

“Use SAS® PROC Tabulate to create a PDF with Bookmarks”

By Barbara Felts, Chuchun Chien and Susan Myers, RTI, International

ABSTRACT

Recently our client requested a report of cross frequencies containing many variables to compare and review their data. Our original thought was to run a cross frequency however this created a very long output. While this initial output met the client's request, the information was difficult to review and did not contain a defined structure.

It was determined the final report needed to provide a structure to the end user which would allow easy access to the comparisons within the report and have a finished appearance. We discovered that you can do so much more to improve the look and readability of the report by using ODS and SAS® Proc tabulate.

This paper will explain our progression from a plain SAS® output of a simple cross frequency to a PDF and then to a more user-friendly PDF with meaningful bookmarks using macros, SAS® ODS and SAS® Proc Tabulate. We will show you how to take your simple Cross Frequencies beyond simple output to this next level of a finished report.

INTRODUCTION

To address a data request involving two variables over 10 years, we had the opportunity to explore several options offered by SAS to create a meaningful report that was easy to navigate. Respondents in our survey were allowed to enter multiple choices in each of our two categories. Because of multiple response options we needed to cross our variables for every possible combination. Our first approach was the simplest, a cross frequency of the requested variables. While this provided the required information, it was lacking in many ways. For one year, the cross frequency resulted in 50 tables since the first category of interest was five variables, and the by group was ten variables. Running this for 10 years would mean scrolling through hundreds of pages.

Our next approach involved Proc Tabulate to an RTF document using ODS. This method greatly improved the appearance but was still hard to navigate. The Table of Contents available with ODS RTF allowed navigation but still had some drawbacks.

Lastly, we transitioned the output to PDF which allowed us to append our results to one document through macro looping. The bookmarks were added and customized by options offered under ODS.

In the following paragraphs we will discuss the constraints and how we dealt with them to make an attractive and useful report for our client. We will go into more detail on options that took our report from a cumbersome, lengthy listing to an attractive document with meaningful bookmarks.

PROC FREQ (CROSS FREQ)

Our first approach to get the requested information for our client was to use a simple Proc FREQ with cross frequency. We created a macro and used variables to loop through the desired years (YR) and a do loop where i incremented to 10 for the 10 different variables (Color1-Color10). We also used a format statement to include text for the answer choices for the two sets of variables which both have the same format 1 = '1-Yes' and 2 = '2-No'. We then invoked the macro (Color) with the desired year passed through as the parameter.

Cross Frequency code:

```
proc format;    /*Format to show meaningful text for responses*/
Value YesNoFMT
    2 = '2-No'
    1 = '1-Yes';
run;

%Macro RunYr (i,yr);
    proc freq data=in.shoes&yr. ;
        tables
            Color&i.*Style1          /*cross freq the variables*/
            Color&i.*Style2
            Color&i.*Style3
            Color&i.*Style4
            Color&i.*Style5
            /list missing;
        format Color&i. Style1-Style5 YesNoFMT.; /*apply formats*/
        title "20&yr. SHOES";
    run;
%mend;

%MACRO Color (yr);
    %do i=1 %to 10;    /*loop through the variables Color1-Color10*/
        %RunYr(&i,&yr.); /*call the RunYr macro for each year*/
    %end;
%mend Color;

%Color(06); /*call the macro for each year*/
```

2006 SHOES						1
The FREQ Procedure						
color3	style2	Frequency	Percent	Cumulative Frequency	Cumulative Percent	
1-Yes	1-Yes	981	98.10	981	98.10	
1-Yes	2-No	17	1.70	998	99.80	
2-No	1-Yes	2	0.20	1000	100.00	
<OVER 3000 LINES OF SIMILAR OUTPUT>						

Figure 1. Output from the cross frequency

The cross frequency produced the requested output (Figure 1) but it was not easy to navigate or to read because there were over 3000 lines of output to wade through. The format made the responses more descriptive, but the variable names did not mean much to the analyst.

PROC TABULATE

Next, we used PROC TABULATE to improve the appearance of the output. We included some macro variables to use as descriptive text for the variable names and the proc format for the response descriptions.

PROC TABULATE code:

```

/*Set up some macro variables to describe the variables*/
%let ColorBox10= 'YELLOW';
%let ColorBox9 = 'ORANGE';
%let ColorBox8 = 'GREEN';
%let ColorBox7 = 'PURPLE';
%let ColorBox6 = 'BLUE';
%let ColorBox5 = 'PINK';
%let ColorBox4 = 'RED';
%let ColorBox3 = 'WHITE';
%let ColorBox2 = 'BLACK';
%let ColorBox1 = 'BROWN';

%let StyleBox5 = 'BOOTS';
%let StyleBox4 = 'SANDALS';
%let StyleBox3 = 'HEELS/WEDGES';
%let StyleBox2 = 'CASUALS/FLATS';
%let StyleBox1 = 'SNEAKERS';

proc format; /*Format to show meaningful text for responses*/
Value YesNoFMT
    2 = '2-No'
    1 = '1-Yes';
run;

%Macro RunYr (i,yr); /*Run for each Color1-Color10 and year*/
%do j=1 %to 5; /*Run for each Style1-Style5*/
    proc tabulate data=in.shoes&yr. missing;
        class Color&i. Style&j.;
        table Color&i.=' ', (Style&j.) *
            (n*f=6.0 rowpctn) / printmiss misstext='0' box=&&ColorBox&i;
        KEYLABEL N = " ";
        label Style&j. = &&StyleBox&j.; /*Apply meaningful Labels*/
        format Color&i. Style1-Style5 YesNoFMT. ; /*Apply format*/
        title "20&yr. SHOES";
    run;
%end;
%mend;

%MACRO Color (yr);
%do i=1 %to 10; /*loop through the variables Color1-Color10*/
    %RunYr(&i,&yr.); /*call the RunYr macro for each year*/
%end;

```

```
%mend Color;

%Color (06); /*call the macro for each year*/
```

2006 SHOES					
WHITE	SNEAKERS				
	1-Yes		2-No		
	RowPctN		RowPctN		
1-Yes	985	98.70	13	1.30	
2-No	1	50.00	1	50.00	

<OVER 5000 LINES OF SIMILAR OUTPUT>

Figure 2. Output from PROC TABULATE

While output was easier to understand because of the descriptive text and layout (Figure 2), it was still hard to navigate through over 5000 lines and 400 pages of output across all of the years.

PROC TABULATE TO RTF

We thought we could get the output to look better and be more user-friendly by creating a RTF document with a table of contents. Utilizing the ODS options CONTENTS=YES and TOC_DATA enables the table of Contents to be created on the first page of the report. The ODS Proc Label with our macro variables allows us to capture the current table variables and their values in the Table of Contents.

PROC TABULATE to RTF Code:

```
/*Set up some macro variables to describe the variables*/
%let ColorBox10= 'YELLOW';
%let ColorBox9 = 'ORANGE';
%let ColorBox8 = 'GREEN';
%let ColorBox7 = 'PURPLE';
%let ColorBox6 = 'BLUE';
%let ColorBox5 = 'PINK';
%let ColorBox4 = 'RED';
%let ColorBox3 = 'WHITE';
%let ColorBox2 = 'BLACK';
%let ColorBox1 = 'BROWN';

%let StyleBox5 = 'BOOTS';
%let StyleBox4 = 'SANDALS';
%let StyleBox3 = 'HEELS/WEDGES';
%let StyleBox2 = 'CASUALS/FLATS';
%let StyleBox1 = 'SNEAKERS';

proc format; /*Format to show meaningful text for responses*/
Value YesNoFMT
    2 = '2-No'
    1 = '1-Yes';
run;
```

```

/*create a RTF file for the output*/
ods rtf file = "/mydir/tabulateBookmark.rtf"
    startpage=no /*prevents each proc from going to a new page*/
    keepn /*Keepn prevents tables from splitting across pages*/
    CONTENTS=YES TOC_DATA; /* For creating a Table of Contents*/

%Macro RunYr (i,yr);
%do j=1 %to 5; /*loop through Style1-Style5*/
    /*proclabel with macro variables create the meaningful bookmarks*/
    ods proclabel="20&yr. &Colorbox&i. by &Stylebox&j.";

    proc tabulate data=in.shoes&yr.
        contents= ' ' /*Gets rid of 2nd bookmark node*/
        missing; /*Include any missing values*/
        class Color&i. Style&j.; /*Variables to be grouped*/
        table Color&i.= ' ' /*Specifies the text for the box above the row*/
            , (Style&j.)*(n*f=6.0 ROWPCTN)/ /*Include the number and freq*/
        printmiss misstext='0'/*Show 0 instead of . for missing counts*/
        box=&&ColorBox&i /*puts meaningful label text*/
        contents= ' '; /*Gets rid of 3rd bookmark node*/
        KEYLABEL N = " "; /* put a blank above the number col*/
        label Style&j. = &&StyleBox&j.; /*puts meaningful label text*/
        format Color&i. Style&j. YesNofmt.; /*formats the responses*/
        title "20&yr. Shoes";

    run;
%end;

%mend;

%MACRO Color (yr);
    %do i=1 %to 10; /*loop through the variables Color1-Color10*/
        %RunYr(&i,&yr.); /*call the RunYr macro for each year*/
    %end;
%mend Color;

%Color (06); /*call the macro for each year*/

```

Table of Contents

2006 'BROWN' by ' SNEAKERS '	7
2006 'BROWN' by ' CASUALS/FLATS'	7
2006 'BROWN' by ' HEELS/WEDGES'	7
2006 'BROWN' by ' SANDALS '	7
2006 'BROWN' by ' BOOTS'	7
2006 'BLACK' by ' SNEAKERS'	8
2006 'BLACK' by ' CASUALS/FLATS'	8

<EXTRA LINES OF TABLE OF CONTENTS DELETED>

WHITE	SNEAKERS			
	1-Yes		2-No	
		RowPctN		RowPctN
1-Yes	985	98.70	13	1.30
2-No	1	50.00	1	50.00

Figure 3. Output from PROC TABULATE to RTF document

After running the program to get the RTF document we then had to do some manual manipulation to the resulting file to populate the Table of Contents (TOC). This was achieved by selecting all of the text by holding down Ctrl-A and then pressing the F9 key to populate the TOC with the individual tables in the body of the report. The output (Figure 3) was an improvement from earlier versions because it allowed the user to jump to the tables directly from the TOC. By holding the CTRL button down and clicking the table name in the TOC the cursor moves directly to the table of interest. This is still hard for the user to maneuver because they would then have to navigate back to the Table of Contents in order to move to a different table in the output. But, it was getting much closer to the user-friendly output we were hoping for.

PROC TABULATE TO PDF

Building upon the Proc RTF version, we decided to create a PDF version of the PROC TABULATE output with meaningful bookmarks in place of the RTF Table of Contents.

Final PROC TABULATE to PDF code:

```
/*Set up some macro variables to describe the variables*/
%let ColorBox10= 'YELLOW';
%let ColorBox9 = 'ORANGE';
%let ColorBox8 = 'GREEN';
%let ColorBox7 = 'PURPLE';
%let ColorBox6 = 'BLUE';
```

```

%let ColorBox5 = 'PINK';
%let ColorBox4 = 'RED';
%let ColorBox3 = 'WHITE';
%let ColorBox2 = 'BLACK';
%let ColorBox1 = 'BROWN';

%let StyleBox5 = 'BOOTS';
%let StyleBox4 = 'SANDALS';
%let StyleBox3 = 'HEELS/WEDGES';
%let StyleBox2 = 'CASUALS/FLATS';
%let StyleBox1 = 'SNEAKERS';

proc format;    /*Format to show meaningful text for the responses*/
Value YesNoFMT
    2 = '2-No'
    1 = '1-Yes';
run;

%Macro RunYr (i,yr);
/*create a PDF file for the output*/
ods pdf file = "/mydir/tabulateBookmark.pdf" pdftoc=1 startpage=never;

%do j=1 %to 5; /*loop through Style1-Style5*/
    /*creates the meaningful bookmarks*/
    ods proclabel="20&yr. &Colorbox&i. by &Stylebox&j.";

    proc tabulate data=in.shoes&yr.
        contents = ''/*gets rid of 2nd bookmark node*/
        missing; /*Include any missing values*/
        Class Color&i. Style&j.; /*Variables to be grouped*/
        table Color&i.=' ' /*Specifies the text for the box above the row*/
            , (Style&j. )*(n*f=6.0 rowpctn)/ /*Include the number and freq*/
        printmiss misstext='0' /*Show 0 instead of . for missing counts*/
        box=&&ColorBox&i /*puts meaningful label text*/
        contents = ''; /*gets rid of 3rd bookmark node*/
        KEYLABEL N = " "; /*put a blank above the number column*/
        label Style&j. = &&StyleBox&j.; /*puts meaningful label text*/
        format Color&i. Style1-Style5 YesNoFMT.; /*formats the responses*/
        title "20&yr. SHOES";
    run;
%end;
%mend;

%MACRO Color (yr);
    %do i=1 %to 10; /*Loop through the variables Color1-Color10*/
        %RunYr(&i,&yr.); /*Call the macro(RunYr) for each year*/
    %end;
%mend Color;
%Color (06); /*call the macro for each year*/

```

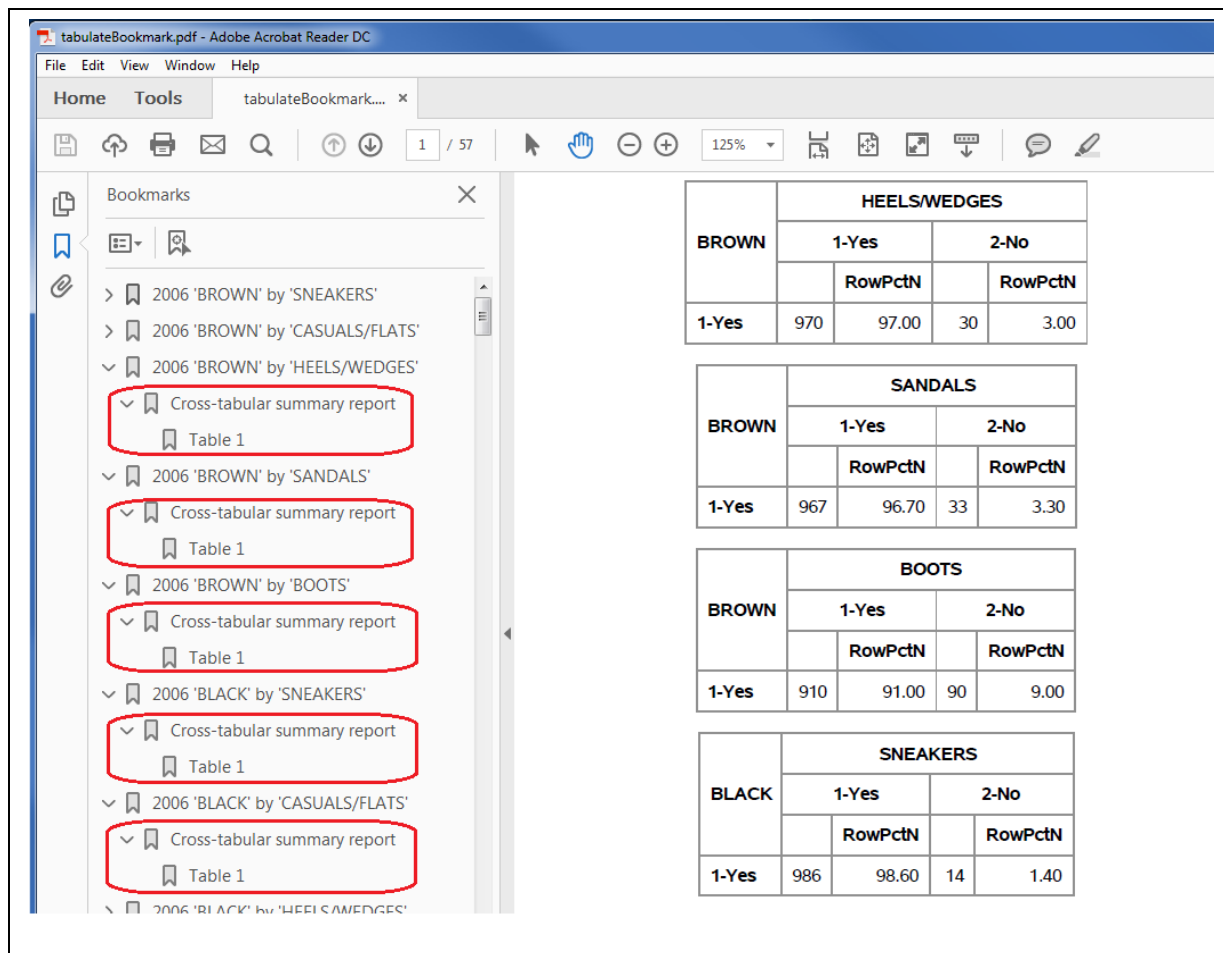


Figure 4. Output from PROC TABULATE to PDF with 3 nodes

A few options allow us to control the appearance of our bookmarks. Without them, three nodes of each bookmark will be displayed. As seen in Figure 4, the second and third levels are repeated throughout for each table created. The ODS option `pdftoc=1` displays the top level bookmark, however the second and third nodes are still present, just not visible without clicking the arrow on the left. To prevent the second node from being created we used the Proc Tabulate option `contents = "`. We needed one more `contents` option to stop the third bookmark level. This time it is an option on the Table statement within Proc Tabulate.

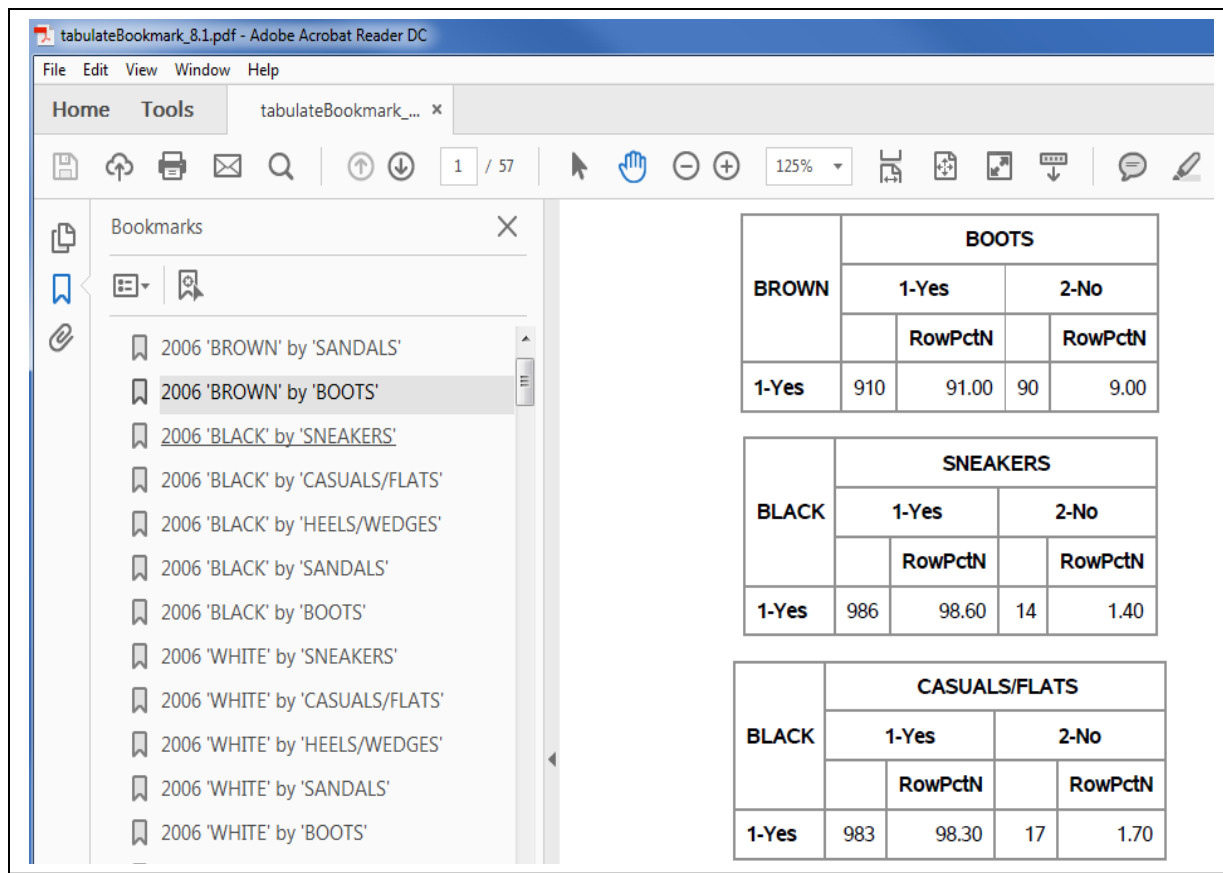


Figure 5. Output from PROC TABULATE to PDF

This final PDF version of the output (Figure 5) was exactly what we were hoping for. The PDF bookmarks allow the user to easily jump directly to the desired table and they remain visible beside the output so the user can easily move throughout the document. The bookmarks and tables include descriptive text instead of just the variable names (color1-color10) and (style1-style5) so that users who aren't familiar with the variables can easily understand the output.

CONCLUSION

While the PROC FREQ and PROC TABULATE produced the requested information the outputs were difficult to read and to maneuver due to the number of tables that were produced. The RTF version of the output was better and contained a table of contents for jumping to the desired table but still not as user-friendly as we desired. The RTF version also required some manual manipulation to populate the Table of Contents. The PDF version of the output was much more user-friendly allowing the user to jump around the tables with ease. The independent side bar of bookmarks in the PDF output is a key benefit. The analyst has the ability to scroll on the navigation pane without changing the position in the actual tables. An additional advantage of the PDF version is that it is not editable without specific PDF editing software so the output will be better preserved.

CONTACT INFORMATION

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

Barbara Felts
RTI, International
b felts@rti.org