Inline Formatting + Long Character Variables = "WOW! That's SAS®?"

Derek Morgan, Washington University Medical School, St. Louis, MO

ABSTRACT

Version 8 of SAS expanded the maximum size of character variables, and this is a practical use of that capacity. When you combine this with a custom ODS template, in-line formatting, and the REPORT procedure, you can build a report that will look as if it came from a word processor.

THE PROBLEM

Several reports were requested as a part of the development of an academic publication tracking database application. These reports went to various destinations, some to the web, some to PDF (sometimes simultaneously) and one to RTF. ODS handles the destinations, and PROC REPORT takes care of the layout and production of the report. However, most reports generated via PROC REPORT have the data neatly arranged in columns.

Although the client (my boss) would have accepted a report of this type (it was better than what they had), one of the standard reporting formats for academic publications has the author name(s), followed by the title of the article and the journal information as shown here:

> Arnett DK, Williams RR, Folsom AR, Rao DC, Heiss G. **Dyslipidemic Hypertension and Familial Aggregation of Coronary Heart Disease: The ARIC/FHS Studies.** Atherosclerosis; 1999, 2, 180-86

While it is easy to track each of these as individual columns in a database, to show it on a report in this format requires a little bit of creativity.

IT'S NOT HARD

First, the application neeeds to make it easy to assemble all of the components into one long column. The title is stored in a long character variable, while author names are stored in 15 separate numeric variables. Each author's name is then formatted, thereby reducing the number of ways that a person's name can be represented, removing redundancy from the author dataset. The same philosophy is applied to the journal name; it is also stored as a numeric variable and formatted when it is used.

The list of author names is created for the report by concatenating the formatted value of each non-missing author name variable and storing it in a long character variable. The journal name is also produced by formatting the value of the journal variable. However, the journal information needs to be standardized because some journals list articles with a volume, while some have both a volume and a number, along with a starting and ending page number. You can concatenate the values stored in the volume, number, starting and ending page variables (again, if they aren't missing) into a long character variable. Now you have the elements of the report ready for the final concatenation into one long character variable.

GETTING PROC REPORT AND ODS TO COOPERATE

It's not very difficult, either. Simply put, all of the information to be printed needs to be in one column. The FLOW option on the DEFINE statement will cause the text to wrap nicely, while the UNIFORM option on the ODS PDF statement will insure that the column takes up the entire width of the page, no matter how wide it may be. I also modified the default PRINTER style with the TEMPLATE procedure. By setting the border color to white (or the same as the backgound of the report) you make the lines around the column cell vanish. To space the output, I defined the cell spacing and cell padding to 12 points, which gives the appearance of double spacing, given the font size used in the report.

IN-LINE FORMATTING

ODS ESCAPECHAR allows you to insert formatting directives such as font size, type, and style into your output. The formatting directives are a part of the data to be printed in the cell. Line 40 is where we assemble the final long output line. The "~S" indicates where a directive is placed. The first directive is ~S={font_weight=bold}, which is placed right before we concatenate the article's title with the author list. The bolding is turned off by using ~S={}, and then we concatenate the citation text we created (in lines 29-34) to complete the output string.

THE CODE

```
1    OPTIONS CENTER NODATE;
2    ODS ESCAPECHAR="~";
3    ODS PATH work.templat(UPDATE) sashelp.tmplmst(READ);
4    PROC TEMPLATE;
5      DEFINE STYLE pdfbiblio; PARENT=styles.printer;
6       REPLACE BODY FROM DOCUMENT /
7       LEFTMARGIN=.75in RIGHTMARGIN=.75in TOPMARGIN=.75in BOTTOMMARGIN=.75in;
8       STYLE SYSTEMTITLE / FOREGROUND=black FONT_FACE=arial FONT_SIZE=4;
9       STYLE SYSTEMFOOTER / FOREGROUND=black FONT_FACE=arial FONT_SIZE=2;
                         /* Double-space cells and remove border */
10      STYLE TABLE / CELLPADDING=12 CELLSPACING=12 BORDERCOLOR=white BORDERWIDTH=0;
11      STYLE HEADER / FOREGROUND=black FONT_FACE=arial FONT_SIZE=3.75;
12      STYLE PAGENO /FONT_FACE=arial FONT_SIZE=2;
13     END;
14   RUN;

15   DATA allpubs;
16   SET data.pubs;
17   LENGTH an_list $ 440 citate $ 500 temp $ 30;
18   ARRAY auth{15} author1-author15;
19   ARRAY jx{3} jyear jvol jno;
20   an_list = LEFT(PUT(author1,authabr.));   /* Assemble author list */
21   DO i=2 TO 15;
22     IF auth{i} NE . THEN
23        an_list = TRIM(an_list) || ", " || LEFT(PUT(auth{i},authabr.));
24   END;
25   DO i=1 TO 3;   /* Assemble Journal Information */
26      IF jx{i} NE . THEN
27        temp = TRIM(LEFT(temp)) || ' ' || TRIM(LEFT(put(jx{i},4.))) || ",";
28   END;
29   citate = TRIM(PUT(journal,jrnl.)) || "; " || LEFT(temp);
30   IF jstrtpage NE . THEN DO;
31     citate = TRIM(citate) || " " || LEFT(PUT(jstrtpage,5.));
32     IF jendpage NE . THEN
33        citate = TRIM(citate) || "-" || LEFT(PUT(jendpage,5.));
34   END;
35   RUN;

36   TITLE "Family Heart Study Bibliography";
37   DATA pdfpublic;
38   SET public;
39   LENGTH biblio $ 2000;   /* Long character variable */
                 /*    Assemble long character variable with in-line formatting.
           ~ is the ODS ESCAPECHAR, which is followed by the S=  style directive    */
40   biblio = TRIM(an_list) || ".  " || "~S={font_weight=bold}" || TRIM(title) ||
            ".~S={}  " ||  TRIM(citate);
41   KEEP pubstat biblio;
42   RUN;
43   ODS PDF FILE="/users/fhscc/public_html/public_pubs.pdf" STYLE=pdfbiblio UNIFORM;
44   PROC REPORT DATA=pdfpublic NOWD NOHEADER STYLE=[OUTPUTWIDTH=100%];
45   COLUMNS pubstat biblio;
46   DEFINE pubstat / ORDER NOPRINT;
47   DEFINE biblio / DISPLAY FLOW " ";
48   RUN;
49   ODS PDF CLOSE;
```

THE RESULT

Family Heart Study Bibliography
Last Updated on: January 19, 2006

FHS Published Manuscripts

Arnett DK, Williams RR, Folsom AR, Rao DC, Heiss G. **Dyslipidemic Hypertension and Familial Aggregation of Coronary Heart Disease: The ARIC/FHS Studies.** Atherosclerosis; 1999, 2, 180-86

Arnett DK, Borecki IB, Ludwig E, Pankow JS, Myers RH, Evans G, Folsom AR, Heiss G, Higgins M. **Angiotensinogen and angiotensin converting enzyme genotypes and carotid atherosclerosis: The Atherosclerosis Risk in Communities and The NHLBI Family Heart Studies.** Atherosclerosis; 1998, 138, 11111-6

Arnett DK, Miller M, Coon H, Ellison R, North K, Province MA, Leppert MF, Eckfeldt JH. **Genome-Wide Linkage Analysis Replicates Susceptibility for Fasting Serum Triglycerides: The NHLBI Family Heart Study.** Human Genetics; 2004, 115, 6, 468-474

Bensen J, Li R, Hutchinson RG, Province MA. **Family History of CHD and Pre-Clinical Atherosclerosis in African Americans and Whites:The ARIC & FHS Studies.** Genetic Epidemiology; 1999, 16, 165-178

Bensen J, Liese A, Rushing J, Province MA, Folsom AR, Rich S, Higgins M. **The Accuracy of Proband Reported Family History: The Family Heart Study (FHS) Experience.** Genetic Epidemiology; 1999, 17, 141-150

Borecki IB, Province MA, Ludwig E, Ellison R, Folsom AR, Heiss G, Lalouel J, Higgins M, Rao DC. **Associations of Candidate Loci AGT and ACE with Severe Hypertension: The NHLBI Family Heart Study.** Annals of Epidemiology; 1997, 7, 3-21

Borecki IB, Higgins M, Schreiner PJ, Arnett DK, Mayer-Davis E, Hunt SC, Province MA. **Evidence for Multiple Determinants of the Body Mass Index: The NHLBI Family Heart Study.** Obesity Research; 1998, 6, 107-224

Bostom AG, Jacques PF, Nadeau MR, Williams RR, Ellison R, Selhub J. **Post-Methionine Load Hyperhomocysteinemia in Persons with Normal Fasting Total Plasma Homocysteine: Initial Results from the NHLBI Family Heart Study.** Atherosclerosis; 1995, 116, 14, 147-151

Coon H, Leppert MF, Province MA, Myers RH, Arnett DK, Eckfeldt JH, Heiss G, Williams RR, Hunt SC. **Evidence for a Major Gene Accounting for Mild Elevation in LDL cholesterol..** Annals of Human Genetics; 1999, 63, 5, 401-412

Coon H, Myers RH, Borecki IB, Arnett DK, Hunt SC, Province MA, Djousse L, Leppert MF. **Replication of Linkage of Familial Combined Hyperlipidemia to Chromosome 1q With Additional Heterogeneous Effect of Apolipoprotein A-I/C-III/A-IV Locus.**   Atherosclerosis, Thrombosis, and Vascular Biology; 2000, 20, 10, 2275-2280

Coon H, Eckfeldt JH, Leppert MF, Myers RH, Arnett DK, Hunt SC. **A Genome-Wide Scan Reveals Evidence for a Locus on Chromosome 11 Influencing Variation in LDL Cholesterol in the NHLBI Family Heart Study.** Human Genetics; 2002, 111, 263-269

Coon H, Singh, Dunn, Eckfeldt JH, Province MA, Hopkins PN, Weiss R, Hunt SC, Leppert MF. **TXNIP gene not associated with Familial Combined Hyperlipidemia: The NHLBI Family Heart Study.** Atherosclerosis; 2004, 174, 2, 357-362

SUMMARY

You can view the current report at: http://www.biostat.wustl.edu/fhs/public_pubs.pdf. Now this could probably be done another way, but the database and the application reside on a LINUX system, and the output is funnelled to the web. The final result is the important thing here. When my supervisor (a very experienced SAS user) saw this, he couldn't believe that no manual intervention was required to produce this. And it wasn't hard to do!

As of version 8, you can store a short essay in a SAS character variable. However, long character variables can be used for more than storing an expansive text description. You can combine several columns from a table and concatenate them into a single column for a report. The FLOW option on PROC REPORT will break this column nicely across output lines, and show it as neatly-formatted text that gives no indication of its humble origins as multiple numeric and/or character columns in a table. If you use ODS ESCAPECHAR, and place formatting directives into the text to be displayed, you can change fonts, font sizes and styles, and colors within the text. All of this makes it seem all the more remote that this is an automated report produced by Base SAS®.

ACKNOWLEDGEMENTS

CONTACT INFORMATION:

Further inquiries are welcome to:

Derek Morgan
Division of Statistical Genomics
Washington University Medical School
Box 8506, 4444 Forest Park Blvd.
St. Louis, MO 63108
Phone: (314) 362-3685   FAX: (314) 362-4227
E-mail: derek@wustl.edu

This and other SAS System examples and papers can be found on the World Wide Web at:

http://www.biostat.wustl.edu/~derek/sasindex.html