

## Annotation utilizations in customized SAS/Graph bar charts

Yong Liu, Centers for Disease Control & Prevention (CDC); Hua Lu, CDC; Liang Wei, DB Consulting Group, Inc.; Xingyou Zhang, CDC; Paul I Eke, CDC; James B Holt, CDC

### Abstract

Bar graphs are generated by using SAS/Gchart to present the distribution of health behaviors or health outcomes among adults aged  $\geq 18$  years by selected characteristics and each of 50 states using 2011 Behavioral Risk Factor Surveillance Survey (BRFSS). Due to missing data or unreliable estimates of parameters, annotation facilities are utilized to make the charts more presentable by adding data labels and footnotes. Further, incorporating a SAS Macro variable into the program can definitely make the development of 50 charts for 50 states more achievable and efficient.

### Introduction

Tables and graphs are frequently used for health researchers or epidemiologists in either federal health agencies or state departments of health to demonstrate health disparities by selected socio-demographic characteristics such as race/ethnicity and educational attainment. SAS/Graph is a powerful tool to fulfill this request especially for a number of states although Microsoft Excel is an alternative when a few charts are needed. In this paper, 50 bar charts with the information of prevalence of health outcomes by racial/ethnic groups and education levels in individual state were generated by using the 2011 Behavioral Risk Factor Surveillance Survey data. All data analyses were implemented using SAS® and SAS-callable SUDAAN (Release 10.0.1, Research Triangle Institute, North Carolina, USA).

**Step 1:** The data of percentages of health outcomes by race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, and others), education level ( $<12$ ,  $=12$ , and  $>12$  years) at the state level as well as U.S. median and state total percentage for 50 states were output through a do loop statement (Wei, 2008). In addition, a macro program was used to generate estimates (percentage) for any health outcomes, i.e. current smokers.

```
%macro pct (var);
```

```
** to get weighted estimates of parameter -median of the U.S., percent;
```

```
proc describe data=bf14 filetype=sas design=wr;
  nest _ststr _psu/missunit;
  weight _llcpwt;
  subpopn _llcpwt>0 and _state<57/name="landline users and cellphone users in 50 states and DC of
  2011 BRFSS";
  subgroup race_c ses smk_cur state;
  levels 4 3 2 50;
  var &var;
  catlevel 1;
  tables state;
  print nsum="sample size" wsum="Population" percent/wsumfmt=f12.1;
  output nsum percent/replace filename=a1 filetype=sas;
  title "weighted percentage of health behaviors in 2011 BRFSS";
run;
```

```
proc univariate data=a1;
  var percent;
  output out=a2 median=rowper;      **output U.S. median prevalence;
  where state>0 and percent>0;
run;
```

```
data a3;
  set a2;
  length label_g $85.;
  label_g='Median U.S.';
  keep label_g rowper;
run;
```

```
%do i=1 %to 50;
```

%let clist=AL AK AZ AR CA CO CT DE FL GA HI ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NE NV NH NJ NM NY NC ND OH OK OR PA RI SC SD TN TX UT VT VA WA WV WI WY; \*\* have to match the order in term of the Federal Information Processing Standard (FIPS) state codes in the BRFSS data;

\*\* to obtain weighted percentage of health outcomes for each state;

```
proc crosstab data=ff14 filetype=sas design=wr;
  nest _ststr _psu/missunit;
  weight _llcpwt;
  subpopn _llcpwt>0 and state=&i/name="landline users +cellphone users in 2011 BRFSS";
  subgroup race_c ses smk_cur;
  levels 4 3 2 ;
  tables race_c*&var;
  print nsum="sample size" wsum="Population" rowper/wsumfmt=f12.1;
  output nsum wsum rowper serow/replace filename=b_%scan(&clist, &i) filetype=sas;
  title "weighted percentage of health behaviors in 2011 BRFSS";
run;
```

```
data b2_%scan(&clist, &i);
  set b_%scan(&clist, &i);
  length label_g $85.;
  if &var=1 and race_c=0 then label_g='Total State';
  if &var=1 and race_c=1 then label_g='White, non-Hispanic';
  if &var=1 and race_c=2 then label_g='Black, non-Hispanic';
  if &var=1 and race_c=3 then label_g='Hispanic';
  if &var=1 and race_c=4 then label_g='Other';
  keep label_g rowper ;
run;
```

Similarly, the information of the percentage by education levels was also generated and output as a dataset with the name of c2\_%scan(&clist, &i). In addition, all datasets were combined together to create a master dataset with all necessary information required to make the chart;

```
data t1_%scan(&clist, &i);
  set a3 b2_%scan(&clist, &i) c2_%scan(&clist, &i);
  if _n_ in (1,2) then group1=1;      **U.S. Median and Total State;
  else if _n_ in (3,4,5,6) then group1=2;  ** racial/ethnic groups;
  else if _n_ in (7,8,9) then group1=3;    **education levels;

  if reliable=2 then total=0.01;      **to assign a much smaller value to the unreliable
  estimate, which is defined if the total number of respondents among the subgroup is less than
  50 or relative standard error ≥30%;
  else total=rowper;
run;
```

**Step 2:** To customize the bar charts using annotate facility or annotate datasets.

The technique of adding useful information such as labels and footnotes into the bar charts differs from that for graphs with continuous data in the vertical and horizontal axis. First, we have to apply 'Midpoint' option along with the position on the X axis to define where labels or footnotes are inserted the bar graphs (Atkinson, 2004). In the following SAS codes, the first annotated dataset was created to add data values for the bars of U.S. median and total state percentages and to add a symbol of 'asterisk' to those groups with unreliable estimates. Specifically, an option of "when='a'" specifies annotations that were drawn after the generation of the chart. In addition, the other options such as "position='2'" along with "xsys='2'" and "ysys='2'" will input the data value into the area of one cell above the bar with a centered position while "position='8'" along with "xsys='5'" and "ysys='5'" will assign the footnotes to the area at one cell below the left-side bottom boundary of the chart in the second annotated dataset (Atkinson, 2004, Stepien, 2006).

\*\*to add data labels for reliable estimates and an asterisk for unreliable estimates;

```
data anno1_%scan(&clist, &i);
  length function style color $8 position $1 text $50;
```

```

retain function 'label' xsys '2' ysys '2' when 'a' ;
set t1_%scan(&clist, &i);
if _n_ in (1,2) then do;
function='label';style='swissb';y=total; hsys='2';group=group1;midpoint=rf;
text=put(rowper,4.1);color='black';size=2;position='2';output;
end;

if reliable=2 then do; **to label unreliable estimates;
function='label';style='swissb';y=total; hsys='2';group=group1;midpoint=label_g;
text='*';color='black';size=2 ;position='2';output;
end;
run;

** to add footnotes;

data anno2_%scan(&clist, &i);
length function style color $8 position $1 text $50;
retain function 'label' x 20 y 2 xsys '5' ysys '5' when 'a' ;
set t1_%scan(&clist, &i);
if reliable=2 then do;
function='label';style='swissb'; text='*unstable estimates due to small sample';
color='black';size=1;position='8' ;output;
end;
run;

** to combine the two annotate datasets;

data anno_%scan(&clist, &i);
set anno1_%scan(&clist, &i) anno2_%scan(&clist, &i);
run;

```

**Step 3:** To output the customized state-specific charts using SAS Output Delivery System (ODS).

SAS ODS can output the products with GIF/PNG format or HTML format to a designated directory for further applications such as presentation (Massengill, 2006, Watts, 2007). The statement of 'Patternid=midpoint' is used to define the order of bars with assigned colors. The options of 'Vbar statement, 'space=2' is used to define the space between each bar (2 points); 'Gspace=6' is used to define the space of bars between groups (6 points); 'Nozero' option is used to control the order of the bars in non-hierarchical bar charts and is ultimately defined by the 'Midpoints' option. The state-specific chart is also named with the corresponding state abbreviated name (a macro variable), which may be useful for further data manipulation or to link to the other products such as tables or maps.

\*\* to create separate macro variables with state full name and state abbreviations for all 50 states in order to assign an appropriate name to the output chart;

```

data bf1;
set brfss2011;
statefullname=fipname1(_state); ** full state name;
st_abbre=fipstate(_state); **abbreviated state name;
run;

data _null_;
set bf1;
call symput('state1',trim(left(statefullname)));
call symput('dsn',trim(left(st_abbre)));
run;

** to output graphs with gif format or html format to a designated directory;

filename graphout 'c:\brfss\figure\';

```

```

options reset=all device=gif noborder gunit=pct cback=white gsfname=graphout gsfmode=replace
htitle=12pt htext=10pt;

axis1 label=(angle=180 f="Arial/Bold" "%") order=(0 to 50 by 10) minor=none value=('0' '10' '20' '30'
'40' '50' h=8pt);      **to define Y-Axis based on the range of percentages of health outcomes;
axis2 label=none split=", " value=(h=6pt);      **X-Axis;
axis3 label=none nobrackets value=(' ' 'Race/ethnicity' 'Education (years)' h=10pt );      **to
define group labels on x-Axis-gaxis and to remove the brackets with 'nobrackets' option;

**to assign each group/each bar with different colors;

pattern1 v=solid color=green;
pattern2 v=solid color=lightgreen;
pattern3 v=solid color=lightpurple;
pattern4 v=solid color=lipk;
pattern5 v=solid color=lightyellow;
pattern6 v=solid color=cyan;
pattern7 v=solid color=lightblue;
pattern8 v=solid color=brown;
pattern9 v=solid color=darkblue;

ods graphics on;
ods html body="&dsn..html" path=graphout style=minimal;
ods listing close;

title1 h=14pt "Figure. Percent of current smokers among adults aged>=18 years ";
title2 h=14pt "by selected characteristics, &state BRFSS 2011";
footnote1 " " h=1pt;
footnote2 " " h=1pt;

**to create a bar graph/chart;
proc gchart data=t1_%scan(&clist, &i);
vbar label_g/ width=6 noframe group=group1 patternid=midpoint sumvar=total space=2 gspace=6
nozero midpoints="Median U.S." "Total State" "White, non-Hispanic" "Black, non-Hispanic"
"Hispanic" "Other" "<12" "12" ">12" coutline=black woutline=2 raxis=axis1 maxis=axis2 gaxis=axis3
annotate=anno_%scan(&clist, &i);
run;
%end;** to end the do loop statement so the graphs for all states have been created and saved into the
directory;
run;
quit;
ods graphics off;
ods html close;
ods listing;

%mend;

%pct(var=smk_cur);      **current smokers for 50 states;

%pct(var=obeses);      **obesity prevalence for 50 states;

```

50 bar charts were generated with a clear, concise depiction of the percentage of health outcomes such as current smokers by geographic areas (the national and state levels), by racial/ethnic subgroups, and at education levels among adults aged 18 years and older (See attached figure).

## Conclusion

SAS/Graphic, ODS, and SAS Macro facility were utilized to create bar graphs with some useful information for all 50 states. The techniques demonstrated in this report allowed researchers or data analysts who are working either in federal agencies or in state departments to make quickly review for the distribution of any health outcomes by selected socio-demographics and also make the comparison of health outcome profiles to the U.S. median among states. Hopefully, these presented tips can save you time and make you enjoy using SAS.

## References

- Wei L, Barker L. (2008). Creating county-level estimates from national weather service data. Proceedings of Annual Southeast SAS Users Group 2008 Conference (SESUG), St Petersburg Beach, Florida. SAS Institute Inc. Available at: <http://analytics.ncsu.edu/sesug/2008/ST-142.pdf>.
- Atkinson M. (2004). Customising SAS/GRAPH charts with annotations. SAS Institute Inc. Available at: <http://www.lexjansen.com/pnwsug/2009/Atkinson,%20Mike%20-%20Customising%20SAS-GRAPH%20Charts%20with%20Annotations.pdf>.
- Stepien S. (2006). How to annotate graphics. Pharmaceutical Users Software Exchange (PhUSE) 2006 Conference. Dublin, Ireland. Available at: <http://www.phusewiki.org/docs/2006/TU05.pdf>.
- Massengill D. (2005). Tips and tricks: Using SAS/Graph effectively. SUGI 30, Philadelphia, PA. SAS Institute Inc. Available at: [http://support.sas.com/rnd/papers/sugi30/p90-30\\_020105.pdf](http://support.sas.com/rnd/papers/sugi30/p90-30_020105.pdf).
- Watts P. (2007). Building a better bar chart with SAS/GRAPH software. Proceedings of the 20<sup>th</sup> Annual Northeast SAS Users Group 2007 Conference. Baltimore, Maryland. SAS Institute Inc. Available at: <http://www.nesug.org/proceedings/nesug07/np/np16.pdf>.

## Contact Information

Yong Liu, Statistician, Epidemiology & Surveillance Branch, Division of Population Health, Centers for Disease Control & Prevention, 4770 Buford Highway, Mail-stop F-78, Atlanta, GA 30341.

Your comments and questions are valued and encouraged. Contact the authors at: (770)488-5528 or e-mail: [IKD8@cdc.gov](mailto:IKD8@cdc.gov).

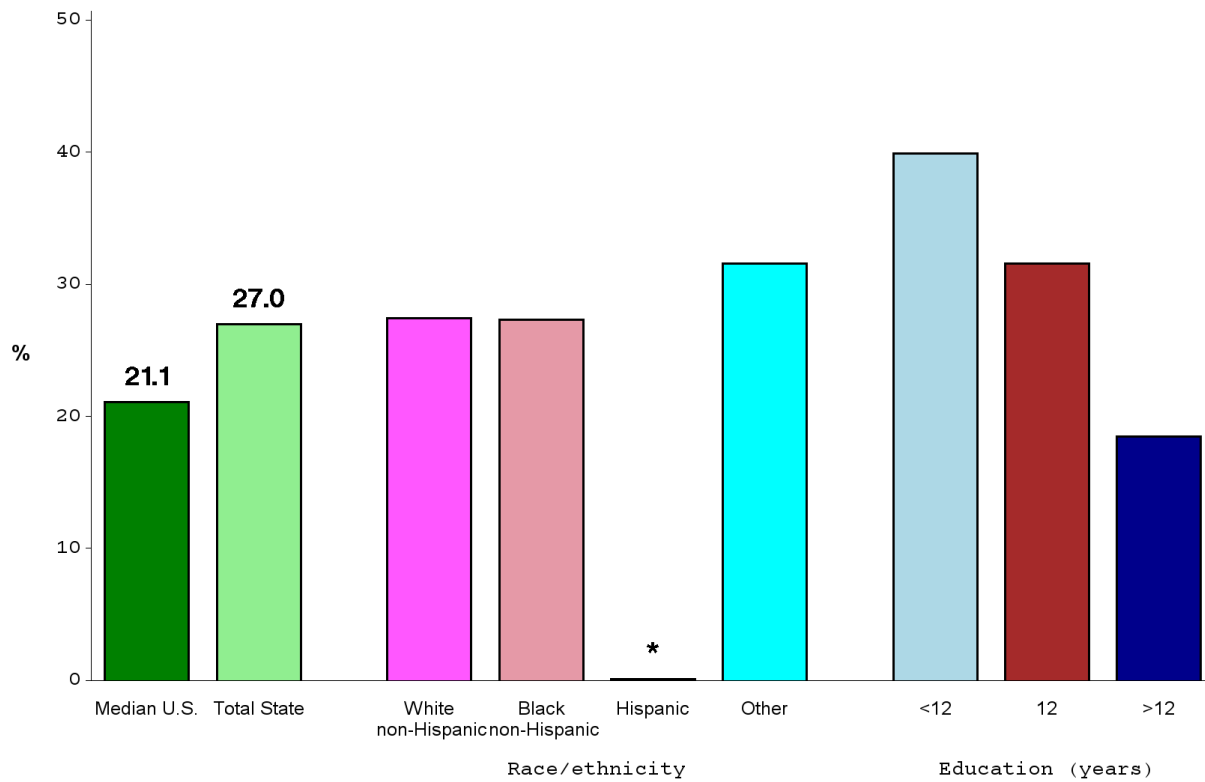
SAS ® and all other SAS Institute Inc. product or service name are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries, ® indicates USA registration.

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

**Disclaimers:** The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Caption: Figure. Weighted prevalence of current smokers among adults aged  $\geq 18$  years, by selected characteristics, Arkansas, Behavioral Risk Factor Surveillance System 2011

Figure. Percent of current smokers among adults aged  $\geq 18$  years by selected characteristics, Arkansas BRFSS 2011



\*unstable estimates due to small sample