Creating Textual Web Reports with Graphics Using a Variety of SAS® Programming Tools
Mary Ellen Davis, U.S. Census Bureau, Washington, D.C.
Nick Spanos, U.S. Census Bureau Washington, D.C.

Abstract
A few years ago, the programming staff of the American Community Survey at the U.S. Census Bureau was asked to produce Narrative Profile reports for display on the internet. Narrative Profiles are easy-to-read reports that use plain-language descriptions and representational graphs to summarize information on a wide array of subjects using words, rather than numbers. Newspaper reporters, grant and loan applicants can quickly obtain information on key topics important to their community, and simple charts and graphs illustrate changes in those communities. This report was first produced for a small number of geographic areas and is now produced for more than 800 geographic areas. A variety of SAS tools are used to produce these reports. A report template and a set of formulas are the foundation on which the textual Web reports are built. The entire report template containing text, numbered place holders, and HTML tags is converted to a SAS dataset. Next, we use a DO LOOP and the TRANWRD function to replace each placeholder with a value. Using DATA _NULL_ and PUT statements, each processed record is output and saved in a HTML file. The graphics used in the report are produced by a separate program using SAS\GRAPH® and SAS\ODS® and saved as .GIF files. This paper and presentation will explain the SAS procedures and programs used to make these reports possible.

Introduction
We were given a report template, a set of formulas, and the mission to create a generalized report that was easy-to-read, used plain-language text, and contained easy-to-read graphs. Using SAS Version 8.2, we created a generalized reporting system that allowed us to produce as many reports, for as many geographies as we wanted.

The project that we work on at the U.S. Census Bureau is the American Community Survey. This survey collects data on a continuous basis, and produces annual and multi-year estimates of demographic characteristics for geographic areas and sub-population groups that meet pre-defined population thresholds. It is an on-going survey that the Census Bureau plans will replace the long form in the 2010 Census. The American Community Survey provides accurate, up-to-date profiles of America’s communities every year. When designing our system, we had to write programs that could be easily updated and re-used every year.

The system that we created uses four programs to build the web reports. All of the programs were written using SAS\Macro. The first two programs reformat the data using the SAS data step and arrays. Two sets of input datasets are created and are used as input by the next set of programs. The second two programs create the textual based report and the graphics. This paper will focus on the second set of programs.
Here are a few examples of the formulas:

NP(1) = PUT((PF1_11), COMMA12.);
NP(2) = PUT((PF1_17), COMMA8.1);
NP(3) = PUT(ROUND(100*(PF1_11 - PF1_18) / PF1_11), WORDS40.);
NP(4) = ROUND(100*(PF1_21 / PF1_1));

These formulas are calculated in another datastep. Some of the formulas produce character values and some produce numeric values. The array value for NP(1) corresponds to the template placeholder [1]. The number produced by the formula for NP(1) replaces the placeholder [1] with a data value.

The first datastep reads in the report template. The second datastep, using do loops, searches the template for "[]" placeholders, sets the brackets to blanks, records the placeholder location, and replaces the numbered placeholders with the data items stored in the array. See code below.

/* Read in Report Template */

data ACSNEW.rec;
infile "&file" lrecl=120
pad missover;
length line 8
inrec $120
dataset $10;
dataset = "NP&NUM";
input @1 inrec $char110.;
line+1;
run;

/* Sort dataset */

proc sort data = ACSNEW.rec;
by dataset;
run;

/* Merge Narrative Profile template with variables produced in a previous dataset */

data ACSNEW.rec2;
merge ACSNEW.rec ACSNEW.NP#
by dataset;
run;

/* Replace items in [brackets] with data values*/

Data ACSNEW.nprec;
set ACSNEW.rec2;
ARRAY NP(110) $220 NP1-NP110;
ARRAY indexa(12) indexa1-indexa12;
ARRAY indexb(12) indexb1-indexb12;

length inrec2 $220;

/* First do loop runs through each line in the document. */

Do i = 1 to _N_;
inrec = inrec;
relen = length(inrec);
/* Second do loop runs through each placeholder per line */

Do j = 1 to 12;
/* The following code stores the location of the placeholders */

indexa(j) = index(inrec2,"[");
indexb(j) = index(inrec2,"]");

/* The following code finds the placeholders "[]" and sets them to blank. */

if (indexa(j) > 0 and indexb(j) > 0) then do;
  if (indexa(j) > 0) and
  substr(inrec2,indexa(j),1) = '
  then substr(inrec2,indexa(j),1) = ';
  if (indexb(j) > 0) and
  substr(inrec2,indexb(j),1) = '
  then substr(inrec2,indexb(j),1) = ';

/* Search for numeric placeholders only. Skip the placeholders with characters. Ignore year. */

if(substr(inrec2,indexa(j)+1,(indexb(j)-indexa(j)-1)) not in ('location','year','State','state','prev year','1999','2000') then do;
  indx = put(substr(inrec2,indexa(j)+1,(indexb(j)-indexa(j))-1),12.);
/* This code replaces all of the placeholders [1] to [110] with values. SASFunction: Tranwrd(source,target,replacement) */
inrec2=tranwrd(inrec2,substr(inrec2,indexa(j)+1,(indexb(j)-indexa(j))-1)),compbl(NP(indx)));
end;
end;
run;
A third data step (similar to the previous code shown) loops through the document and converts all of the character placeholders to a set of values stored in a dataset. The [location] and [state] placeholders are replaced with county and state names using the TRANWRD function. This datastep also changes all of the graphic names to actual filenames, so that they can be displayed by HTML.

**SAS Tools: SAS Functions**

The following SAS functions were used in the program above.

INDEX: Searches a character string for a character or set of characters. When it finds the character that it is searching for, the function generates the character's position in the string.

TRANWRD: Replaces all occurrences of a word, a letter, or a number in a character string.

SUBSTR: Subsets a character string.

COMPBL: Removes two or more blanks from a character string and replaces them with 1 blank.

**Creating Graphics**

The graphics used in the report are produced and saved as .GIF's using SAS\GRAPH and SAS\ODS. The specifications call for bar charts, so we used Proc GCHART. Since the bar charts are required to have percentages displayed after each bar, the Annotate facility is used.

See sample graph code below:

```sas
%Macro crgraphs;
    /* Set the general graphics environment. */
    goptions reset=global gunit=pct
    htitle=6 pct htext=4 pct
    hsize=6 in vsize=4 in
    vpos=30 hpos=30
    ftitle='Arial' ftext='Arial';
    ods listing;
    filename grafout "C:\Narrative Profiles\2000\C2SS\6SUMLEV\img\ag&num2..gif";
    goptions device=gif noborder display
gsfname=grafout gsfmode=replace;
    /* 1. The Age Distribution of People in [location], [state] in [year] x-axis: Percent of population y-axis: Age category */
    /* Create Annotate Dataset */
    Data anno2;
    length function color style text $8;
    retain function 'label' color 'black' when 'a'
    style 'zapf'
xsys ysys '2' position '+' size 4 hsys '3'
    rotate 0 angle 0;
    set GNP.AG&NUM;
    midpoint=asrt;
    group=age;
    text=right(' '||put((age/100),PERCENT7.));
y=age/2;
    run;
    /* Set title and footnote. */
    title1 height=4 justify=center font='Arial/bo'
    "The Age Distribution of People in &name, &state in 2000";
    footnote1 justify=left "Source: U.S. Census Bureau";
    /* Set labels, tick marks, and value ranges. */
    axis1 label = (font='Arial/bo' a=90 r=0 'Age category')
    value = (tick=1 "65 and over"
    tick=2 '45 to 64'
    tick=3 '25 to 44'
    tick=4 '18 to 24'
    tick=5 'Under 18');
    axis2 label = (font='Arial/bo' "Percent of population")
    order = (0 to 100 by 10) offset=(,.25)in;
    /* Create a bar chart */
    proc gchart data = GNP.AG&NUM;
    hbar asrt/ sumvar=age
    discrete
    nostat
    anno=anno2
    width=4
    maxis=axis1
    raxis=axis2
    space=5
    cutline=black;
    run;
    quit;
    %Mend;
```

The following HTML code attaches the graph to the Web report:

```html
<p><img border="0" src="img/ag##.gif" width="570" height="380"&nbsp;></p>
```
SAS Tools: SAS\GRAPH PROC GCHART Options

The following list summarizes the SAS\GRAPH PROC GCHART options and their function in the program above:

**SUMVAR:** Calculates the values used by each bar in the bar chart. The data values are represented by the length of the bars in the chart.

**DISCRETE:** Treats a numeric variable as a discrete variable. The default is continuous. PROC GCHART creates a separate bar for each chart value.

**NOSTAT:** Suppresses default table statistics.

**ANNO:** Tells PROC GCHART to use a user defined annotate dataset for certain chart settings.

**WIDTH:** Sets the width of the bars.

**MAXIS:** Allows for user defined axis definitions to the midpoint axis.

**RAXIS** Allows for user defined labels, tick marks and value ranges.

**SPACE:** Sets spaces between the bars of the bar chart.

**COUNTLINE:** Outlines the bars of the bar chart.

Conclusion

Using SAS\Macro, Datastep programming, SAS\ODS, and SAS\GRAPH, we created a reporting system that is flexible, easy-to-use, and produces publication quality reports.

There are many advantages that our system and methodology give us when processing Web reports. The HTML placeholders in the reports automatically format the text and graphics for the reports. An unlimited number of reports can be created using a few report templates. Reports are quick and easy to produce using our system. Making changes to reports, graphics, or formulas can be done quickly. Finally, using an HTML front end, containing a map and a geography pick list, the reports look dynamic even though they are static.

We encountered a few disadvantages when using our system. All of the reports and graphics are static and not dynamic. This meant that all of the reports had to be produced at one time. The report and graphic programs produced 844 HTML files and 6,752 GIF files.

We have a few planned upgrades to our system. When time permits, we will attempt to create dynamic textual and graphical reporting system using SAS\IntrNet ® software.

A finished report is located at the end of this paper as an attachment.

References


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies

Acknowledgments

We wish to thank the following individuals for their valuable assistance to this paper:

John Stiller, James Clement, Howard McGowan and Doug Hillmer.

Contact Information

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

Mary Ellen Davis
U.S. Census Bureau
4700 Silver Hill Road, Stop 8400
Washington, D.C. 20233-8400
301-763-8508
Mary.Ellen.Davis@ccmail.census.gov

Nick Spanos
U.S. Census Bureau
4700 Silver Hill Road, Stop 8400
Washington, D.C. 20233, 8400
301-763-6841
Nicholas.M.Spanos@ccmail.census.gov
From the Census 2000 Supplementary Survey:

Population and Housing Profile: the United States

NOTES: The population estimates in this profile are limited to the household population and exclude the population living in institutions, college dormitories, and other group quarters.

In the narrative profile below, if one sample estimate is larger than another, it does not necessarily mean that the same holds for their true values for the entire population. These estimates are derived from a sample of housing units and measure the true values with a degree of uncertainty. In the tabular profiles this uncertainty is represented by the 90 percent confidence interval given for each estimate. For further information on confidence intervals see the Accuracy of the Data document.

Caution should be used when comparing data by race for years before 1999 with those for 2000 due to the Census Bureau's implementation of the October 1997 revised standards for data on race and ethnicity. For the 2000 data, respondents may report one or more races and there are seven (American Indian and Alaska Native, Asian, Black or African American, Native Hawaiian and Other Pacific Islander, White, Some other race, and Two or more races) instead of five racial tabulation categories. The two or more races category includes all respondents who reported more than one race. See the methodology section for more details regarding these changes.

In addition, caution should be used when comparing population numbers by age, race, and sex for 2000 with all earlier years. The numbers for 2000 have been weighted to be generally consistent with Census 2000 counts.

For information on confidentiality protection, sampling error, nonsampling error, and definitions, go to the Census 2000 Supplementary Survey web site www.census.gov/c2ss/www/ and click on Methodology.

POPULATION OF the United States: In 2000, the United States had a household population of 273,643,274 - 140,091,913 (51 percent) females and 133,551,361 (49 percent) males. The median age was 35.4 years. Twenty-six percent of the population were under 18 years and 12 percent were 65 years and older.

The Age Distribution of People in the United States in 2000

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Percent of Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 18</td>
<td>28%</td>
</tr>
<tr>
<td>18 to 24</td>
<td>9%</td>
</tr>
<tr>
<td>25 to 44</td>
<td>30%</td>
</tr>
<tr>
<td>45 to 64</td>
<td>22%</td>
</tr>
<tr>
<td>65 and over</td>
<td>12%</td>
</tr>
</tbody>
</table>

Source: U.S. Census Bureau, Census 2000 Supplementary Survey
For people reporting one race, 79 percent were White; 12 percent were Black or African American; 1 percent were American Indian or Alaska Native; 4 percent were Asian; less than 0.5 percent were Native Hawaiian or Other Pacific Islander, and 4 percent were Some other race. Two percent reported Two or more races. Thirteen percent of the people in the United States were Hispanic or Latino. Seventy percent of the people in the United States were not Hispanic or Latino, White alone. People of Hispanic or Latino origin may be of any race.

HOUSEHOLDS AND FAMILIES: In 2000 there were 106,905,819 households in the United States. The average household size was 2.56 people.

Families made up 68 percent of the households in the United States that year. This figure includes both married couple families (51 percent) and other families (17 percent). Non-family households made up 32 percent of all households in the United States. Most of the non-family households were people living alone, but some were people living in households maintained by non-relatives.

The Types of Households in the United States in 2000

- Married-couple families: 51%
- Other families: 17%
- People living alone: 26%
- Other nonfamily households: 6%

Source: U.S. Census Bureau, Census 2000 Supplementary Survey

NATIVITY AND LANGUAGE: Eleven percent of the people living in the United States in 2000 were foreign born. Eighty-nine percent were native, including 67 percent who were born in their state of residence.

Among people at least five years old living in the United States in 2000, 18 percent spoke a language other than English at home. Of those speaking a language other than English at home, 60 percent spoke Spanish and 40 percent spoke some other language; 43 percent reported that they did not speak English "very well."

GEOGRAPHIC MOBILITY: In 2000, 84 percent of the people at least one year old living in the United States were living in the same residence one year earlier; 10 percent had moved during the past year from another residence in the same county, 3 percent
from another county in the same state, 3 percent from another state, and 1 percent from abroad.

**Geographic Mobility of Residents of the United States in 2000**

- Same residence: 84%
- Different residence, same county: 10%
- Different county, same state: 3%
- Different state: 3%
- Abroad: 1%

Source: U.S. Census Bureau, Census 2000 Supplementary Survey

**EDUCATION:** In 2000, 82 percent of people 25 years and over had at least graduated from high school and 25 percent had a bachelor’s degree or higher. Among people 16 to 19 years old, 11 percent were dropouts; they were not enrolled in school and had not graduated from high school.

The total school enrollment in the United States was 72,722,615 in 2000. Preprimary school enrollment was 8,349,711 and elementary or high school enrollment was 48,769,926 children. College enrollment was 15,602,978.

**The Educational Attainment of People in the United States in 2000**

- Graduate or professional degree: 9%
- Bachelor’s degree: 36%
- Associate degree: 6%
- Some college, no degree: 21%
- High school diploma or equivalency: 30%
- Less than high school diploma: 18%
DISABILITY: In the United States, among people at least five years old in 2000, 16 percent reported a disability. The likelihood of having a disability varied by age - from 7 percent of people 5 to 20 years old, to 14 percent of people 21 to 64 years old, and to 41 percent of those 65 and older.

INDUSTRIES: In 2000, for the employed population 16 years and older, the leading industries in the United States were Educational, health and social services, 19 percent, and Manufacturing, 14 percent.

OCCUPATIONS AND TYPE OF EMPLOYER: Among the most common occupations were: Management, professional, and related occupations, 33 percent; Sales and office occupations, 27 percent; Service occupations, 15 percent; Production, transportation, and material moving occupations, 15 percent; and Construction, extraction and maintenance occupations, 9 percent. Eighty percent of the people employed were Private wage and salary workers; 14 percent were Federal, state, or local government workers; and 6 percent were Self-employed.

TRAVEL TO WORK: Seventy-six percent of the United States workers drove to work alone in 2000, 11 percent carpooled, and 5 percent took public transportation. Three percent worked at home. Among those who commuted to work, it took them on average 24 minutes to get to work.

INCOME: The median income of households in the United States was $41,433. Eighty-one percent of the households received earnings and 17 percent received retirement income other than Social Security. Twenty-six percent of the households received Social Security. 

Percent of people 25 years and over

Source: U.S. Census Bureau, Census 2000 Supplementary Survey

Employment by Industry in the United States in 2000

<table>
<thead>
<tr>
<th>Type of Industry</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture, forestry, fishing and hunting, and mining</td>
<td>2%</td>
</tr>
<tr>
<td>Construction</td>
<td>7%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>14%</td>
</tr>
<tr>
<td>Wholesale trade</td>
<td>4%</td>
</tr>
<tr>
<td>Retail trade</td>
<td>12%</td>
</tr>
<tr>
<td>Transportation and warehousing, and utilities</td>
<td>5%</td>
</tr>
<tr>
<td>Information</td>
<td>3%</td>
</tr>
<tr>
<td>Finance, insurance, real estate and rental and leasing</td>
<td>7%</td>
</tr>
<tr>
<td>Professional and business services</td>
<td>9%</td>
</tr>
<tr>
<td>Educational, health and social services</td>
<td>13%</td>
</tr>
<tr>
<td>Leisure and hospitality</td>
<td>8%</td>
</tr>
<tr>
<td>Other services (except public administration)</td>
<td>5%</td>
</tr>
<tr>
<td>Public administration</td>
<td>5%</td>
</tr>
</tbody>
</table>

Percent of employed population 16 years and over

Source: U.S. Census Bureau, Census 2000 Supplementary Survey

Note: The Professional and business services category includes the following industries: Professional, scientific, management, administrative and waste management services. The Leisure and hospitality category includes the following industries: Arts, entertainment, recreation, accommodation and food services.
Security. The average income from Social Security was $11,365. These income sources are not mutually exclusive; that is, some households received income from more than one source.

POVERTY AND PARTICIPATION IN GOVERNMENT PROGRAMS: In 2000, 13 percent of persons were in poverty. Seventeen percent of related children under 18 were below the poverty level, compared with 11 percent of people 65 years old and over. Ten percent of all families and 27 percent of families with a female householder and no husband present had incomes below the poverty level. Seventeen percent of the households in the United States received means-tested public assistance or noncash benefits.

![Poverty Rates in the United States in 2000](image)

Source: U.S. Census Bureau, Census 2000 Supplementary Survey

HOUSING CHARACTERISTICS: In 2000, the United States had a total of 115,904,650 housing units, 9.6 percent of which were vacant. Of the total housing units, 66 percent were in single-unit structures, 27 percent were in multi-unit structures, and 7 percent were mobile homes. Sixteen percent of the housing units were built since 1990.

![The Types of Housing Units in the United States in 2000](image)
OCCUPIED HOUSING UNIT CHARACTERISTICS: The occupied units consisted of 66 percent that were owner occupied and 34 percent that were renter occupied. There were 529,339 (0.5 percent) and 624,524 (0.6 percent) occupied units lacking plumbing and kitchen facilities, respectively. Three percent of the households did not have telephone service and 9 percent of the households did not have access to a car, truck, or van for private use. Multi-vehicle households were not rare. Thirty-nine percent had two vehicles and another 18 percent had three or more.

HOUSING COSTS: The median monthly housing costs for (specified) mortgaged owners was $1,085, nonmortgaged owners $287, and (specified) renters $612. Twenty-seven percent of owners with mortgages, 11 percent of owners without mortgages, and 41 percent of renters in the United States spent 30 percent or more of household income on housing.

Source: U.S. Census Bureau, Census 2000 Supplementary Survey