MAP CONSTRUCTION AND GEOLOCATIONAL ANALYSES WITH SAS AND A GEOZIP FILE

Steven R. Borbash
West Virginia University

1. Introduction

1.1 Overview

A system for attaching geolocational data including coordinates, county and district names and numbers to a list of zip codes is described. This system is called the GEOZIP/SAS system because it uses the Statistical Analysis System (SAS) together with a specially constructed GEOZIP data file.

The GEOZIP/SAS system is similar in purpose to the DIME/ADMATCH system developed by the U. S. Census Bureau, except that GEOZIP/SAS is not limited to Standard Metropolitan Statistical Areas (SMSAs). The two systems are compared briefly in this paper.

The construction of a GEOZIP data file for the State of West Virginia is described. The file contains one record for each postal zip code within a state. Each record contains county and district names and numbers, the place name of the post office and the latitude and longitude of the post office.

Several GEOZIP/SAS applications are discussed and illustrated with SAS language statements. These applications begin by attaching some or all of the GEOZIP file data to either a list of zip codes or to a list of place names covered by the GEOZIP file. By assuming the location of individuals to be approximately that of their post offices, a simple list of zip codes can (after GEOZIP/SAS processing) be converted to maps and totals by counties or districts. This provides a rapid and simple method for analyzing population distributions. SAS-72 forms the basis for nearly all examples, but recoding into SAS-75 should present no problems.

1.2 GEOZIP/SAS and DIME/ADMATCH Comparisons

For Standard Metropolitan Statistical Areas (SMSAs) the U. S. Census Bureau has developed a computerized method for attaching approximate coordinates (usually of block centroids) to the street address portion of individual records. This is done with the DIME files [5] and the ADMATCH software package [4]. However, for many purposes coordinates are desired for addresses outside the SMSAs. This is especially true for studies of rural populations, such as those dealing with clinic locations, emergency medical services, etc.

The GEOZIP file concept was originated by the author to provide a fast and convenient computer method of attaching approximate coordinates and other geolocational data to individual zip codes or place names either inside or outside SMSAs. The attached coordinates are those of the U. S. Post Office identified with the zip code. The SAS (Statistical Analysis System) computer software [6] is used to attach GEOZIP file information to the zip codes and to carry out subsequent analyses of the data. The GEOZIP/SAS system can also be used to attach zip codes to unzipped place names, but other commercial systems already exist to do this, with much better accuracy within SMSAs. See for example reference [3].

Important differences between the SAS/GEOZIP and DIME/ADMATCH systems are listed below.

1. Coordinates are attached to zips in the GEOZIP/SAS system and to street addresses in the DIME/ADMATCH system.

2. Coordinates can be attached for zips inside or outside SMSA's with the GEOZIP/SAS system, but are limited to the SMSA boundaries with the DIME/ADMATCH system.

3. Coordinates of individual locations are less accurate with the GEOZIP/SAS system (where individuals are assigned the coordinates of their post office) than with the DIME/ADMATCH system (where individuals are usually assigned the coordinates of their block antroids).

The loss in coordinate locational accuracy in the GEOZIP/SAS system is often unimportant where large areas (such as a state) are considered. A major portion of the work done with the West Virginia GEOZIP file has been the assignment of individuals to counties and their political Minor Civil Divisions (MCDs) for comparison to census data for county and districts. Here again, locational accuracy is of minor importance. The loss of locational accuracy also helps protect individual privacy. As a result, obtaining a list of zip codes for a group of individuals is often easier than obtaining their street addresses.

2. The West Virginia GEOZIP File

2.1 Definitions of Variables

The SAS INPUT statement for the West Virginia GEOZIP file is shown on Figure 1. There are 1398 records in the file, and each record is on a single punch card. Special variables DUPCITY and DUPZIP (either blank or '1') appear in columns 78 and 79 respectively. These variables are included because the GEOZIP file data is attached to records from another file through a SAS MERGE either BY ZIP or BY PLACE. For attachment BY ZIP no duplicate zip records must appear in GEOZIP, and for attachment BY PLACE no duplicate place names must appear in GEOZIP. (Duplicates are allowed in the other file mentioned in the MERGE statement).

Before a MERGE by ZIP all records with DUPZIP = 1 are eliminated from the GEOZIP file, and before a MERGE by PLACE all records with DUPCITY = 1 are eliminated. Examples of how the variables DUPZIP and DUPCITY are assigned to records in the GEOZIP file are shown on Figure 2.

For cities such as Charleston, a large number of distinct zip codes are used for the same post office (typically for box numbers, special buildings, and branches served from a main post office). All of these are identified with the same PLACE name, and DUPCITY = 1 for all of these records except one which is chosen as a master record. All records with DUPCITY = 1 are eliminated before the MERGE by PLACE leaving only the one master record in GEOZIP from the group of zips with the same PLACE name. Typical records of this type are shown at the top of Figure 2.

Some main post offices service postal branches in neighboring towns. These branches are identified with distinct town names, but their zip code remains the same as the main post office. See for example the bottom of Figure 2, where the Morgantown post office services branches of Westover, Star City and Sabraton. Commonly used misspellings of PLACE names are treated as distinct branches. For attaching GEOZIP data by PLACE all these records are retained, because their PLACE names are different, but for attaching GEOZIP data by ZIP only one can be retained to avoid duplicates and this is MORGANTOWN in the examples. Deleting all records with DUPZIP = 1 eliminates all but one master record from a group with duplicate zips.

2.2 GEOZIP File Construction

The West Virginia GEOZIP file was constructed one county at a time. Each post office and branch within a county (including those discontinued) is listed by county in the Directory of Post Offices published by the U. S. Postal Service. All these post offices were located with colored marker pens on detailed county and city maps. The latitude and longitudes of the post offices were scaled directly from the maps (post office locations are marked on the detail maps). The district names (political subdivisions of counties, often called minor civil divisions or MDC's) were also obtained directly from the county maps. The district

FIGURE 1

SAS INPUT STATEMENT FOR
WEST VIRGINIA GEOZIP FILE

```
INPUT
  CNTYNAME $ 1-10
  CNTYNUM $ 12-14
  DISTNAME $ 16-30
  DISTNUM 32-34
  PLACE $ 36-50
  LNGDEG 52-53
  LNGMIN 55-59
  LATDEG 60-61
  LATMIN 63-67
  ZIP 68-72
  DUPCITY 78
  DUPZIP 79;

COMMENT

    CNTYNUM AND DISTNUM ARE THREE DIGIT US CENSUS BUREAU NUMBERS

    DISTNAME IS THE US CENSUS NAME FOR A COUNTY POLITICAL SUBDIVISION
             ALSO CALLED A MINOR CIVIL DIVISION OR MCD

    PLACE IS THE POST OFFICE NAME

DUPCITY = 1 WHEN ANOTHER RECORD HAS THE SAME PLACE NAME AS THIS
            ONE, AND THIS RECORD MUST BE ELIMINATED TO GET RID
            OF DUPLICATE PLACE NAMES IN THE FILE.  (DUPCITY IS
            MISSING OTHERWISE)

DUPZIP = 1  WHEN ANOTHER RECORD HAS THE SAME ZIP AS THIS ONE, AND
            THIS RECORD MUST BE ELIMINATED TO GET RID OF
            DUPLICATE ZIP CODES IN THE FILE.  (DUPZIP IS MISSING
            OTHERWISE);
//
//
```

FIGURE 2

EXAMPLES OF DUPLICATE NAMES AND
ZIPS IN THE GEOZIP FILE

ZIP	PLACE	DUPZIP	DUPCITY
25301	CHARLESTON		1
25305	CHARLESTON		1
25306	CHARLESTON		1
25312	CHARLESTON		1
25300	CHARLESTON		

COMMENT

TO KEEP ONLY ONE RECORD WITH PLACE='CHARLESTON' USE

--IF DUPCITY=1 THEN DELETE--

DUPCITY=1 FOR BRANCH POST OFFICES WITH THE SAME NAME,
PO BOXES, ETC.--USUALLY IN CITIES;

ZIP	PLACE	DUPZIP	DUPCITY
26505	MORGANTOWN		
26505	STAR CITY	1	
26505	WESTOVER	1	
26505	SABRATON	1	

COMMENT

TO KEEP ONLY ONE RECORD WITH ZIP=26505 USE
--IF DUPZIP=1 THEN DELETE--

DUPZIP=1 FOR BRANCH POST OFFICES WITH DISTINCT NAMES AND FOR
'ACCEPTABLE' MISSPELLINGS OF PLACE NAMES;

//
//

and county in which the Post Office was physically located was attached to each GEOZIP record, as well as the three-digit code numbers for counties and MCD's which were obtained from U. S. Census tape summaries for West Virginia. All data was entered on work sheets for each West Virginia zip and then punched onto cards, one county at a time.

2.3 Checking the File

After the file was keypunched, it was sorted 3 ways, printed, and scanned for errors.

(1) BY CNTYNAME DISTNAME for comparison with U. S. Census names and numbers of counties and MCD's.

(2) BY CNTYNAME PLACE for comparison with the U. S. Post Office Directory listing by counties.

(3) BY PLACE for comparison with the U. S. Post Office directory listing by post office name.

Identification and correction of all obvious omissions, deletions and misspellings was followed by a mapping of all counties using PROC PLOT which identified the post offices separately by using the first letter of PLACE as a printer symbol. See Figure 3 for the SAS program statements. The general placement pattern of points on the SAS county scatter plots was compared with the post office locations marked with colored ink on the county maps. Because of the automatic scaling features of PROC PLOT one or more points with a highly deviant LAT or LNG would cause the plot to be distorted, with the valid points packed into a small area of the plot. After these errors in LAT and LNG were found and corrected, another county scatter plot was run to identify LAT and LNG values in error by smaller amounts. Figure 4 is a final SAS scatter plot of the entire WV GEOZIP file, showing all postal locations in West Virginia.

At this point, the GEOZIP file was merged with a large test file of West Virginia place names. This test file has 18,000 place names and was obtained from an agency of State government. It represented a sample from an even larger file used for correspondence with State residents.

All place names in this test file which did not match a place name from the GEOZIP file were printed and considered individually. (See Figure 5 for the SAS program statements). Place names in the test file, but without a post office were included in the GEOZIP file along with the zip code of the nearest post office. The variable DUPZIP was set to 1 for these new entries. The remaining non-matching place names were the result of spellings which were in disagreement with those in GEOZIP. When the common-usage spelling was judged reasonable, it was included in GEOZIP with the same record parameters as the preferred spelling record except for the variables PLACE and DUPZIP (=1). For example there are more than 20 records for White Sulphur Springs, each representing a different common usage abbreviation of this West Virginia place name.

FIGURE 3

CHECKING THE GEOZIP FILE
WITH PROC PLOT

```
DATA GEOZIP;
  INPUT DDNAME=X1
     CNTYNAME $ 1-10
        .
        .
        .
     DUPZIP 79;
  LAT=LATDEG+LATMIN/60;
  LNG=-(LNGDEG+LNGMIN/60);

PROC SORT;
  BY CNTYNAME;

COMMENT
    EACH COUNTY WILL BE PLOTTED SEPARATELY;

PROC PLOT;
  BY CNTYNAME;
  VAR LAT LNG;
  ID PLACE;


COMMENT
    PLOT THE WHOLE STATE;

PROC PLOT ROWS=70;
  VAR LAT LNG;

//
//
```

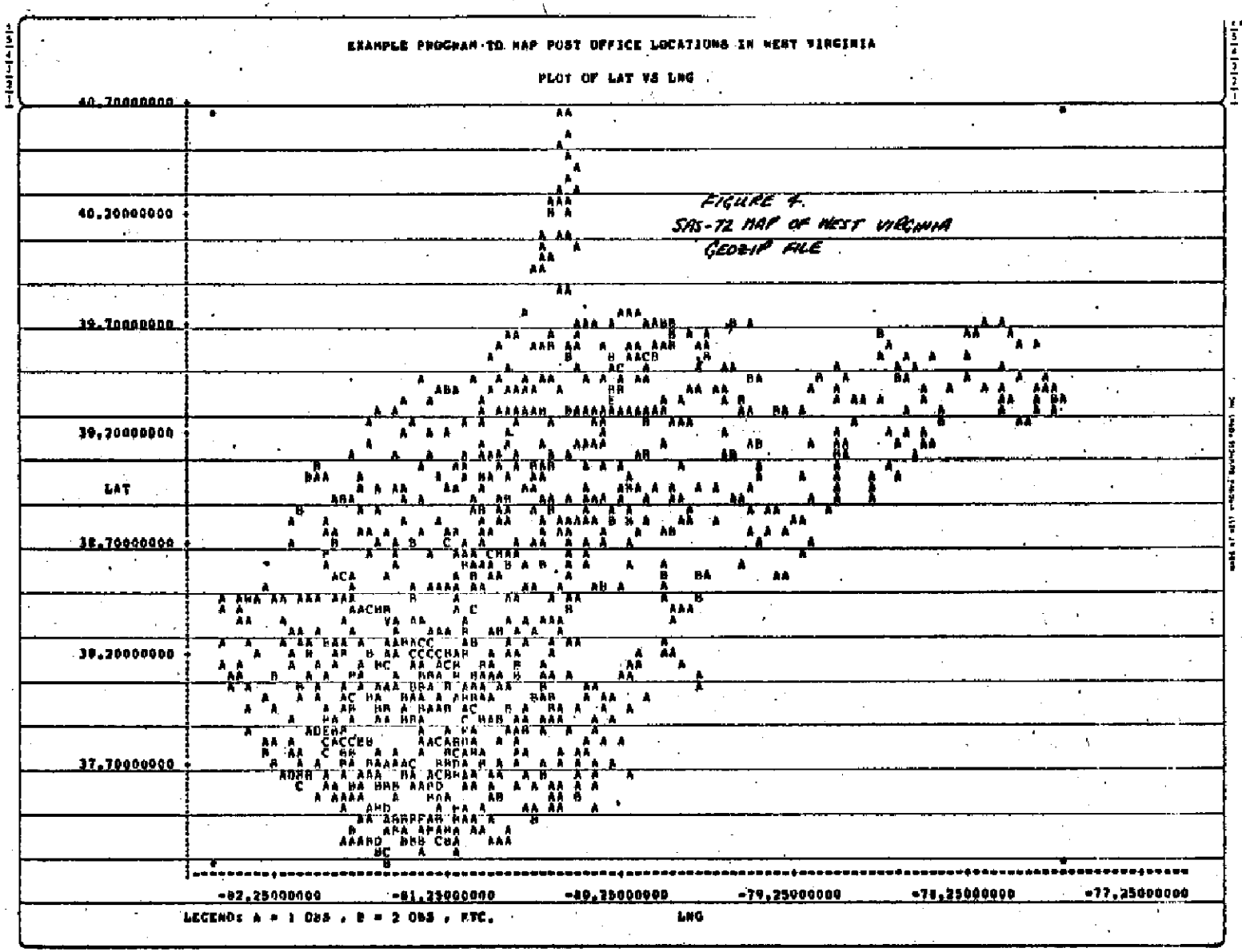


FIGURE 4.
SAS-72 MAP OF WEST VIRGINIA
GEOBIP FILE

The completed WV GEOZIP file has 1398 records. When those with DUPZIP = 1 are eliminated, there are 1180 records remaining. When those with DUPCITY = 1 are eliminated from the complete file, there are 1357 records remaining.

3. Uses of the GEOZIP File

The GEOZIP file is easily used with SAS. Examples showing typical applications are listed below and discussed individually in the text which follows.

1. Attaching GEOZIP data to records from other files by PLACE and by ZIP.
2. Aggregating attached records by county and district.
3. Printer maps using SAS.
4. Printer maps using SAS/GRIDS.
5. Locating all records in a file within a given radial distance of a reference point.

3.1 Attaching GEOZIP Data to Records by PLACE and ZIP

Before doing any sort of locational analysis on records containing West Virginia zip codes or place names, it is necessary to attach GEOZIP data to these records. Figure 5 shows how GEOZIP data is attached to a list of PLACE names. The GEOZIP file is read in, records with DUPCITY = 1 are eliminated with a program statement and then the file is sorted BY PLACE. Next the file PLACELIST which is to have GEOZIP data attached is read in and sorted by PLACE. A convenience variable KOUNT is added here for use in aggregating records by PROC MEANS later.

The GEOZIP data attachment is done with the statements MERGE GEOZIP ZIPLIST; BY PLACE; to create the final data set called COMBINED. It is necessary that GEOZIP (the left member of the pair of data sets being merged) have no duplicate PLACE records. This is guaranteed by deleting records with DUPCITY = 1. The file PLACELIST (right member of the merged pair) may have duplicate PLACE records. Only those records which have matching values of PLACE (MERGE = 3) will appear in the COMBINED file. Records from PLACELIST which have no matching GEOZIP record (MERGE = 2) indicate an error, and are printed with the PUT statement. When MERGE = 1, an unused PLACE name is indicated. Records with MERGE = 1 and MERGE = 2 are deleted. The attachment of GEOZIP data by zip is very similar and is illustrated in Figure 6.

3.2 Aggregating Records with Attached GEOZIP Data

Figure 7 shows a SAS program segment which uses PROC FREQ and PROC MEANS to aggregate records by districts, counties and for the entire state in a file (COMBINED) with attached GEOZIP data.

FIGURE 5

ATTACHING GEOZIP DATA TO A LIST
OF WV PLACE NAMES

```
COMMENT
    ATTACHMENT OF WV GEOZIP DATA TO A LIST OF WV PLACE NAMES;

DATA GEOZIP;
  INPUT DDNAME=X1
    CNTYNAME $ 1-10
        .
        .
        .
    DUPZIP 79;
  IF DUPCITY=1 THEN DELETE;
PROC SORT;
   BY PLACE;

DATA PLACELIST;
   INPUT DDNAME=X2
      PLACE $ 1-15;
    KOUNT=1;

PROC SORT;
  BY PLACE;

DATA COMBINED;
   MERGE GEOZIP PLACELIST;
       BY PLACE:

  IF MERGE=1 THEN DELETE;
  IF MERGE=2 THEN PUT 'NO MATCH FOR WV PLACE' PLACE;
  IF MERGE=2 THEN DELETE;
  IF MERGE=3 THEN OUTPUT;

COMMENT
    MERGE=1 FOR A RECORD FROM GEOZIP WITH NO MATCHING RECORD FROM
            PLACELIST.  THIS MEANS NO ONE IN PLACELIST USED THIS
            PLACE.
    MERGE=2 FOR A RECORD FROM PLACELIST WITH NO MATCH IN GEOZIP.
            THIS IS AN ERROR.  IF GEOZIP HAS ALL VALID PLACES
            THEN THE ONE IN PLACELIST IS INVALID.
    MERGE=3 FOR A MATCHING RECORD FROM BOTH GEOZIP AND PLACELIST.
            THE GEOZIP DATA IS EFFECTIVELY ATTACHED.
//
//
```

FIGURE 6

ATTACHING GEOZIP DATA TO A LIST OF
WV ZIP CODES

```
COMMENT
    ATTACHMENT OF WV GEOZIP DATA TO A LIST OF WV ZIP CODES;

DATA GEOZIP;
  INPUT DDNAME=X1
    CNTYNAME $ 1-10
        .
        .
        .
     DUPCITY 78;
  IF DUPZIP=1 THEN DELETE;
    KOUNT=1;

PROC SORT;
   BY ZIP;

DATA ZIPLIST;
  INPUT DDNAME=X2
     ZIP 11-14;

PROC SORT;
  BY ZIP;

DATA COMBINED;
   MERGE GEOZIP ZIPLIST;
       BY ZIP;

  IF MERGE=1 THEN DELETE;
  IF MERGE=2 THEN PUT 'NO MATCH FOR WV ZIP' ZIP;
  IF MERGE=2 THEN DELETE;
  IF MERGE=3 THEN OUTPUT;

COMMENT
    MERGE=1 FOR A RECORD FROM GEOZIP WITH NO MATCHING RECORD FROM
            ZIPLIST.  THIS MEANS NO ONE IN ZIPLIST USED THIS
            ZIP.
    MERGE=2 FOR A RECORD FROM ZIPLIST WITH NO MATCH IN GEOZIP.
            THIS IS AN ERROR.  IF GEOZIP HAS ALL VALID ZIPS
            THEN THE ONE IN ZIPLIST IS INVALID.
    MERGE=3 FOR A MATCHING RECORD FROM BOTH GEOZIP AND ZIPLIST.
            THE GEOZIP DATA IS EFFECTIVELY ATTACHED.
//
//
```

FIGURE 7

AGGREGATION OF ATTACHED RECORDS
BY COUNTY AND DISTRICT

```
PROC FREQ DATA=COMBINED;
  TABLES CNTYNAME CNTYNAME*DISTNAME;
  TITLE 'COUNTY AND DISTRICT TOTALS';

COMMENT
    SORT BEFORE PROC MEANS;

PROC SORT DATA=COMBINED;
     BY CNTYNAME DISTNAME;

COMMENT
   GET RECORD COUNT FOR DISTRICTS;

PROC MEANS NOPRINT SUM OUT=X3;
  BY CNTYNAME DISTNAME;
  VAR KOUNT;
PROC PRINT;
   VAR CNTYNAME DISTNAME KOUNT;
   TITLE 'DISTRICT TOTALS';

COMMENT
    GET COUNTY TOTALS;

PROC MEANS NOPRINT SUM OUT=X4;
     BY CNTYNAME;
     VAR KOUNT;
PROC PRINT;
   VAR CNTYNAME KOUNT;
   TITLE 'COUNTY TOTALS';

COMMENT
    NOW GET STATE TOTAL COUNT;

PROC MEANS NOPRINT SUM OUT=X5;
     VAR KOUNT;
DATA X6;
SET X5;
  TOTALS='WHOLE STATE';
PROC PRINT;
  VAR TOTALS KOUNT;
  TITLE 'STATE TOTALS';

//
//
```

PROC FREQ generates tables with county and district totals without a preceding PROC SORT, but gives no output data sets. PROC MEANS requires PROC SORT beforehand but will create resultant summary data sets. PROC MEANS operates on the convenience variable KOUNT = 1 introduced in Figures 5 and 6.

3.3 Printer Maps Using SAS

Figure 8 shows an outline map of the State of West Virginia. The state can be enclosed in a rectangle which is 236.32 miles high and 264.62 miles wide. The ratio of N-S boundaries to E-W boundaries is 0.893. The latitudes and longitudes of the four corners are as shown, giving approximate scale factors of 68.98 miles/degree latitude (vertically) and 54.03 miles/degree longitude (horizontally). Latitude increases from bottom to top and longitude increases from right to left.

Figure 9 shows programs (for SAS-72 and SAS-75) which use a file (COMBINED) which has latitude and longitude attached to each record from the GEOZIP file. Latitude and longitude are converted to degrees and fractions of degrees. Longitude is multiplied by (-1) because PROC PLOT assumes increasing values of the horizontal variable from left to right, which is the opposite of the longitude variable. This multiplication is necessary to avoid inverted maps.

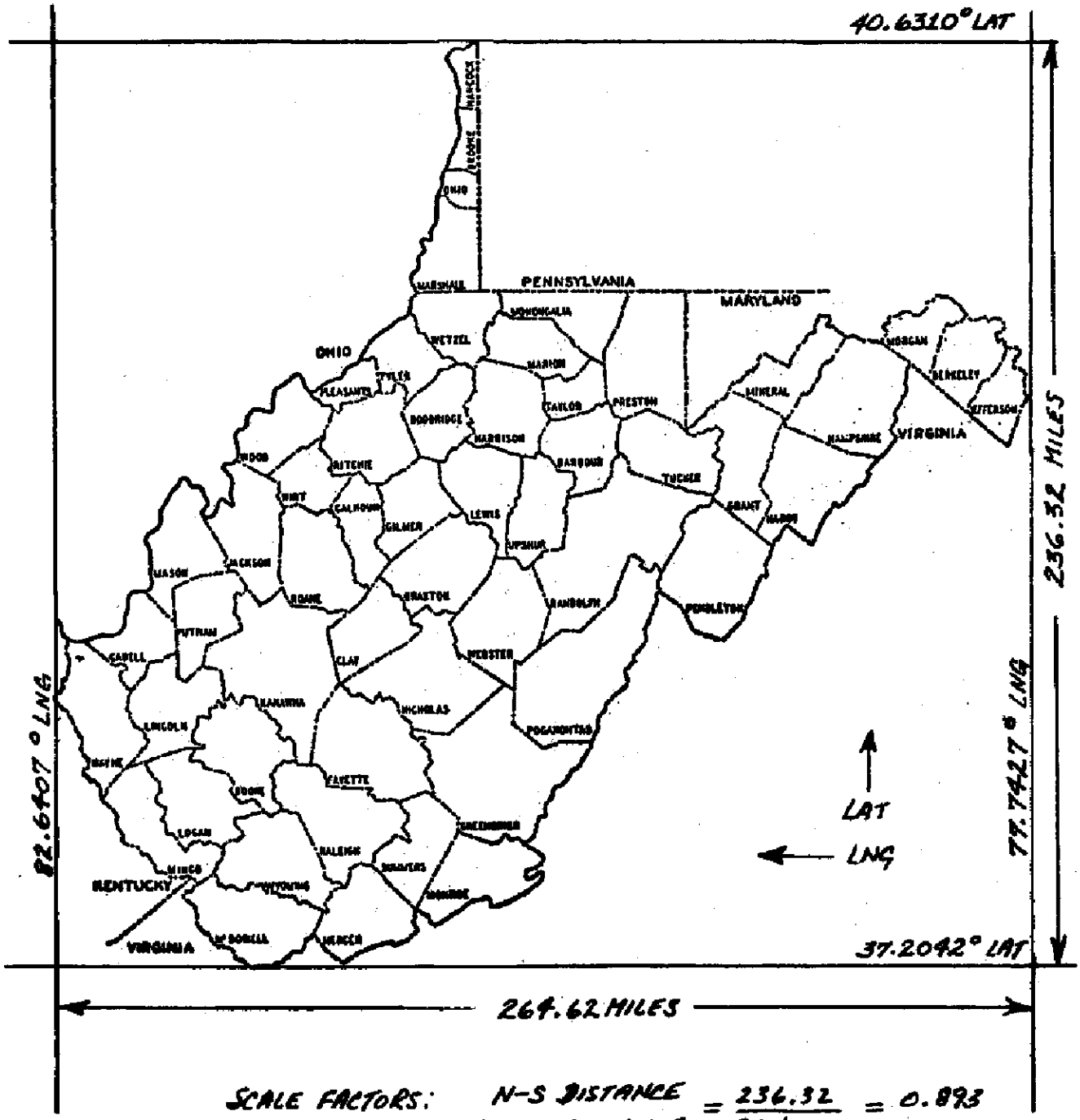
Then PROC PLOT is used (SAS-72) with LAT as the vertical variable and LNG as the horizontal variable. The ROWS option of the PROC PLOT statement is used to adjust the vertical dimension of the map (the horizontal dimension is not under programmer control in SAS-72) so the ratio of the vertical to horizontal dimension of the final plot is as close as possible to 0.893. For West Virginia, ROWS = 70 gives a very acceptable map. Figure 4 was generated in this manner with PROC PLOT. The four corners of the map boundary are read in as a card data set and added to the main file with a SET statement. This allows an easy check on map dimensions, and provides reference points to align a plastic overlay with a map outline if desired, before xeroxing.

For SAS-75 PROC SCATTER is used instead of PROC PLOT. For PROC SCATTER, the horizontal and vertical scale factors are both under programmer control. A suggested procedure is to first adjust the horizontal map dimension, and then hold this constant while the vertical dimension is adjusted.

3.4 Printer Maps Using SAS/GRIDS

In this two-step application, SAS is used in the first job step to condition a file with attached GEOZIP data. This conditioned file is passed to a second job step where a special map program called GRIDS generates one or more printer maps. The GRIDS processor and documentation [2] is available from the U. S. Census Bureau for a modest fee. GRIDS creates shaded maps and allows the aggregation cell sizes to be adjusted (as opposed to SAS, where the cell sizes are always one printer symbol in area). Whereas SAS customarily uses the convention A,B,...etc. for cell print symbols corresponding to a cell count of 1,2,...etc.,

FIGURE 8
WEST VIRGINIA
MAP DIMENSIONS AND SCALE FACTORS



SCALE FACTORS: $\frac{\text{N-S DISTANCE}}{\text{E-W DISTANCE}} = \frac{236.32}{264.62} = 0.893$

$$\frac{\text{MILES}}{^\circ\text{LAT}} = \frac{236.32}{(40.6310 - 37.2042)} = \frac{236.32}{3.4260} = 68.98$$

$$\frac{\text{MILES}}{^\circ\text{LNG}} = \frac{264.62}{(82.6407 - 77.7427)} = \frac{264.62}{4.898} = 54.03$$

SRB/1/21/96

FIGURE 9

PRINTER MAPS USING SAS-72

```
DATA GEOZIP;
   SET COMBINED;
        LAT=LATDEG+LATMIN/60;
        LNG=(LNGDEG+LNGMIN/60);
   KEEP LAT LNG;

DATA CORNERS;
  INPUT LAT 1-10
        LNG 11-20;
     LNG=-LNG;
CARDS;

    (4 CORNER CARDS HERE)

DATA MAP;
  SET COMBINED;
  SET CORNERS;

PROC PLOT ROWS=70;
  VAR LAT LNG;
  ID PLACE;
```

PRINTER MAPS USING SAS-75

```
DATA GEOZIP;
  SET COMBINED;
     LAT=LATDEG+LATMIN/60;
     LNG=-(LNGDEG+LNGMIN /60);
    KEEP LAT LNG;
DATA CORNERS;
  INPUT LAT 1-10
        LNG 11-20;
    LNG=-LNG;
CARDS;

    (4 CORNER CARDS HERE)

DATA MAP;
  IF ENDA THEN GO TO B;
  SET COMBINED END=ENDA;
  RETURN;
B:SET CORNERS;

PROC SCATTER;
 PLOT LNG*LAT/VMIN=37  VMAX=41  VINC=0.5  VPOS=64
              HMIN=-83 HMAX=-77 HINC=1    HPOS=96;
//
//
```

GRIDS overprints selected characters to achieve a shaded cell effect. A GRIDS map is shown in Figure 11. The SAS-72 and GRIDS instructions to produce the map are shown in Figure 10.

The SAS job steps read in a file (DDNAME = X1) which is assumed to have GEOZIP data already attached to each of the records. LAT and LNG are expressed in degrees and fractions of degrees, but multiplication of LNG by (-1) is not required in this application, because GRIDS adjusts for this. The single-character variable TAG (=blank) and the numeric variable KOUNT (=1) are added to all input records. KOUNT will be used by GRIDS as the aggregation variable for each cell, and TAG is an end-of-file indicator, which is blank for all records except a last dummy record where it assumes the value 'L'. To attach the last dummy record to the data set, a one-record data set called LASTOBS is created by reading one card with TAG = 'L'. This is placed at the end of the data set MAP. The SAS step terminates by taking the data set MAP out of SAS format (SASFMT) and putting the data records out on a disk in fixed block format (FBFMT). This is done with PROC OUTPUT, a SAS procedure written by Mr. Daniel Chilko of the West Virginia University Computer Center [1]. Its function is similar to the SAS-75 PUT statement.

The second job step executes the GRIDS procedure, which uses the SAS output file as input. The GRIDS statement XINCR ='LEFT' tells the processor that longitude increases from right to left. A detailed explanation of the other GRIDS statements can be found in reference [2].

3.5 Locating All Records in a File Within a Given Radial Distance of a Reference Point

Figure 12 shows a portion of a SAS program which identifies and prints all records in the input file (specified with DDNAME = X1) whose radial distance from a reference point is less then R_LIMIT = 50 miles. The differences between the reference latitude and longitude and the latitudes and longitudes specified for each individual record are converted to miles with conversion factors Y_CONST and X_CONST. Finally, these X and Y distances are converted to a radial distance by taking the square root of the sum of the squares. This program is useful for studying service area problems and/or facilities location problems.

4. Conclusions

The West Virginia GEOZIP file used with SAS has been a fast and useful tool for summarizing data about population locations within the State. The approximation errors of assuming individual residences at post office locations has been of little concern when analyses are made over the relatively large statewide area.

The GEOZIP file in conjunction with SAS has been used extensively for nearly two years by health planners and data analysts with the West Virginia Regional Medical Program who are concerned primarily with problems related to statewide distributions of health care providers and recipients.

FIGURE 10

PRINTER MAPS USING SAS/GRIDS

```
//*
//S1 EXEC SAS
//*
//SAS.X1 DD DSN=OLDJOB,UNIT=2314, VOL=SER=1EDISK,
//       DISP=(OLD,KEEP,DELETE),DCB=(RECFM=FB,LRECL=100,BLKSIZE=2000)
//* RECORDS WITH ATTACHED GEOZIP DATA
//*
//SAS.X2 DD DSN=PASSDATA,UNIT=2314,VOL=SER=IEDISK,
//      DISP=(NEW,PASS),DCB=(RECEM=FB,LRECD=25,BLKSIZE=2500),
//      SPACE=(TRK,(5,1),RLSE)
//* DATA SET TO BE PASSED TO GRIDS WITH 2000 RECORDS
//*      CREATED BY PROC OUTPUT;
//*
//SAS.SYSIN DD *
 DATA GEOMAP;
   INPUT DDNAME=X1
    LNGDEG 52-53
    LNGMIN 55-59
    LATDEG 60-61
    LATMIN 63-67;
   LAT=LATDEG+LATMIN/60;
   LNG=LNGDEG+LNGMIN/60;
   TAG=' ';
   KOUNT=1;
   DROP LATDEG LATMIN LNGDEG LNGMIN;
COMMENT
    CREATE A LAST DUMMY RECORD WITH TAG='L';
 DATA LASTOBS;
   INPUT TAG $ 5;
CARDS;
    L
;
COMMENT
     PUT THIS LAST RECORD AT THE END OF THE DATA SET;
DATA MAP;
   SET GEOMAP;
   SET LASTOBS;
COMMENT
    TAKE DATA SET OUT OF SASFMT AND INTO FBFMT TO PASS TO GRIDS;

PROC OUTPUT;
 PARMCARDS;
X2(25)/LNG(1-10 5) LAT(11-20 5) KOUNT(21-23 1) TAG($ 25)
        ;
/*
```

FIGURE 10 (continued)

PRINTER MAPS USING SAS/GRIDS

```
//S2 EXEC GRIDS
//*
//GRIDS.FT10F001 DD DSN=*.S1.SAS.X2,DISP=(OLD,DELETE,KEEP)
//* DATA PASSED FROM SAS
//GRIDS.SYSIN DD *
CNTRL='FMT'
DATAUNIT=10
TX=(1,10)
TY=(11,10)
TO=(21,3)
EOF=(25,1)
VEOF='L'
CNTRL='DD'
CNTRL='MAP'
MAPIDENT=1
MAPTYPE='SHADED'
MIN=(77.7427,37.2042)
MAX=(82.6417,40.6310)
CELLWDTH=8
CELLHGHT=8
PRINTER=8
INCH=(.5384,.4129)
XINCR='LEFT'
YINCR='TOP'
SHADES=(0,2,4,6,7,11,16,18,20)
RANGES=(0.5,25.1,50.1,75.1,100.1,150.1,200.1,150.1)
MAPTITLE='YES'
WEST VIRGINIA PHYSICIANS COUNT SHADED MAP
=
/*
//
```

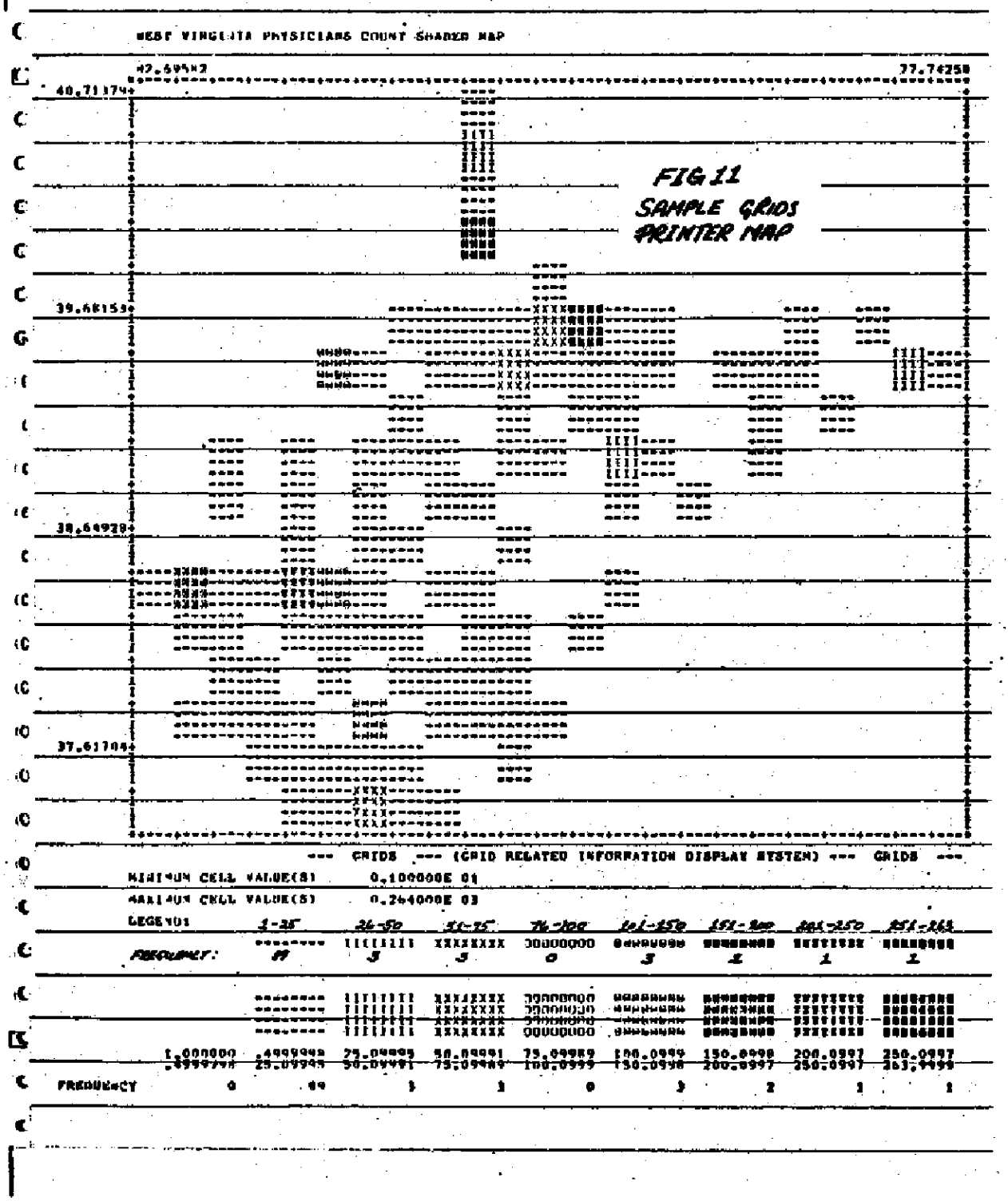

WEST VIRGINIA PHYSICIANS COUNT SHADED MAP
82.69582
77.74258
40.71174
39.66159
38.64929
37.61704
FIG 11
SAMPLE GRIDS
PRINTER MAP
--- GRIDS --- (GRID RELATED INFORMATION DISPLAY SYSTEM) --- GRIDS ---
MINIMUM CELL VALUE(S) 0.100000E 01
MAXIMUM CELL VALUE(S) 0.244000E 03
LEGEND 1-25 26-50 51-75 76-100 101-150 151-200 201-250 251-264
FREQUENCY: 19 3 3 0 3 2 1 1
1.000000 .4999998 25.00000 50.00001 75.00000 100.0000 150.0000 200.0000 250.0000
.9999998 25.00000 50.00001 75.00000 100.0000 150.0000 200.0000 250.0000 263.9999
FREQUENCY 0 19 3 3 0 3 2 1 1

FIGURE 12

CREATING A FILE SUBSET WHOSE
RECORDS ARE ALL WITHIN A RADIAL DISTANCE OF A
REFERENCE POINT

```
COMMENT
    READ IN FILE WITH ATTACHED GEOZIP DATA;

DATA GEOMAP;
  INPUT DDNAME=X1
        NAME $ 1-20
        LNGDEG 52-53
        LNGMIN 55-59
        LATDEG 60-61
        LATMIN 63-67;-
   LAT=LATDEG+LATMIN/60;
   LNG=LNGDEG+LNGMIN/60;

   LATREF=38.5;
   LNGREG=81.5;
   R_LIMIT=75.0;
   Y_CONST=68.98;
   X_CONST=54.03;

COMMENT
    LATREF AND LNGREF DEFINE A REFERENCE POINT.
    Y_CONST AND X_CONST CONVERT ONE DEGREE OF LATITUDE
    AND LONGITUDE TO MILES,RESPECTIVELY.
    R_DIST IS THE RADIAL DISTANCE IN MILES BETWEEN THE
    REFERENCE POINT AND EACH RECORD IN THE FILE.  ONLY
    THOSE RECORDS WITH R_DIST < R_LIMIT WILL BE OUTPUT TO THE DATA
    SET GEOMAP;

   Y_DIST=ABS(LATREF-LAT)*Y_CONST;
   X_DIST=ABS(LNGREF-LNG)*X_CONST;

   R_DIST=SQRT(Y_DIST**2+X_DIST**2);
   IF R_DIST < R_LIMIT THEN OUTPUT;
DROP R_LIMIT Y_CONST X_CONST Y_DIST X_DIST;

PROC SORT;
  BY R_DIST;
PROC PRINT;
   VAR R_DIST NAME;
//
//
```

One major advantage of the GEOZIP file is that it requires only a list of zip codes or a list of place names to produce maps and distribution tables by county, etc. A simple list of zip codes is often obtainable quickly and without the necessity of elaborate individual privacy guarantees.

Funds to construct the West Virginia GEOZIP file and to develop SAS applications programs were provided by a Grant from the West Virginia Regional Medical Program to the West Virginia University Industrial Engineering Department.

References

[1] Chilko, Daniel M. "SAS Output Procedure", West Virginia University Computer Center Academic Services Instruction 42.24, (Sept.,1973), Morgantown, WV 26506

[2] U. S. Bureau of the Census, Census Use Study: GRIDS, A Computer Mapping System, (April, 1972), Washington, DC 20233

[3] List Processing Co., ZAP: Zip Code Attachment Program, (1974), List Processing Co., 683 Larch Avenue, Elmhurst, Illinois 60126

[4] U. S. Bureau of the Census, Census Use Study OS ADMATCH: An Address Matching System, (1972), Washington, DC

[5] U. S. Bureau of the Census, Census Use Study: The DIME Geocoding System, Report No. 4, (1970), Washington, DC

[6] SAS Project Group, SAS - A Users Guide to the Statistical Analysis System, (Aug., 1972), Institute of Statistics, P. O. Box 5457, Raleigh, NC 27607