Creating a Distance Matrix from Census Data with SAS[1]

Many summary census files have information about the longitude and latitude of the center geographic units. These can be used to calculate the distance between the centers geographic units.

Following is a SAS program that uses data from ICPSR study 6054 (Census of Population and Housing, 1990 [United States]: Summary Tape File 3C) to calculate the distance between the centers of states. The data set used has 51 cases (50 states plus DC), and this N of Cases is reflected in the array statements that create variables (for example: `array dlon dlon1 - dlon51 ;`). If a different level of geography is used (for example, counties) then the number of elements in each array must be modified to reflect the number of cases.

```
** distance_matrix ;

* get longitude and latitude in decimal degrees - from study 6054;
title1 'distance_matrix' ;
LIBNAME LIBRARY 'C:\all\help\helpnew\distance\SASLIB' ;
options linesize = 80 pagesize = 70 ;
data longlat ;
infile 'C:\all\help\helpnew\distance\state.data' ;
MERGVAR=1 ;
** varvar is created here just to show how a variable of interest ;
** would be used.  This variable could be %urban, for example ;
varvar = ranuni(333) ;
label varvar = 'A variable of interest' ;
input stab $ statefip INTPTLAT INTPTLNG ; run ;

PROC TRANSPOSE DATA= longlat OUT=stabdat PREFIX= st ; VAR stab      ; run ;
PROC TRANSPOSE DATA= longlat OUT=Y        PREFIX=  Y ; VAR INTPTLNG ; run ;
PROC TRANSPOSE DATA= longlat OUT=X        PREFIX=  X ; VAR INTPTLAT ; run ;
PROC TRANSPOSE DATA= longlat OUT=v        PREFIX=  b ; VAR varvar   ; run ;

DATA xyv ; MERGE stabdat Y X V;
MERGVAR=1 ;

data LIBRARY.DISTANCE_MATRIX ; merge longlat xyv ; by MERGVAR ;

lngcur = (INTPTLNG * 3.1415926535 ) / 180 ;
latcur = (INTPTLAT * 3.1415926535 ) / 180 ;

** lat of other states ;
ARRAY X X1-X51;
** long of other states ;
ARRAY Y Y1-Y51;
** array for longitude (empty now) ;
ARRAY LNG LNG1-LNG51;
** array for latitude (empty now) ;
ARRAY LAT LAT1-LAT51;
** array for differences between longitudes (empty now) ;
array dlon dlon1 - dlon51 ;
** array for differences between latitudes (empty now) ;
array dlat dlat1 - dlat51 ;
** array for distance (empty now) ;
ARRAY DIS DIS1-DIS51;
```

```
* http://www.census.gov/cgi-bin/geo/gisfaq?Q5.1
* Haversine Formula (from R.W. Sinnott, "Virtues of the Haversine",
* Sky and Telescope, vol. 68, no. 2, 1984, p. 159): ;

do over DIS ;
LNG = (Y * 3.1415926535 ) / 180 ;
LAT = (X * 3.1415926535 ) / 180 ;
dlon = lngcur - lng ;
dlat = latcur - lat ;
a = (sin(dlat /2))**2
        + cos(LAT)
        * cos(latcur)
        * (sin(dlon/2))**2 ;
c = 2 * arsin(min(1,sqrt(a))) ;
DIS = 3956*c ;
END ;
drop a c y1-y51 x1-x51 lng1-lng51 LAT1-LAT51 dlon1-dlon51 dlat1-dlat51
mergvar _name_ _label_ INTPTLAT INTPTLNG lngcur latcur varvar B1-B51 ;

proc print uniform ; id stab ; var  st1  dis1 ; format dis1-dis51 6.0 ; RUN ;
proc print uniform ; id stab ; var  st2  dis2 ; format dis1-dis51 6.0 ; RUN ;
proc print uniform ; id stab ; var  st3  dis3 ; format dis1-dis51 6.0 ; RUN ;
proc print uniform ; id stab ; var  st4  dis4 ; format dis1-dis51 6.0 ; RUN ;
proc print uniform ; id stab ; var st48 dis48 ; format dis1-dis51 6.0 ; RUN ;
proc print uniform ; id stab ; var st51 dis51 ; format dis1-dis51 6.0 ; RUN ;
```

The program and data for testing can be downloaded from:

http://staff.washington.edu/glynn/state.data

http://staff.washington.edu/glynn/distance_matrix.sas

On the next page you will find output from the above program (put into excel for easier viewing) so that distances can be checked.

stab	Dis1 to AL	Dis2 to AK	Dis3 to AZ	Dis4 to AR	...	Dis48 to WA	...	Dis51 to WY
AL	0	3,459	1,485	347		2,027		1,293
AK	3,459	0	2,656	3,184		1,695		2,297
AZ	1,485	2,656	0	1,147		962		655
AR	347	3,184	1,147	0		1,694		961
CA	1,883	2,325	442	1,538		685		790
CO	1,143	2,554	490	797		932		272
CT	984	3,333	2,197	1,163		2,353		1,727
DE	760	3,412	2,056	973		2,309		1,642
DC	685	3,380	1,976	891		2,245		1,570
FL	389	3,840	1,811	720		2,411		1,678
GA	192	3,564	1,674	530		2,184		1,453
HI	4,317	3,052	2,836	3,983		2,696		3,235
ID	1,673	2,025	672	1,335		367		384
IL	524	2,953	1,324	404		1,625		921
IN	480	3,066	1,491	484		1,787		1,090
IA	734	2,725	1,151	506		1,376		681
KS	747	2,806	837	409		1,284		552
KY	335	3,221	1,531	437		1,905		1,191
LA	333	3,416	1,215	256		1,877		1,154
ME	1,287	3,199	2,376	1,425		2,400		1,844
MD	708	3,387	1,999	916		2,263		1,591
MA	1,064	3,356	2,276	1,247		2,414		1,798
MI	823	2,801	1,587	775		1,681		1,062
MN	1,019	2,458	1,265	814		1,252		669
MS	163	3,384	1,327	212		1,906		1,173
MO	490	2,976	1,143	238		1,549		820
MT	1,541	1,986	891	1,228		537		319
NE	926	2,584	840	609		1,105		378
NV	1,745	2,217	444	1,398		528		568
NH	1,107	3,272	2,265	1,266		2,364		1,766
NJ	840	3,366	2,096	1,034		2,308		1,657
NM	1,104	2,822	381	768		1,154		577
NY	918	3,187	2,049	1,054		2,185		1,565
NC	468	3,519	1,879	743		2,267		1,557
ND	1,241	2,218	1,096	977		949		447
OH	573	3,119	1,671	654		1,921		1,244
OK	638	2,994	857	293		1,438		716
OR	1,961	1,882	778	1,619		207		682
PA	742	3,240	1,932	895		2,143		1,488
RI	1,029	3,374	2,256	1,218		2,411		1,787
SC	345	3,555	1,798	651		2,243		1,521
SD	1,069	2,405	950	779		1,011		354
TN	208	3,289	1,472	340		1,916		1,190
TX	747	3,206	800	476		1,575		910
UT	1,452	2,386	334	1,105		703		351
VT	1,093	3,196	2,207	1,232		2,289		1,699
VA	565	3,414	1,900	791		2,220		1,528
WA	2,027	1,695	962	1,694		0		734
WV	525	3,284	1,776	695		2,077		1,388
WI	817	2,701	1,397	686		1,502		868
WY	1,293	2,297	655	961		734		0