

Don't Dither About Your Data, Let SAS/GRAPH® Trending Box Plots Drive Decision Making

Charlotte F. Carroll, M.S., Virginia Health Quality Center, Glen Allen, VA

ABSTRACT

Box and Whisker plots (box plots) are excellent tools for communicating point estimates, as well as information on the spread of data. Box plots, which are plotted over time, show changes in the median, range, and quartiles of the data over a specific time period. The ability to view changes in the median, as well as changes in the spread of the data over time, can enable the decision maker to see the effects of interventions more clearly. This paper demonstrates how the GPLOT procedure in SAS/GRAPH® can be utilized to create boxplots that trend over time. Multiple SYMBOL, AXES, and PLOT statements, as well as the OVERLAY, statement are employed to enhance the graphic by including dithered data points on the boxplots and adding additional trend lines to the plot.

INTRODUCTION

The Virginia Health Quality Center (VHQC), winner of the 2002 U.S. Senate Productivity and Quality Award for Virginia, is a health care quality improvement organization that assists health care providers in making successful and meaningful changes in the way care is delivered and in improving outcomes of that care, especially for the Medicare community of Virginia.

The box plot graphic described in this paper is used by the VHQC for communicating results of quality improvement projects to participating providers. Additionally, the box plots are used for targeting counties or providers who fall above or below the population of their peers over time. Counties or providers which fall below their peers are targeted for additional intervention. Counties or providers which fall above their peers are a focus for study into best practices, which can then be communicated to the Virginia health provider community.

This paper provides the SAS/GRAPH® code a user would need to create a trending box plot with added dithered data points and an additional trend line.

SAS/GRAPH® code highlighted in the creation of the box plots includes:

- The SYMBOL statement INTERPOL=BOXT## for creation of the box plot.
- Additional SYMBOL statements for dithered points and added trend line.
- Multiple graphs in one PLOT statement with the OVERLAY statement, so that the box plot and dithered data points appear on the same graph.
- PLOT2 statement to add the trend line to the graph.

CREATING THE BOX PLOT

Box plots were created using the PLOT statement from Proc GPLOT. A SYMBOL statement with INTERPOL=BOXT10 was used to create box and whisker plots with whiskers at the 10th and 90th percentiles of the dataset. SAS® allows for the adjustment of the percentiles where the whiskers will end. The author has selected the 10th and 90th percentiles for the purposes of demonstration.

The data points were added to the plot by adding an additional graph to the PLOT statement and utilizing the OVERLAY statement to place the graphs on top of one another. An

additional SYMBOL statement was created for the data points using a SYMBOL2 statement and specifying the VALUE=DOT.

The data points used to create the box plot were dithered in a prior data step by adding a random amount to the value of TIME and creating the variable TIME2. With dithering, the data points are separated and identifiable horizontally rather than superimposed and non-identifiable along a single vertical straight line.

The PLOT2 statement was included to add a trend line to the graph. The SYMBOL3 statement defined the VALUE=TRIANGLE and INTERPOL=JOIN for the trend line.

AXIS statements were utilized to define the values and labels of the x- and y-axes, and to ensure consistency of the y-axis values for the PLOT and PLOT2 statements.

The LEGEND statement was used to add the definition of the added trend line to the plot.

The experimental SAS extended meta file (SASEMF) device and the Output Delivery System RTF capabilities were utilized to automate the production of the graphic into a file which is easily read by the user.

The macro listed below which creates the box plots can be easily edited by the reader to automate production of box plot graphics over several variables and provider types.

Data used in this example are for illustration purposes only and do not represent actual county or statewide clinical results.

SAS Code:

```

/* Read in dataset with County Rates */
data county_all2;
  set county_all;
  by county;
  if cntyname not in ( ' ', 'UNKNOWN' );

/* Add random Dithering to data points */
/* uniform between -0.1 and +0.1 */
time2=time - .1 + .2*uniform(234);

label perc2="Statewide Rate";

run;

proc format;
  value tft 0=" "
           1="Baseline"
           2="Interim 1"
           3="Interim 2"
           4=" ";
run;

/* Macro which creates graph */
%macro graph(dat,var,min,tit1,tit12);

```

```

/* Set Gplot options */
goptions reset=global border htitle=1.5
htext=1.3 ftitle=complex ftext=simplex
device=sasemf;

/* Set up symbol for Boxplot */
symbol1 interpol=boxt10
      mode=exclude
      value=none
      co=black
      cv=black
      height=.6
      bwidth=15
      width=5;

/* Set up Symbol for Data Points */
symbol2 ci=blue
      value=dot
      h=.5;

/* Set up symbol for Statewide Rate */
symbol3 ci=red
      i=join
      value=triangle
      h=1.7
      line=8
      width=2;

/* Define X-axis */
axis1 split='' label=none
      order=(0 to 4 by 1)
      length=100
      ;

/* Define Y-axis */
axis2 label=(h=1.5 '(%)')
      order=(&MIN to 100 by 10);

/* Define Legend */
legend1 position=(middle bottom outside)
      frame
      across=2
      label=none
      shape=line(10)
      value=(h=1.3);

run;

/* Define Output File */
ods rtf file=
"c:\boxplots\LargeCountyPlotFor%CMPRES(&var).rtf";

/* Plot Graph */
proc gplot data=&DAT;
  where label="&var";

  plot rate*time rate*time2/ overlay haxis=axis1
      vaxis=axis2;

  plot2 perc2*time/overlay haxis=axis1
      vaxis=axis2 legend=legend1;

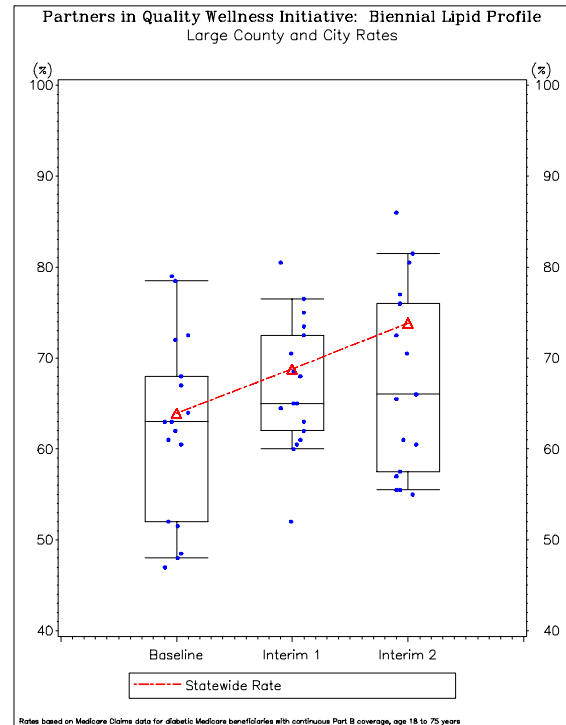
  format time tft.;
run;

ods rtf close;
%mend graph;

/* Call MACRO */
%graph(county_all3,totlipd,40,Biennial Lipid
  Profile,Large County and City Rates)
quit;

```

Example Graphic :



CONCLUSION

This paper demonstrates a method for using SAS/GRAPH® to create trending box and whisker plots. The method highlights use of the SYMBOL, PLOT, OVERLAY, and PLOT2 statements associated with PROC Gplot. Additional information is provided to the reader to allow for the inclusion of dithered data points and added trend lines on to the box plot graphic.

REFERENCES

<http://www.vhqc.org>.

SAS Institute Inc. 1999. *SAS/GRAPH Software: Reference, Version 8, Volumes 1 and 2*.

ACKNOWLEDGEMENTS

I would like to thank Bob Boudreau, PhD, and Barbara Okerson, PhD for assistance with the dithering algorithm and inspiration and guidance in the writing and review of this paper.

Data sets used for example were generated used SAS/STAT Proc SURVEYSAMPLE.

SAS, SAS/GRAPH, and SAS/STAT are registered trademarks of SAS Institute Inc. of Cary, North Carolina in the USA and other countries.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The analyses upon which this publication is based were performed under contract no. 500-99-VA-03, titled "Utilization and Quality Control Peer Review Organization for the Commonwealth

of Virginia," sponsored by the Centers for Medicare & Medicaid Services (CMS), Department of Health and Human Services (DHHS). The content of this publication does not necessarily reflect the view or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The Virginia Health Quality Center (VHQC) assumes full responsibility for the accuracy and completeness of the ideas presented. This publication is a direct result of the Health Care Quality Improvement Program initiated by CMS, which has encouraged identification of quality improvement projects derived from analysis of patterns of care and therefore required no special funding on the part of VHQC. Ideas and contributions to the VHQC concerning experience in engaging with the issues presented are welcomed.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. For more information contact:

Charlotte F. Carroll, M.S.
Biostatistician
Virginia Health Quality Center
4510 Cox Road, Suite 400
Glen Allen, VA 23060
Phone: 804.289.5320
e-mail: vapro.ccarroll@sdps.org



VIRGINIA HEALTH QUALITY CENTER