Surveys: Finding the message in the tables

Alice Feldesman, U.S. Government Accountability Office

ABSTRACT

The Executive
As an executive you want to manage your organization using data. It gives you the bottom line and tells you if you are heading in the right direction.   Sometimes you call in your research team and you ask them questions that are on your mind---and they say "You'll need to do a survey to answer that one".

The Research Team

We've lined up people who know how to write the questions and we have someone to write the code to analyze the survey. So for each question with a typical 5 option rating and a no basis to judge and some people left the question blank. So all together for each question there would be 7 percentages.   But let's make it more informative by showing for each question the demographics. In the case study described in this article there would be 22 rows for each of 100 questions. Let's see 100 question times 22 rows times 7 response options. That would mean we'd be delivering 15,400 numbers.

My experience

This happened to me exactly like it was described above. Fortunately, the boss was transparent and would take those 15,400 tables and try over a few weekends to find a few hours to transcribe the relevant numbers into a spreadsheet so maybe the message would jump out. How many times have you said to yourself "A computer can do what you are doing manually? Please stop! Let me figure out how to get the computer to do what you are doing." This article describes how to get to the message when you are faced with 15,400 percentages.

INTRODUCTION

Executives are relying on data to manage their organizations and often surveys are conducted. The research team typically generates many tables of the results with demographic analyses. This paper demonstrates a way of simplifying the presentation so that the message is more easily found. This enables the executive to not only get to the message more quickly, but the data become more relevant since they are more timely.

THE EXECUTIVE

As an executive you want to manage your organization using data. It gives you the bottom line and tells you if you are heading in the right direction.   Sometimes you call in your research team and you ask them questions that are on your mind---and they say "You'll need to do a survey to answer that one".   So while they take time and can be somewhat costly, you totally get their value to you and the organization. So off the research team goes and a few weeks or even months later they show up with lots of tables and say "here's the results of the survey". You thank them profusely and then just stare at all of those tables.

THE RESEARCH TEAM

As a member of the research team we are glad the boss values our contribution and so we were busy doing this next survey. We've lined up people who know how to write the questions and we have someone to write the code to analyze the survey. So we plan to deliver a table for each question in the survey. And, each question had a typical 5 option rating and a no basis to judge and some people left the question blank. So all together for each question there would be 7 percentages.   But as a research team

we decided to make it more informative by showing for each question the demographics. So if we have 4 demographic questions, we should have a table for each question that has the results for everyone, but also for each of the demographics. In the case study described in this article there would be 22 rows for each of 100 questions. This indicated that for 100 question times 22 rows times 7 response options. That would mean we'd be delivering 15,400 numbers. As a research team we highly respected our boss and we know the organization values our contribution.   We expect that some how the boss and the rest of the team will figure out the message.

MY EXPERIENCE

The scenario presented above in The Executive and The Research Team, happened exactly as describe.

Fortunately, the boss was transparent and would take those 15,400 table percentages and try over a few weekends to find the relevant numbers into a spreadsheet.

How many times have you said "A computer can do what you are doing manually? Please stop! Let me figure out how to get the computer to do what you are doing."

This paper describes how to get to find the message when faced with multiple percentages in multiple tables. Taking the time to do the additional analysis described below will help your boss find the message from the data. You will want all the tables as backup if you need to deep dive into any questions the boss may have, finding the message in the data will be easier.

This analysis will help you find potentially important messages that you may have missed when looking at multiple tables. So take this journey and read this paper and implement this analysis. The boss will love you for it. Trust me.

ANALYSIS

Case Study

This case study will be using SAS® code to describe how this process was used for a survey at the Government Accountability Office (GAO) on its telework evaluation.   As the agency introduced teleworking, it was important to the executives to periodically take the "pulse" of the roll out to see if any adjustments were needed.   So it was very important to get to the message as quickly as possible, since employees were waiting to start teleworking as soon as possible.

Set up of the table for the executive

Presenting the data to an executive needs to be well thought out. Instead of setting up the table as rows for the demographics variables, let the rows to be the questions on the survey and the columns to be the demographic variables. For example if a survey had 100 questions have 100 rows. Since a piece of paper can accommodate 100 rows the research team can generate two pieces of paper rather than 100 pages of tables to tell the story of the data.

If there were 22 demographic variables, the columns will be the 22 demographics which may fit nicely on a page depending on the number of categories for each demographic. The research team will need to determine what percentages for each demographic variable needed.

For example, if a variable is a 7-category variable, it is unlikely that all 7 percentages need to be presented. Rather, as typically seen in survey result presentation only the positive, neutral, negative, and "other" responses need to be presented. So at most 4 response options need to be presented for each variable. For 22 demographic questions this would result in 88 columns of numbers. This could be further reduced to the strengths (positives) and weakness (negatives or opportunities for improvement) or 44 columns of numbers.   In the end, only have a few pieces of paper will be presented to the executive to aid in determining the message rather than 100 pages of 15,400 numbers.

Additionally, color can be help in the presentation of the results. Color can be intuitive. For example, Green is good and Red is bad. Having the research team choose a cutpoint for color coding will aid in

the presentation. For example, if the positives are greater than 85% color the number green. If the negatives are greater than 15% color the number red. These numbers can be determined collaboratively with the executive and the research team. The color also aids in determining if demographic groups differ. As an executive reads down a column, they can see if a demographic group holds similar views across a variety of questions.

SAS SOFTWARE CODE

The traditional way of creating tables using PROC TABULATE.

1. Choose a path to create the an1 file.
2. Create a macro where I can specify any skip patterns, the survey question, and the title for the table
3. Repeat the invocation of the macro for each of the variables in the survey.

Tip: Before examining the output I have a tip.

Tip1: To create the invocation the line that begins %an3 for each of the questions in the survey, use the output of PROC CONTENTS to have SAS software write its own code.   If you do it this way you can never have the incorrect title on a table

```
PROC CONTENTS NOPRINT DATA=FIN2 OUT=NEWONE; RUN;

DATA NEWONE2; SET NEWONE; KEEP NAME VARNUM LABEL; IF FORMAT NE ' '; RUN;

DATA NEWONE3; LENGTH NAME $6.;  SET NEWONE2;
PT1="%AN1 (VAR=";
PT2=PT1||LEFT(TRIM(NAME));
PT3=PT2||', TITLE="';
PT4=PT3||LEFT(TRIM(LABEL));
PT5=PT4||'");RUN;

PROC SORT DATA=NEWONE3; BY VARNUM;
PROC PRINT DATA=NEWONE3; VAR PT5; ;RUN;

FILENAME OUT 'G:\TELEWORKEREVAL\ANALYSIS\DATA.TXT';

DATA OUT; FILE OUT LRECL=32532; SET NEWONE3;
PUT
PT5 '09'X;
RUN;
```

Traditional Method

Below is the code for the traditional PROC TABULATE for the first set of questions from the GAO telework survey. This code produced ten pages of tables from this tabulate. Of course, many of you would expect to recode the data and collapse categories. But initially I believe you would have this as your first deliverable.

```
ODS HTML BODY="S:\CDMA\FELDESMANA\TELEWORKEVAL\ANALYSIS\TABULATE
 _EXAMPLE1.HTML" (TITLE="POSITION TELEWORKER 2018 SURVEY')
HEADTEXT="<STYLE> HR {PAGE-BREAK-AFTER;ALWAYS} THEAD {DISPLAY;TABLE-HEADER-
GROUP}  @MEDIA ALL {FONT {FONT-SIZE=80%}} </STYLE>" STYLE=STYLES.TEST1;
```

```
%MACRO AN1 (VAR=VAR, TITLE=TITLE);
PROC TABULATE DATA=FIN MISSING; CLASS &VAR/PRELOADFMT;
WHERE 1<=&VAR<=6 AND FINISH=1 ;
CLASS Q17;

TABLES ALL='TOTAL'   Q17='POSITION' ,
(&VAR=' '*(PCTN<&VAR>='%'*F=6.1 N='NUMBER OF CASES'*F=COMMA8.) ALL*N='TOTAL
CASES'*F=6.0) /RTS=55 PRINTMISS MISSTEXT='0';
TITLE &TITLE;  RUN;
%MEND AN1;

%AN1 (VAR=Q1A,TITLE= 'Q1A. TOPIC 1'); RUN;
%AN1 (VAR= Q1B,TITLE= 'Q1B. TOPIC 2'); RUN;
%AN1 (VAR= Q1C,TITLE= 'Q1C. TOPIC 3'); RUN;
%AN1 (VAR= Q1D,TITLE= 'Q1D. TOPIC 4'); RUN;
%AN1 (VAR= Q1E,TITLE= 'Q1E. TOPIC 5'); RUN;
%AN1 (VAR= Q1F,TITLE= 'Q1F. TOPIC 6'); RUN;
%AN1 (VAR= Q1G,TITLE= 'Q1G. TOPIC 7'); RUN;
%AN1 (VAR= Q1H,TITLE= 'Q1H. TOPIC 8'); RUN;
%AN1 (VAR= Q1I,TITLE= 'Q1I. TOPIC 9'); RUN;
%AN1 (VAR= Q1J,TITLE= 'Q1J. TOPIC 10'); RUN;
%AN1 (VAR= Q1K,TITLE= 'Q1J. TOPIC 11'); RUN;
%AN1 (VAR= Q1L,TITLE= 'Q1J. TOPIC 12'); RUN;

ODS HTML CLOSE; RUN;
```

The output of the tabulate appears at the end of this paper. (I am only showing you the table of topic 1. There would be a separate table for topic 2 to topic 12.) While you will need this detailed table for your follow on analyses, you will never figure out the message in a timely way from this complex table.

Alternative Method

Below is the SAS code for my alternative method. While this code is more complex, the interpretation of the tables is considerably easier.

The basic approach is:

- Create data sets of the total and for each of the subgroups through transforming your rows into columns
- Create a "place holder" since not all response categories might be selected. If a category is missing, then PROC TRANSPOSE will shift your data and your results will be wrong.
- Set cuptoints. The cut points of 5 percent negative and 85 percent positive are arbitrary. You might even consider running several versions of these cut points to get to a variety of messages.
- Determine the order of presentation. The presentation is made in the order of the survey. Consider sorts by the values in any of the columns and make tables so that you can readily see the most extreme values.

SAS CODE

OPTIONS SOURCE NOSTIMER NOCENTER PAGESIZE = 60 LINESIZE = 132 NOQUOTELENMAX;

/*COMMENT: THIS TEMPLATE CONTROLS THE APPEARANCE OF THE FONTS AND USES A MINIMAL STYLE SHEET*/

PROC TEMPLATE;

```
DEFINE STYLE STYLES.TEST;
PARENT=STYLES.MINIMAL;
STYLE SYSTEMTITLE FROM SYSTEMTITLE /
  FONT_FACE=HELVETICA FONT_SIZE=3
  FONT_WEIGHT=BOLD JUST=L;
STYLE SYSTEMFOOTER FROM SYSTEMFOOTER /
 FONT_FACE=HELVETICA FONT_SIZE=3;
STYLE HEADER FROM HEADER/
 FONT_FACE=HELVETICA
  JUST=L VJUST=B;
STYLE DATA FROM DATA/
 FONT_FACE=HELVETICA;
STYLE ROWHEADER FROM ROWHEADER/
 FONT_FACE=HELVETICA;
 STYLE TABLE FROM TABLE /
   JUST=LEFT; END; RUN;
```

/*Comment: This template controls the appearance of the fonts and uses a sasweb style sheet. This is desirable as it shades every other row for readability*/

```
ODS PATH WORK.TEMPLAT(UPDATE) SASHELP.TMPLMST(READ);
PROC TEMPLATE;
 DEFINE STYLE STYLES.TEST2;
 PARENT=STYLES.SASWEB;
 CLASS ROWHEADER /
 PROTECTSPECIALCHARS=OFF;
  END;
RUN;
```
/*Comment: The libname shows where the data are stored. The formats are the translation of the response options to words to represent the scale respondents were given.*/

```
LIBNAME QPLDIR 'S:\CDMA\FELDESMANA\TELEWORKEVAL\ANALYSIS';
```

```
PROC FORMAT;
    VALUE  _0001_
             1 = 'VERY POSITIVE IMPACT'
             2 = 'GENERALLY POSITIVE IMPACT'
             3 = 'NO IMPACT'
             4 = 'GENERALLY NEGATIVE IMPACT'
             5 = 'VERY NEGATIVE IMPACT'
             6 = 'NOT APPLICABLE/NO BASIS TO JUDGE'
             7=  'NO RESPONSE';


        VALUE   _0010_
             1='PDP, BAND I OR BAND II ANALYST, ANALYST-RELATED (SPECIALIST),
                COMMUNICATIONS ANALYST (BAND I OR II)'
             2='ADMINISTRATIVE PROFESSIONAL AND SUPPORT STAFF (APSS) (AC OR PT)'
             3='BAND III ANALYST, ANALYST-RELATED (SPECIALIST),  SUPERVISORY
                COMMUNICATIONS ANALYST (BAND III), SUPERVISORY OR NON-SUPERVISORY
                ATTORNEY IN PA PAYPLAN, OR MANAGERIAL AND SUPERVISORY (MS)'
             4='SES OR SENIOR LEVEL (SL); INCL ATTORNEYS IN SES PAYPLAN';
               RUN;
```

/*Comment: This is the format that will be used to do the color coding*/

```
PROC FORMAT;
        VALUE NEG  5-HIGH='RED';
        VALUE POS  75-HIGH ='LIGHT GREEN'; RUN;
```
/*Comment: I set one=1 because I want a generally place holder for any where clauses*/

```
DATA FIN; SET QPLDIR.FIN; ONE=1;  RUN;
```

/*Comment: These are the placeholders. When they are merged with the data set all response options
will be filled either with a zero or a value. If you do not perform this step the proc transpose will shift your
data*/


```
DATA HOLDERPOS; INPUT COLS VAL  @@; CARDS;
1 0 2 0 3 0
;
DATA HOLDERPOS2; INPUT COLS VAL  @@; CARDS;
1 0 2 0 3 0 4 0 5 0 6 0 7 0 8 0 9 0 10 0 11 0 12 0
;
RUN;
```
/*Comment: This is the macro that creates the collapsing of the scale into fewer response options*/


```
%MACRO CH (VAR=VAR,DEMO=DEMO,LAB=LAB, WHERE=WHERE);
DATA FINY;   SET FIN;  DEMO=&DEMO; IF FINISH=1;
IF 1<=&VAR<=3 THEN NEW&VAR=1; ELSE IF 4<=&VAR<=5 THEN NEW&VAR=2; ELSE
NEW&VAR=3;
```

/*Comment: This frequency produces an agency wide percentages. The sparse option provides all
possible combinations of levels of the variables in the table, even when some combination levels do not
occur in the data.*/


```
PROC FREQ DATA=FINY NOPRINT; TABLES NEW&VAR/OUT=GAO&VAR SPARSE; WHERE
```

&WHERE; DATA GAO&VAR; SET GAO&VAR; DROP COUNT; RUN;

```
DATA GAO&VAR; SET GAO&VAR;
IF NEW&VAR=1 THEN COLS=1;
IF NEW&VAR=2 THEN COLS=2;
IF NEW&VAR=3 THEN COLS=3;
RUN;
```

PROC SORT DATA=GAO&VAR; BY COLS;PROC SORT DATA=HOLDERPOS; BY COLS; RUN;

/*Comment: This is where the place holder is merged with the data set so that all response options will occur in the data set*/

```
DATA GAO&VAR; MERGE GAO&VAR HOLDERPOS; BY COLS;
IF PERCENT=. THEN PERCENT=0;
IF COLS=1 AND NEW&VAR=. THEN NEW&VAR=1;
IF COLS=2 AND NEW&VAR=. THEN NEW&VAR=2;
IF COLS=3 AND NEW&VAR=. THEN NEW&VAR=3;
RUN;
```

/*Comment: This transpose makes the percentages become columns rather than rows*/

PROC TRANSPOSE DATA=GAO&VAR OUT=TGAO&VAR; RUN;

/*Comment: This data step keeps only the percentage information. In order to keep track of which percentages are gao-wide versus by demographics, the variable is renamed perct*/

DATA TGAO&VAR; SET TGAO&VAR; IF _NAME_='PERCENT'; IF _NAME_='PERCENT' THEN _NAME_='PERCT'; RUN;

/*Comment: In order to keep track of which percentages are gao-wide versus by demographics, the columns 1 to 3 are renamed 13 to 15. The selection of 13 to 15 is dependent on the number of levels of the demographic variable being analyzed. In this case we are looking at a demographic variable with 4 levels with the collapsing of the response options into 3 levels. Hence, the percentages for the demographic variable will take on 12 levels. That is why the agency wide results are coded 13 to 15.*/

```
DATA TGAO&VAR; SET TGAO&VAR;
COL13=COL1;
COL14=COL2;
COL15=COL3;

DROP COL1 COL2 COL3 _LABEL_; RUN;
PROC SORT DATA=FINY;BY &DEMO;  RUN;
```

/*Comment: This is the end of the creation of the agency wide results. The same approach is used for the demographic variable that has 4 levels.*/

/*Comment: We begin again with the frequencies being retained in a data set using the sparse option. You will need frequencies as the percentages will be calculated in a later data step*/

PROC FREQ DATA=FINY NOPRINT; TABLES &DEMO*NEW&VAR/OUT=&VAR&DEMO SPARSE; WHERE 1<=&DEMO<=4 AND &WHERE; RUN;

/*Comment: This takes the data set and creates 12 columns of the results. One for each of the 3 recodes of the response options with each of the 4 levels of the demographic variable*/

DATA &VAR&DEMO; SET &VAR&DEMO;

```
IF NEW&VAR=1 AND &DEMO=1 THEN COLS=1;
IF NEW&VAR=2 AND &DEMO=1 THEN COLS=2;
IF NEW&VAR=3 AND &DEMO=1 THEN COLS=3;
IF NEW&VAR=1 AND &DEMO=2 THEN COLS=4;
IF NEW&VAR=2 AND &DEMO=2 THEN COLS=5;
IF NEW&VAR=3 AND &DEMO=2 THEN COLS=6;
IF NEW&VAR=1 AND &DEMO=3 THEN COLS=7;
IF NEW&VAR=2 AND &DEMO=3 THEN COLS=8;
IF NEW&VAR=3 AND &DEMO=3 THEN COLS=9;
IF NEW&VAR=1 AND &DEMO=4 THEN COLS=10;
IF NEW&VAR=2 AND &DEMO=4 THEN COLS=11;
IF NEW&VAR=3 AND &DEMO=4 THEN COLS=12;
RUN;


PROC SORT DATA=&VAR&DEMO; BY COLS;

PROC SORT DATA=HOLDERPOS2; BY COLS;  RUN;
```

/*Comment: Here is where the place holder is used so that all combinations will either be 0 or some value. This ensures that the proc transpose does not distort the data*/

```
DATA &VAR&DEMO; MERGE &VAR&DEMO HOLDERPOS2; BY COLS;

IF PERCENT=. THEN PERCENT=0;
IF COLS=1 AND NEW&VAR=. AND &DEMO=1 THEN NEW&VAR=1;
IF COLS=2 AND NEW&VAR=. AND &DEMO=1 THEN NEW&VAR=2;
IF COLS=3 AND NEW&VAR=. AND &DEMO=1 THEN NEW&VAR=3;
IF COLS=4 AND NEW&VAR=. AND &DEMO=2 THEN NEW&VAR=4;
IF COLS=5 AND NEW&VAR=. AND &DEMO=2 THEN NEW&VAR=5;
IF COLS=6 AND NEW&VAR=. AND &DEMO=2 THEN NEW&VAR=6;
IF COLS=7 AND NEW&VAR=. AND &DEMO=3 THEN NEW&VAR=7;
IF COLS=8 AND NEW&VAR=. AND &DEMO=3 THEN NEW&VAR=8;
IF COLS=9 AND NEW&VAR=. AND &DEMO=3 THEN NEW&VAR=9;
IF COLS=10 AND NEW&VAR=. AND &DEMO=4 THEN NEW&VAR=10;
IF COLS=11 AND NEW&VAR=. AND &DEMO=4 THEN NEW&VAR=11;
IF COLS=12 AND NEW&VAR=. AND &DEMO=4 THEN NEW&VAR=12;
RUN;

PROC SORT DATA=&VAR&DEMO; BY &DEMO; RUN;
```

/*Comment: This is the place where the denominator is calculated so that the percentages for each of the demographic levels can be calculated*/

```
PROC MEANS DATA=&VAR&DEMO SUM NOPRINT; VAR COUNT; BY &DEMO; OUTPUT
OUT=S&VAR&DEMO SUM=SCOUNT; RUN;
DATA S&VAR&DEMO; SET S&VAR&DEMO; DROP _TYPE_ _FREQ_; RUN;
PROC SORT DATA=S&VAR&DEMO; BY &DEMO; RUN;
```

/*Comment: This is where the percentages are calculated for each of the 12 columns*/

```
DATA M&VAR&DEMO;  MERGE &VAR&DEMO S&VAR&DEMO;BY &DEMO;
PERCT=(COUNT/SCOUNT)*100;
IF PERCT=. THEN PERCT=0;
RUN;
```

```
PROC SORT DATA=M&VAR&DEMO; BY COLS; RUN;
```

/*Comment: Here is where the transpose is conducted so that the data set will have columns rather than rows of the percentages.*/

```
PROC TRANSPOSE DATA=M&VAR&DEMO OUT=T&VAR&DEMO; RUN;

data t&var&demo; length lab $200.; set t&var&demo; if _NAME_='perct'; lab=&lab;  run;
```

/*Comment: Finally the agency wide and the demographic percentages are combined.*/

```
DATA &VAR&DEMO; MERGE TGAO&VAR T&VAR&DEMO; RUN;

%MEND CH;
```

/*Comment: Here is where each of the survey questions are run through the macro along with the specific demographic variable. To maintain the integrity of the data I needed to specify simply Topics 1 to 12. In the actual analysis the title of the specific questions are used.*/

```
%CH (WHERE=ONE=1,VAR=Q1A, DEMO=Q19, LAB='Q1A. TOPIC 1');RUN;
%CH (WHERE=ONE=1,VAR=Q1B, DEMO=Q19,LAB  = 'Q1B. TOPIC 2'); RUN;
%CH (WHERE=ONE=1,VAR=Q1C, DEMO=Q19,LAB  = 'Q1C. TOPIC 3'); RUN;
%CH (WHERE=ONE=1,VAR=Q1D, DEMO=Q19,LAB  = 'Q1D. TOPIC 4'); RUN;
%CH (WHERE=ONE=1,VAR=Q1E, DEMO=Q19,LAB  = 'Q1E. TOPIC 5'); RUN;
%CH (WHERE=ONE=1,VAR=Q1F, DEMO=Q19,LAB  = 'Q1F. TOPIC 6'); RUN;
%CH (WHERE=ONE=1,VAR=Q1G, DEMO=Q19,LAB  = 'Q1G. TOPIC 7'); RUN;
%CH (WHERE=ONE=1,VAR=Q1H, DEMO=Q19,LAB  = 'Q1H. TOPIC 8'); RUN;
%CH (WHERE=ONE=1,VAR=Q1I, DEMO=Q19,LAB  = 'Q1I. TOPIC 9'); RUN;
%CH (WHERE=ONE=1,VAR=Q1J, DEMO=Q19,LAB  = 'Q1J. TOPIC 10'); RUN;
%CH (WHERE=ONE=1,VAR=Q1K, DEMO=Q19,LAB  = 'Q1K. TOPIC 11'); RUN;
%CH (WHERE=ONE=1,VAR=Q1L, DEMO=Q19,LAB  = 'Q1L. TOPIC 12'); RUN;
```

/*Comment: Here is where all of the survey questions are stacked on top of each other so that the rows are the survey questions and the columns are the percentages.*/

```
DATA ALL;  SET Q1AQ19 Q1BQ19 Q1CQ19 Q1DQ19 Q1EQ19 Q1FQ19 Q1GQ19 Q1HQ19 Q1IQ19 Q1JQ19 Q1KQ19 Q1LQ19; RUN;

ODS LISTING CLOSE;
```

/*Comment: This is a code to write out a file that I called Position_example_2018.html*/

```
ODS HTML
BODY="S:\CDMA\FELDESMANA\TELEWORKEVAL\ANALYSIS\POSITION_EXAMPLE_2018.HTML"
(TITLE='POSITION TELEWORKER 2018 SURVEY') HEADTEXT="<STYLE> HR {PAGE-BREAK-AFTER:ALWAYS} THEAD {DISPLAY:TABLE-HEADER-GROUP}  @MEDIA ALL {FONT {FONT-SIZE=80%}} </STYLE>" STYLE=STYLES.TEST;
```

/*Comment: This is a basic proc report to create a table*/

```
PROC REPORT DATA=ALL NOWINDOWS SPACING=1 PSPACE=1 SPLIT='*' MISSING HEADLINE;
COLUMN LAB COL13 COL14 COL15 COL1 COL2 COL4 COL5 COL7 COL8 COL10 COL11;
```

/*Comment: Here you will the use of the style column background where I use the pos. format for positive color coding and neg. format for the negative color coding*/

```
DEFINE LAB /DISPLAY ' ' FORMAT=$220. STYLE ={JUST=LEFT CELLWIDTH=2IN};
DEFINE COL13/DISPLAY 'GAO-WIDE*POS*1' FORMAT=4.1
STYLE(COLUMN)={BACKGROUND=POS.};
```

```
DEFINE COL14/DISPLAY 'GAO-WIDE*NEG*2' FORMAT=4.1
STYLE(COLUMN)={BACKGROUND=NEG.};
DEFINE COL15/DISPLAY 'GAO-WIDE*UNK*3' FORMAT=4.1
STYLE(COLUMN)={BACKGROUND=OTH.};
DEFINE COL1/DISPLAY 'STAFF*POS*1' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=POS.};
DEFINE COL2/DISPLAY 'STAFF*NEG*2' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=NEG.};
DEFINE COL4/DISPLAY 'SUPPORT*POS*1' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=POS.};
DEFINE COL5/DISPLAY 'SUPPORT*NEG*2' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=NEG.};
DEFINE COL7/DISPLAY 'MANAGER*POS*1' FORMAT=4.1
STYLE(COLUMN)={BACKGROUND=POS.};
DEFINE COL8/DISPLAY 'MANAGER*NEG*2' FORMAT=4.1
STYLE(COLUMN)={BACKGROUND=NEG.};
DEFINE COL10/DISPLAY 'SES*POS*1' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=POS.};
DEFINE COL11/DISPLAY 'SES*NEG*2' FORMAT=4.1 STYLE(COLUMN)={BACKGROUND=NEG.};
TITLE1 'ALL RESPONDENTS-BY POSITION'; RUN;
```

/*Comment: Here is the more tradition code that uses tabulate to create a table for a total and each of the demographic variables. To maintain the integrity of the data I needed to specify simply Topics 1 to 12. In the actual analysis the title of the specific questions are used.*/


```
%MACRO AN1 (VAR=VAR, TITLE=TITLE);

PROC TABULATE DATA=FIN MISSING; CLASS &VAR/PRELOADFMT; WHERE 1<=&VAR<=7 AND
FINISH=1; CLASS Q19;

TABLES ALL='TOTAL' Q19='POSITION',

(&VAR=' '*(PCTN<&VAR>='%'*F=6.1 N='NUMBER OF CASES'*F=COMMA8.) ALL*N='TOTAL
CASES'*F=6.0) /RTS=55 PRINTMISS MISSTEXT='0';TITLE &TITLE;  RUN;

%MEND AN1;


%AN1 (VAR=Q1A, TITLE  = 'Q1A. TOPIC 1'); RUN;
%AN1 (VAR=Q1B, TITLE  = 'Q1B. TOPIC 2'); RUN;
%AN1 (VAR=Q1C, TITLE  = 'Q1C. TOPIC 3'); RUN;
%AN1 (VAR=Q1D, TITLE  = 'Q1D. TOPIC 4'); RUN;
%AN1 (VAR=Q1E, TITLE  = 'Q1E. TOPIC 5'); RUN;
%AN1 (VAR=Q1F, TITLE  = 'Q1F. TOPIC 6'); RUN;
%AN1 (VAR=Q1G, TITLE  = 'Q1G. TOPIC 7'); RUN;
%AN1 (VAR=Q1H, TITLE  = 'Q1H. TOPIC 8'); RUN;
%AN1 (VAR=Q1I, TITLE  = 'Q1I. TOPIC 9'); RUN;
%AN1 (VAR=Q1J, TITLE  = 'Q1J. TOPIC 10'); RUN;
%AN1 (VAR=Q1K, TITLE  = 'Q1K. TOPIC 11'); RUN;
%an1 (var=Q1L, title  = 'Q1l. Topic 12'); run;


ODS HTML CLOSE; RUN;
```


The table that is generated now fits on one page and is color coded.

- When you read across a row you are comparing the groups on a single question.
- When you read down a column you are examining how a specific group views the various topics.
- And because this is an easy to manage single table, you can read this table into excel and calculate any differences between groups with basic excel formulas.

All respondents-by position

	GAO-wide pos 1	GAO-wide neg 2	GAO-wide unk 3	Staff pos 1	Staff neg 2	Support pos 1	Support neg 2	Manager pos 1	Manager neg 2	SES pos 1	SES neg 2
Q1a. Topic 1	81.1	17.7	1.2	83.0	16.3	88.6	6.4	75.5	23.6	66.3	32.6
Q1b. Topic 2	85.8	12.4	1.8	89.4	9.2	86.4	7.3	78.3	20.8	73.9	23.9
Q1c. Topic 3	78.4	19.7	1.9	80.0	18.4	86.4	6.8	73.8	25.5	63.0	35.9
Q1d. Topic 4	89.9	8.5	1.6	92.2	7.0	88.2	5.0	86.5	12.2	81.5	16.3
Q1e. Topic 5	68.2	29.7	2.2	70.8	27.5	79.5	13.2	60.9	38.0	43.5	54.3
Q1f. Topic 6	88.4	8.0	3.6	90.7	6.7	87.3	2.7	85.8	10.7	75.0	21.7
Q1g. Topic 7	94.8	3.1	2.1	96.0	2.5	91.4	2.3	93.6	4.9	93.5	3.3
Q1h. Topic 8	94.6	2.5	2.9	95.7	2.6	88.2	4.1	95.7	1.7	90.2	1.1
Q1i. Topic 9	94.4	3.1	2.5	95.5	3.0	88.6	3.6	95.3	3.2	92.4	0.0
Q1j. Topic 10	95.0	3.3	1.7	96.0	2.8	90.9	2.7	94.0	4.9	94.6	3.3
Q1k. Topic 11	93.5	4.6	2.0	95.0	3.8	88.6	3.6	92.1	6.4	91.3	5.4
Q1l. Topic 12	89.4	9.4	1.3	91.8	7.4	86.8	7.3	84.5	14.8	83.7	16.3

Q1a. Topic 1

	Very positive impact		Generally positive impact		No impact		Generally negative impact		Very negative impact		Not applicable/No basis to judge		No response		All
	%	Number of Cases	%	Number of Cases	%	Number of Cases	%	Number of Cases	%	Number of Cases	%	Number of Cases	%	Number of Cases	Total Cases
Total	11.6	247	14.4	306	55.1	1,172	16.1	343	1.6	33	1.1	24	0.1	2	2127
Position	5.6	2	13.9	5	55.6	20	22.2	8	0.0	0	2.8	1	0.0	0	36
0															
PDP, Band I or Band II Analyst, Analyst-Related (Specialist), Communications Analyst (Band I or II)	11.7	154	14.2	187	57.0	749	14.9	196	1.4	18	0.6	8	0.1	1	1313
Administrative Professional and Support Staff (APSS) (AC or PT)	21.4	47	23.2	51	44.1	97	6.4	14	0.0	0	5.0	11	0.0	0	220
Band III Analyst, Analyst-Related (Specialist), Supervisory Communications Analyst (Band III), Supervisory or non-Supervisory Attorney in PA payplan, or Managerial and Supervisory (MS)	8.4	39	11.8	55	55.4	258	21.0	98	2.6	12	0.6	3	0.2	1	466
SES or Senior Level (SL); including Attorneys in SES payplan	5.4	5	8.7	8	52.2	48	29.3	27	3.3	3	1.1	1	0.0	0	92

CONCLUSION

I hope my technique of switching rows and columns will enable you to quickly understand the results of your surveys. It makes the analysis easier, quicker, and more actionable.   I have found that my colleagues have quickly adapted to using this approach, but I do encourage them to also look at the overall tabulate results. The tabulate will give you additional information regarding your decision on how to best collapse your scales as well as important information about the number of respondents. If you aren't careful about how you collapsed your scales, then important information about such issues as item nonresponse and no basis to judge/not applicable might not be discovered.   If you ignore the number of respondent information, you might think you have an important message, but then you might also see that the number of respondents could be rather small.

I would welcome any comments. I can be reached at Feldesmana@gao.gov.