, obtain a frequency table of the group variable's values. Reordering the labels using the LABELS parameter WILL NOT reorder the groups in the picture; instead, it simply mislabels the groups.

YLENGTH = number
Example: YLENGTH=4.5
Optional. Specifies the length of the vertical axis in inches. The default is 3.0. Increase this number to obtain more vertical distance between rows of dots. For PROC PLOT pictures, this parameter is ignored.

YVAR = variable name
Example: YVAR=GRESULT
Required regardless of picture type. Specifies a single response variable name. This variable must be numeric. There is no default. The values of this variable will appear on the vertical axis, with tick marks chosen by PRoE PLOT and SAS/GRAPH in the macro. If this variable has many levels, consider grouping them. The user may supply an AXIS2 statement to override the one provided by the macro.

ERRORS
If publication-quality graphics are being produced, the macro will not produce a picture if the frequency counts for a group are so large that they would run into the dots from the next group. An error message is printed, along with the suggestion that the GRPSPACE parameter value be increased.

If you have many groups (more than 5, say), consider extending the horizontal axis by adding a statement such as the following to the program:

GOPTIONS HSIZE = 17 HPOS = 160;
This makes the picture 17 inches wide and increases the number of horizontal positions to 160 to allow for more groups.

MISSING VALUES
If either the grouping variable or the response variable are missing, they will be excluded from the picture. Special missing values are also excluded.

MACRO DEFINITION CODE

%macro DOlDlAGI
DATAIN= _LAST_, /* data set name for analysis */
DOTSIZE=LARGE, /* large or small dots */
DOTSPACE=0.5, /* distance between dots */
GPLOT=NO, /* create a GPLOT picture? */
GROUP=GROUP, /* name of grouping variable */
GRPSPACE=10, /* distance between groups */
HVALUE=H=.7, /* specs for group labels */
MLENGTH=3 IN, /* length of vertical axis */
HLENGTH=6 IN, /* length of horizontal axis */
YVAR=, /* name of Y axis (analysis) variable */
LABELS = /* group labels */
)
%
let TOOmany=0; /* flag too many dots to fit width */
let GPLOT=%upcase(&GPLOT);
* Check for missing LABELS option;
%if &GPLOT=NO and &LABELS= %STR() %then
%do;
%let ERROR: NO LABELS = OPTION SPECIFIED ON MACRO CALL;
%put %str(ERROR: NO LABELS = OPTION SPECIFIED ON MACRO CALL);
%put %str(HORIZONTAL AXIS LABELS WILL NOT REPRESENT);
%put %str(GROUP VALUES FROM INPUT DATA SET.);
%end;
* Save name of _last_ data set to reset option at end of macro;
%let SAVENAME=&SYSLAST;
* Count the number of dots at each Y value, lor each group;
proc summary data=&DATAIN nway;
class &GROUP &YVAR;
output out=_COUNTS (keep=&GROUP &YVAR fREQ_ rename=tFREQ_ =_COUNT);
run;
* Output one observation per dot, offsetting them a little;
data _RECODE;
set _COUNTS end:: _LASTREC:
by &GROUP;
* Check for too many dots on a line for the current DOTSPACE and GRPSPACE specifications to handle without running into the next group;
if _DOTSPACE * _COUNT > &GRPSPACE then
do;
call symput('TOOMANY', '1');
stop;
end;
* Set up midpoints for the groups, and count the number of groups;
if FIRST.&GROUP then
do;

1174
* Set horizontal distance between dots:

    retain _INCR &DOTSPACE;

* Output first dot: This will be the only dot if there is only one observation at this Y value. Otherwise, it will be the dot appearing at the far left, for the current Y value, and for this group. The variable _X will contain the horizontal coordinate of the group, though the actual values of _X (10, 20, etc.) will be masked. The purpose of the _X variable is to base the horizontal axis on a numeric variable, so that the dots can be offset by a numeric increment (_INCR);

    if _COUNT > 1 then do; /* more than one dot on this row */
        /* strip out left-hand side of row;*/
        _X = MIDPT - (int(_COUNT/2) * INC);
        /* move 1/2 space to right to center the dot if there are an even number of dots;*/
        if mod(_COUNT,2)=0 then _X = _X + (_INC/2);
        output _RECODE;
    end;
    else /* count=1 -- only one dot at this Y value */
        /* put dot exactly in center;*/
        _X = MIDPT;
        output _RECODE;
    end;

* Output remainder of dots (for this row, and for this group);

    if _COUNT > 1 then do; /* more than one dot on this row */
        _1 = 1 to _COUNT-1;
        _X = _X + _INCR;
        output _RECODE;
    end;

* Create the macro variable GRPCOUNT from the data step variable _GRPCNT, containing the final value of the variable _X for the dummy tick.

    if _LASTREC then do;
        _GRPCNT = (_GRPCNT + 1) * &GRPSPACE;
        call symput('GRPCOUNT,put(_GRPCNT,5.1I);
    end;

* Plot the points:

    proc plot data=_RECODE;
    plot &YVAR * _X;
    run;

    %if &GPLOT=YES %then
        %do; /* use GPLOT to produce picture */
    proc gplot data=_RECODE;
    plot &YVAR X / vminor=1; /* one minor (small) tick mark between major ones */
    axis=axis2; /* use AXIS2 statement to define vertical axis */
    haxis=axis1; /* use AXIS1 statement to define horizontal axis */
    frame; /* close the box around the picture */
    axis2 length=&LENGTH; /* length of vertical axis */
    axis1 order=0 to &GRPCOUNT by &GRPSPACE /* force axes to use group mklpoints as major tick marks */
    length=&LENGTH; /* length of horizontal axis */
    major=none; /* no major (large) tick marks */
    minor=none; /* no minor (small) tick marks */
    label=none; /* suppress "X", the name of the horiz. var. */
    value=HVALUE; /* options for tick mark text */
    /* blank label for the zero tick mark */
    &LABELS /* user-defined group labels */
    &VLabels /* suppress label for last tick mark */
    ;

* The SYMBOL statement specifies the plotting symbol used to create the dot;

    %if proc.cat&DOTSIZE)==LARGE %then
        %do;
        symbol h=0.3 v=dot;
        %end;
    %else
        %do;
        symbol v=point;
        %end;
    %end; /* use GPLOT to produce picture */

* Note that no RUN statement is included to allow the user to follow the macro call with additional SAS/GRAPH statements;

    %end; /* continue */

* Reset option last to name of most recently created data set;

    %reset _last_='&SAVENAME';

%end DOTDIAG;

AUTHOR
Karen Malloy
ARC Professional Services Group, Inc.
1301 Piccard Dr., Rm. 202
Rockville, MD 20850
(301)258-8826