

A SAS® SYSTEM: FROM CD TO GIS MAPS

Samuel D. Calhoun, Economic Research Service, USDA

ABSTRACT

Once a year the Rural Business and Development Policy Branch, Economic Research Service, USDA, publishes an issue of Rural Condition and Trends (RCaT), dealing with Federal Programs. This issue describes Federal program and policy changes affecting rural America. The original data comes from the Bureau of the Census in CD-ROM format. The data is in ASCII format. It consist of over 350,000 records. A transformation process of the data takes place before the RCaT is produced. The majority of this process uses base PC-SAS version 6.12 for Windows Next Technology (WINNT) and some mainframe SAS programs. In addition to the SAS programs there are some FORTRAN® programs. The purpose of this paper is to show how SAS plays a major role in manipulating the data in order to produce maps (see figures 1 and 2) for a report.

INTRODUCTION

The principal data source is the Consolidated Federal Funds Reports (CFFR) data from the U.S. Department of Commerce, Bureau of the Census, Governments Division. For each Fiscal year, four data files are constructed from the CFFR data. CFFR data is reported for 3,141 counties in the 50 States as well as for subcounty units of government, the District of Columbia, and U.S. Territories. Our data uses 3,085 counties. In order to match the CFFR data with other data sets, 56 counties in the CFFR data need to be consolidated. These 56 counties, which are separate in the GIS system, have been combined with other counties in our data. For example, some of the GIS counties in Alaska have been combined to form 14 county groups; 38 independent cities in Virginia have been

combined with the surrounding counties, resulting in only 98 counties for Virginia; five New York City boroughs are combined into just one. In Hawaii, Montana, and Wisconsin, two GIS counties have been combined into one county. The following are descriptions of the four raw data files.

1. County file - The county file has individual counties as observations, sorted by the 5-digit Federal Information Processing Standard (FIPS) State/county code.
2. County/program file - The county/program file has individual programs as observations, sorted by program ID-code.
3. State file -The State file has individual States as the unit of observation, sorted by State two-digit FIPS code.
4. State/program file - The State/program file has programs as observations, sorted by program ID.

Each program has individual characteristics that affect the way the data can be used to interpret geographic patterns.

CENSUS CD-ROM FORMAT

The CD-ROM contains fixed length ASCII text files for all fiscal years from 1983 to 1997. Each year is accompanied by four files:

1. Data file - Each data file contains government identification code, program identification code, object/assistance type code, amount in dollars, and FIPS geographic code.
2. Geographic Reference file - The Geographic

Reference file contains the name and governmental unit code for all State, county and subcounty areas in the Nation.

3. Program Identification file - The Program Identification file contains program identification codes and their respective program titles.

4. Agency code file - The Federal Agency file contains all four-digit codes and their respective agency names.

The CFFR data is a presentation of Federal Government expenditures or obligations in State, county, and subcounty areas of the United States. U.S. Territories and the District of Columbia are included.

COUNTY PROCEDURES

1. Sort the data by FIPS code using PROC SORT.

2. Run PROC COMPARE twice to compare both the previous year's program ID to the present year's program ID by program name and the previous year's program name to the present year's program name by program ID. This lets you know whether the program ID has a program name that is different from that of the previous ID and the same goes for the program name that has a different program ID. This procedure is used to create a unique code called the function code, used to determine the type of assistance. An example of a function code is Agriculture and natural resources function.

3. Run PROC FREQ by program name and object code. The object code is also an indicator of the type of assistance. An example of an object code is SW (Salaries and Wages). I call this procedure CROSSTAB.

4. Run a program that combines the GIS FIPS to the codes that are used in the Federal Funds.

5. Upload the data from the PC to the Mainframe.

6. Submit job to create county aggregate file. This program aggregates dollars allotted to Federal programs. Census reports the data to the subcounty level. This job aggregates the data to the county level.

7. Download the FORTRAN output to the PC to use the SAS program to retrieve the data.

8. Run the retrieval program to extract the data for input into the GIS system.

COUNTY/PROGRAM PROCEDURES

1. Submit job to create the county program file from the county aggregate file.

2. Download the FORTRAN output to PC in order to use the SAS program to retrieve the data.

STATE PROCEDURES

1. Submit job to create the State file.

2. Download the FORTRAN output to PC in order to use the SAS program to retrieve the data.

STATE/PROGRAM PROCEDURES

1. Submit job to create the State/program file.

2. Download the FORTRAN output to PC in order to use the SAS programs to retrieve the data.

CONCLUSION

The use of base SAS is a useful tool in creating and retrieving the Federal Funds (CFFR) data. SAS provides an alternative to interface with the FORTRAN output data. The retrieval process is easy to implement and speeds up creation of the data files. It has enhanced the overall publication

of the RCaT.

APPENDIX A: LISTING OF COUNTY AND COUNTY/PROGRAM RETRIEVE PROGRAMS

The mainframe programs are the ones with Job Control Language (JCL). The PC-SAS programs do not use JCL.

SORT PROGRAM

```
FILENAME INDAT1
'D:\CFFR1997\DATA1997.DAT' LRECL = 48;
FILENAME OUT1
'D:\CFFR1997\FIPSORT.DAT' LRECL = 48;
DATA CFFR97;
  INFILE INDAT1 ;
  INPUT PART1 $CHAR33. IFIP 34-38 REST
$CHAR10.;
PROC SORT;
  BY IFIP;
DATA OUT1;
  SET CFFR97;
  FILE OUT1;
  PUT PART1 $CHAR33. IFIP 34-38 REST
$CHAR10.;
RUN;
```

COMPARE PROGRAM #1

```
OPTIONS LABEL;
DATA FF96;
INFILE INDAT1;
INPUT PGID $1-6 NAME $7-81;
PROC SORT;
BY NAME;
DATA FF97;
INFILE INDAT2;
INPUT PGID $1-6 NAME $7-81;
PROC SORT;
BY NAME;
PROC COMPARE DATA=FF97
COMPARE=FF96;
```

```
VAR PGID NAME;
ID NAME;
RUN;
```

COMPARE PROGRAM #2

```
OPTIONS LABEL;
DATA FF96;
INFILE INDAT1;
INPUT PGID $1-6 NAME $7-81;
PROC SORT;
BY PGID;
DATA FF97;
INFILE INDAT2;
INPUT PGID $1-6 NAME $7-81;
PROC SORT;
BY PGID;
PROC COMPARE DATA=FF97
COMPARE=FF96;
VAR PGID NAME;
ID PGID;
RUN;
```

CROSSTAB PROGRAM

```
OPTIONS LABEL LINESIZE=133;
FILENAME INDAT1
'D:\CFFR1997\FIPSORT.DAT';
DATA CFFR97;
  INFILE INDAT1;
  INPUT PGMID $13-18 OBJECT $19-20;
PROC FREQ;
  TABLES PGMID*OBJECT;
RUN;
```

COMBINE PROGRAM

```
FILENAME INDAT1
'D:\CFFR1997\DATA1997.DAT' LRECL = 48;
FILENAME OUT1
'D:\CFFR1997\FIPSORT.DAT' LRECL = 48;
DATA CFFR97;
INFILE INDAT1 ;
INPUT PART1 $CHAR33. IFIP 34-38 REST
```

```

$CHAR10.;
BEAINCOME 95 ;
IF (IFIP = 2013 OR IFIP = 2016 OR IFIP = 2060
OR IFIP = 2070 OR IFIP = 2164)
  THEN IFIP = 2010;
  ELSE IF IFIP = 2185 OR IFIP = 2188 OR IFIP
= 2290 OR IFIP = 2068 THEN IFIP = 2050;
  ELSE IF (IFIP = 2201) THEN IFIP = 2130;
  ELSE IF (IFIP = 2220 OR IFIP = 2232 OR IFIP
= 2282) THEN IFIP = 2100 ;
  ELSE IF (IFIP = 2240) THEN IFIP = 2090;
  ELSE IF (IFIP = 30113) THEN IFIP = 30067;
  ELSE IF (IFIP = 36005 OR IFIP = 36047 OR
IFIP = 36081 OR IFIP = 36085)
THEN IFIP = 36061;
  ELSE IF (IFIP = 51510) THEN IFIP = 51013;
  ELSE IF (IFIP = 51515) THEN IFIP = 51019;
  ELSE IF (IFIP = 51520) THEN IFIP = 51191;
  ELSE IF (IFIP = 51530) THEN IFIP = 51163;
  ELSE IF (IFIP = 51540) THEN IFIP = 51003;
  ELSE IF (IFIP = 51560 OR IFIP = 51580)
THEN IFIP = 51005;
  ELSE IF (IFIP = 51570) THEN IFIP = 51041;
  ELSE IF (IFIP = 51590) THEN IFIP = 51143;
  ELSE IF (IFIP = 51595) THEN IFIP = 51081;
  ELSE IF (IFIP = 51600 OR IFIP = 51610)
THEN IFIP = 51059;
  ELSE IF (IFIP = 51620) THEN IFIP = 51175;
  ELSE IF (IFIP = 51630) THEN IFIP = 51177;
  ELSE IF (IFIP = 51640) THEN IFIP = 51035 ;
  ELSE IF (IFIP = 51650) THEN IFIP = 51199 ;
  ELSE IF (IFIP = 51660) THEN IFIP = 51165 ;
  ELSE IF (IFIP = 51670) THEN IFIP = 51149 ;
  ELSE IF (IFIP = 51678) THEN IFIP = 51163 ;
  ELSE IF (IFIP = 51680) THEN IFIP = 51031 ;
  ELSE IF (IFIP = 51683 OR IFIP = 51685)
THEN IFIP = 51153 ;
  ELSE IF (IFIP = 51690) THEN IFIP = 51089 ;
  ELSE IF (IFIP = 51700) THEN IFIP = 51199 ;
  ELSE IF (IFIP = 51700 OR IFIP = 51735)
THEN IFIP = 51199 ;
  ELSE IF (IFIP = 51710 OR IFIP = 51740)
THEN IFIP = 51550 ;
  ELSE IF (IFIP = 51720) THEN IFIP = 51195 ;

```

```

  ELSE IF (IFIP = 51730) THEN IFIP = 51053 ;
  ELSE IF (IFIP = 51750) THEN IFIP = 51121 ;
  ELSE IF (IFIP = 51760) THEN IFIP = 51087 ;
  ELSE IF (IFIP = 51770 OR IFIP = 51775)
THEN IFIP = 51161 ;
  ELSE IF (IFIP = 51780) THEN IFIP = 51083 ;
  ELSE IF (IFIP = 51790 OR IFIP = 51820)
THEN IFIP = 51015 ;
  ELSE IF (IFIP = 51830) THEN IFIP = 51095 ;
  ELSE IF (IFIP = 51840) THEN IFIP = 51069 ;
  ELSE IF (IFIP = 55078) THEN IFIP = 55115 ;
PROC SORT;
  BY IFIP;
DATA OUT1;
  SET CFFR97;
  BY IFIP;
  FILE OUT1;
  PUT PART1 $CHAR33. IFIP 34-38 REST
$CHAR10.;
RUN;

```

COUNTY DATA RETRIEVE PROGRAM

```

OPTIONS LABEL;
FILENAME INDAT1
'D:\CFFR1997\COUNTFIL.DAT'
LRECL=17700;
DATA FF97;
INFILE INDAT1;
INPUT FIPS 1-5 ST $7-8 COUNTY $10-37
BEALE93 38-39 FM 41 MI 43
MF 45 GV 47 TS 49 NS 51 RT 53 FL 55 CM 57
PV 59 TP 61 REGION 63 INC95 65-75
PERCAP95 77-86 POP97 88-97 POP97 100-110
(PG1-PG1256) (1256*13.0);
PUT FIPS "'ST'" "'COUNTY'" BEALE93
POP97 PG443 PG1256;
RUN;

```

COUNTY/ PROGRAM DATA RETRIEVE PROGRAM

```

OPTIONS LABEL;
FILENAME INDAT1

```

```
'D:\FEDFUNDS\CFFR1997\CPROGFIL.DAT'
LRECL=1210;
DATA FF97;
INFILE INDAT1;
INPUT PGID $1-6 OBJCODE $7-8 FUNTCODE
10-12 DISPCODE 14 PRGNAME $16-85
(PG1-PG45) (15.0) (PG46-PG90) (10.2);
PUT PGID OBJCODE FUNTCODE DISPCODE
PRGNAME PG1 PG46;
RUN;
```

APPENDIX B: LISTING OF STATE AND STATE/PROGRAM DATA RETRIEVE PROGRAMS

STATE DATA RETRIEVE PROGRAM

```
OPTIONS LABEL;
FILENAME INDAT1
'D:\FEDFUNDS\CFFR1997\STATEFIL.DAT'
LRECL=29097;
DATA FF97;
INFILE INDAT1;
INPUT STFIPS 1-2 STATENAM $3-19
REGION 20 INCOME97 21-31 POP97 32-42
APCAPITA 43-50 RST 51 UM 52 OTHER 53 FD
54 PP 55 RD 56 FL 57 (P1-P1320) (14.0)
(P1321-P2640) (8.2);
PUT STATENAM STFIPS P1 P1321;
RUN;
```

STATE /PROGRAM DATA RETRIEVE PROGRAM

```
OPTIONS LABEL;
FILENAME INDAT1
'D:\FEDFUNDS\CFFR1997\SPROGFIL.DAT'
LRECL=410;
DATA FF97;
INFILE INDAT1;
INPUT PGID $1-6 OBJCODE $7-8 FUNTCODE
10-12 DISPCODE 14 PRGNAME $16-85
(PG1-PG13) (15.0) (PG14-PG26) (10.2);
PUT PGID OBJCODE FUNTCODE DISPCODE
```

```
PRGNAME PG1 PG14;
RUN;
```

ACKNOWLEDGMENTS

The following staff at the Economic Research Service contributed to the idea of this paper:

Richard Reeder, Economic Research Service
Faqr Bagi, Economic Research Service

REFERENCES

SAS is a registered trademark 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.

AUTHOR'S CURRENT ADDRESS

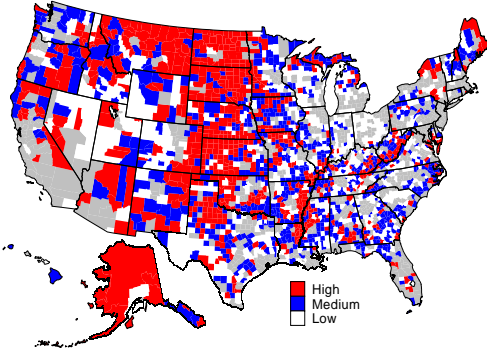
Samuel D. Calhoun, Mathematician
U.S. Department of Agriculture
Economic Research Service
1800 M Street, N.W., Rm. N2146
Washington, D.C. 20036

Phone: (202)-694-5339

Fax: (202)-694-5364

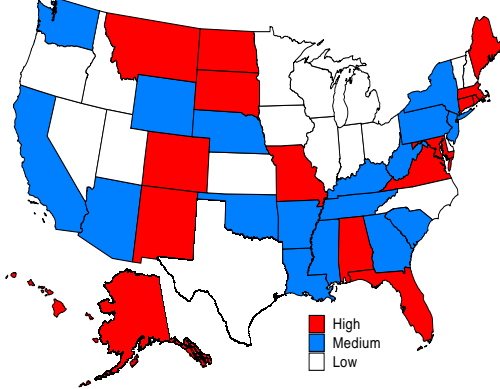
Internet: SCALHOUN@ECON.AG.GOV

Figure 1
County per capita Federal funds, fiscal year 1997



Source: Calculated by ERS using Federal funds data from the Bureau of the Census.

Figure 2
State per capita Federal funds, fiscal year 1997



Source: Calculated by ERS using Federal funds data from The Bureau of the Census.