

Beginnings in Spatial Analysis and GIS in SAS

Doug Dover

Public Health Surveillance and Environmental Health Branch
Public Health Division

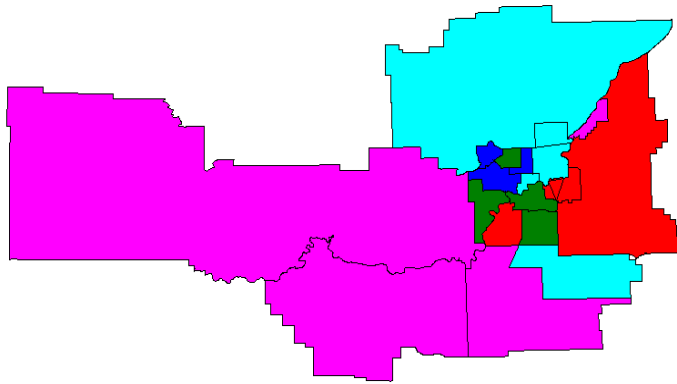
What I wanted to do...

- Spatial Smoothing of Rates by Small Regions
 - Implementation of the BYM model

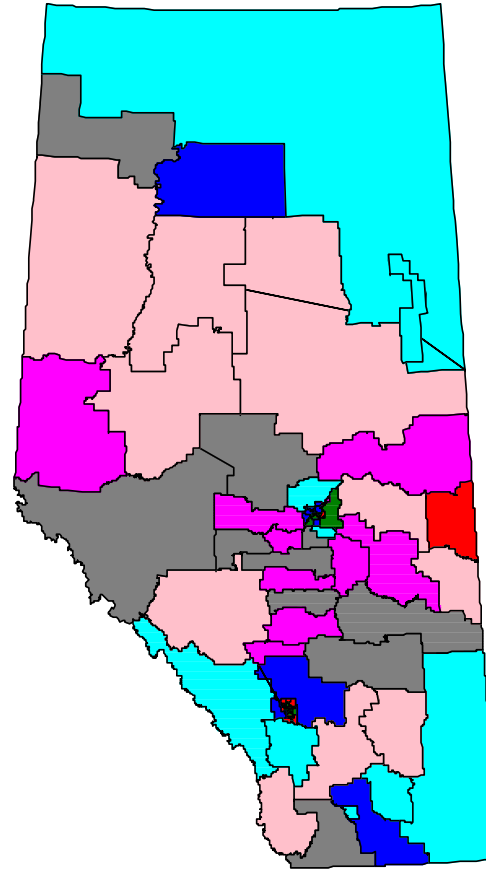
- Map Smoothed Rates

Small Regions

Edmonton subRHAs



all subRHAs



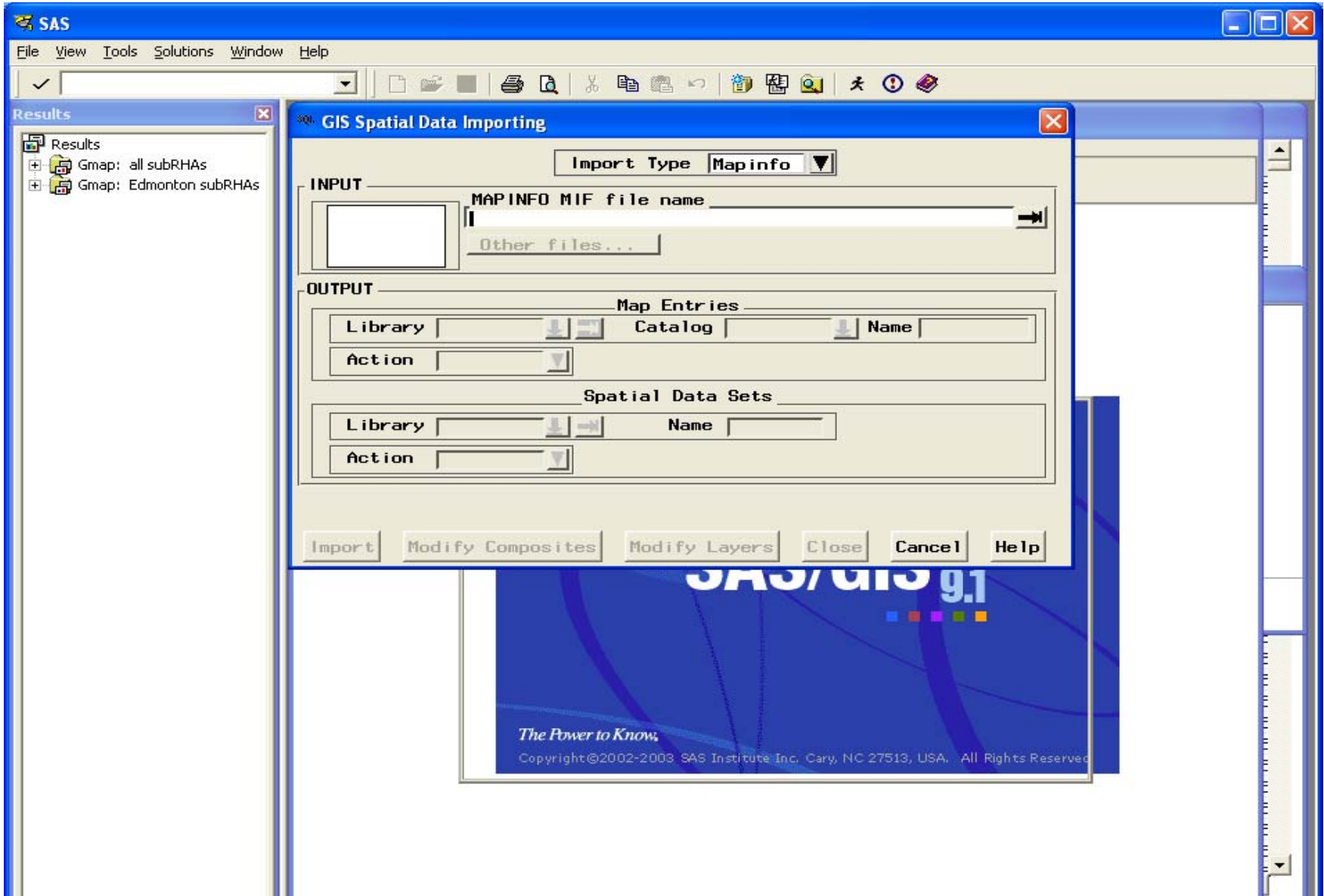
Map Boundary Files

- In a GIS format (MapInfo, ESRI)
- So it should be easy to import

- SAS has that GIS wizard solution

- GIS Spatial Data Importing...

Map Boundary Files



The image shows a screenshot of the SAS GIS Spatial Data Importing dialog box. The dialog is titled "GIS Spatial Data Importing" and has a blue header bar. The main area is divided into several sections:

- Import Type:** A dropdown menu set to "Mapinfo".
- INPUT:** A section with a "MAPINFO MIF file name" text box and a file selection button labeled "Other files...".
- OUTPUT:** A section with two sub-sections:
 - Map Entries:** Fields for "Library", "Catalog", and "Name", each with a download icon. An "Action" dropdown menu is below.
 - Spatial Data Sets:** Fields for "Library" and "Name", each with a download icon. An "Action" dropdown menu is below.

At the bottom of the dialog are buttons for "Import", "Modify Composites", "Modify Layers", "Close", "Cancel", and "Help".

In the background, the SAS main window is visible, showing a menu bar (File, View, Tools, Solutions, Window, Help) and a toolbar. A "Results" window on the left shows a tree view with "Results" expanded, containing "Gmap: all subRHAs" and "Gmap: Edmonton subRHAs".

At the bottom of the screen, there is a blue banner with the text "SAS/GIS 9.1" and "The Power to Know." Below this, it says "Copyright ©2002-2003 SAS Institute Inc. Cary, NC 27513, USA. All Rights Reserved."

Map Boundary Files



```
*this is a nice summary file*;
Data whois;
infile kramer dlm=",";
format subRHA_name $60.;
input subRHA subRHA_text $ subRHA_name $ junk1 junk2;
subRHA_text = dequote(subRHA_text);
subRHA_name = dequote(subRHA_name);
RHA = int(subRHA/100);
RHA_text = substr(subRHA_text,1,2);
myregion = _N_; **note: myregion is 1..70 in the data order,
not RHA order*;
subRHAorder = input(put(subRHA,sRHAord.),5.0); *1..70 in
same order as subRHA values*;
drop junk;;
run;
```

```
*this is the raw text boundary file - ugly*;
Data boundaryraw;
infile gordon missover;
format t1 $255.;
format t2 $255.;
input t1 $ t2 $;
run;
```

```
*process the file*;
Data boundary2;
Set boundaryraw;
retain region;
retain myregion 0;
retain mifcount;
true=0;
if t1="Region" then do;
region=input(t2,3.0);
myregion + 1;
end;
lagt1=lag(t1);
if lagt1="Region" then do;
mifcount=input(t1,6.0);
end;
if (-1000 < input(t1,best.) < -1) and (0 < input(t2,best.) < 1000)
then do;
lat = input(t2,best.);
lon = input(t1,best.);
true=1;
end;
run;
```

```
*remove junk lines*;
Data boundary;
Set boundary2;
if true then output;
run;
```

```
*QA only - no idea what mifcounts are, but they uniquely identify
subRHAs*;
proc tabulate data=boundary;
class myregion region;
var mifcount;
table myregion*region, n mifcount*mean;
run;
```

```
*merge only boundaries*;
Data tempmymap;
Set boundary;
x=lon;
y=lat;
keep x y myregion;
run;
```

```
Data tempmymap;
Merge tempmymap whois;
By myregion; *merge by order in raw data*;
drop myregion;
run;
```

```
*Create mymap dataset*;
Proc GProject Data=tempmymap
Out=subRHAmmap
project=lambert
polelong=50
polelat=50

eastlong
degree
;
id subRHA;
run;
```

```
***Create Adjacency Matrix from Boundary Data***;
Data tempadj;
Set tempmymap;
*x y myregion*;
run;
```

```
Proc Sort Data=tempadj;
By x y subRHAorder;
run;
```

Projecting the Map

Create mymap dataset;

```
Proc GProject Data=tempmap Out=subRHAmmap
```

```
  project=lambert
```

```
  polelong=50 polelat=50
```

```
  eastlong
```

```
  degree;
```

```
  id subRHA;
```

```
run;
```

Mapping

Map Something;

Proc GMap Data=subRHArates

map=subRHA_projected;

id subRHA;

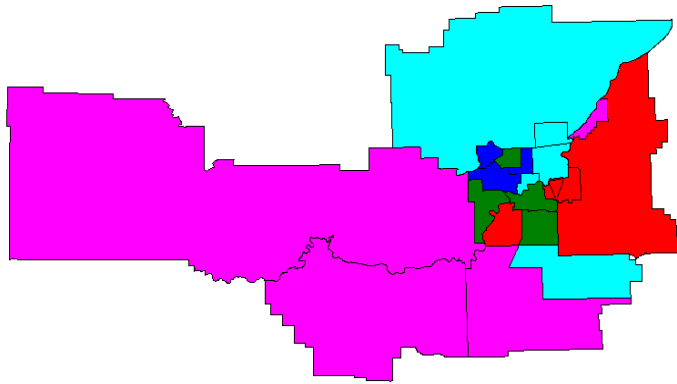
choro pred;

run;

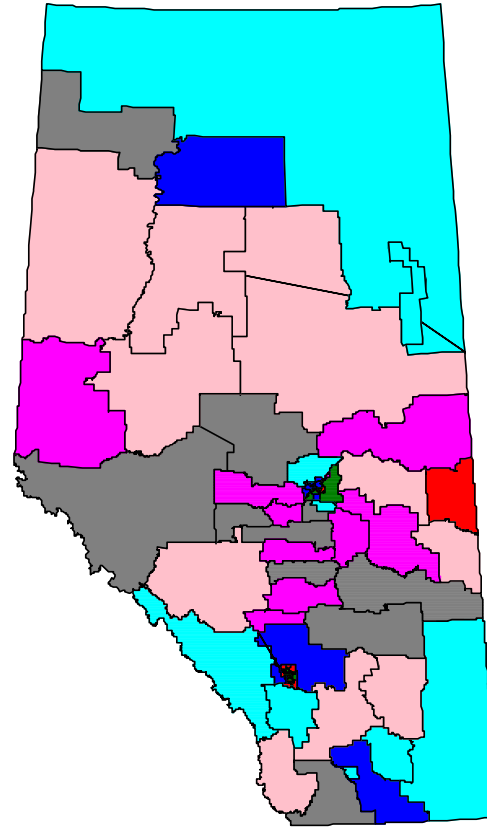
quit;

It all works!

Edmonton subRHAs



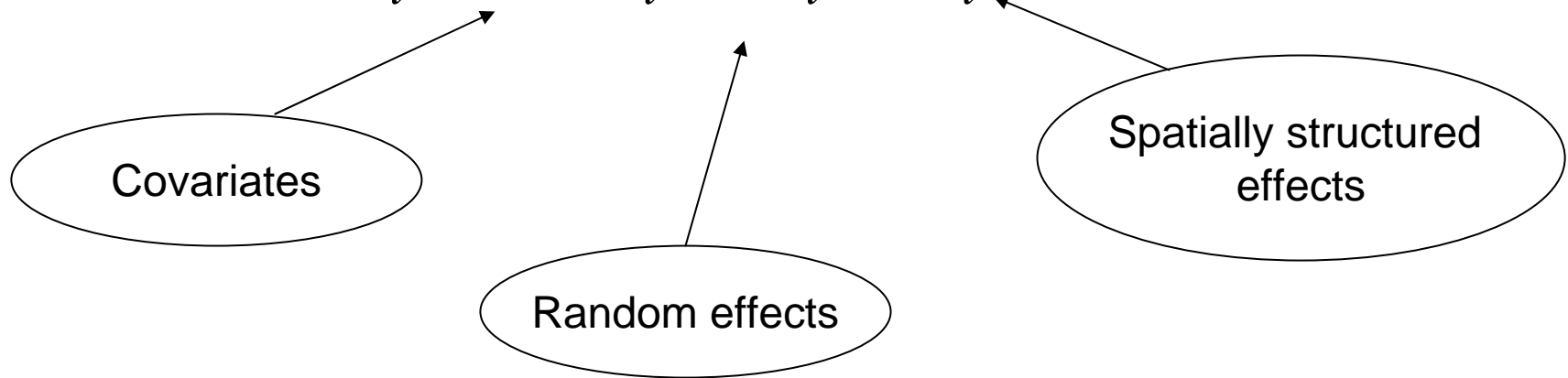
all subRHAs



BYM model in brief

- Besag-York-Mollie (1991)
- Rates in each region i are related by:

$$rates_i \propto \beta x_i + u_i + v_i$$



Modeling Spatial Relationships

- Rates in each region i are related by:

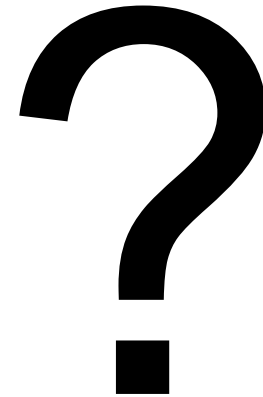
$$rates_i \propto \beta x_i + u_i + v_i$$

	101	102	201
101		1	0
102	1		1
201	0	1	

Spatially structured effects

Adjacency Matrix in SAS

	101	102	201
101		1	0
102	1		1
201	0	1	



Adjacency Matrix in SAS

Stop-gap solution was to use the GIS map data and find regions that shared latitude-longitude points.

```
Proc SQL;  
  Create Table adjacency as  
  /*cross product of all possible coordinates*/  
  Select a.subRHA as reg1, b.subRHA as reg2  
  from tempadj as a, tempadj as b  
  /*save same coord with different subRHAs*/  
  where a.x = b.x and a.y=b.y and a.subRHA ne b.subRHA  
  ;  
quit;
```

Modeling time

```
title "full BYM model";
%glimmix(data=subRHArates, procopt=covtest,

stmts=%str( class subRHAorder;
            model avgcount = / solution ddfm=sat;
            random subRHAorder / ldata=ICAR type=lin(1) solution;
            random subRHAorder / solution;
            parms (0.5) (0.3) (1)/eqcons=3;
            ),

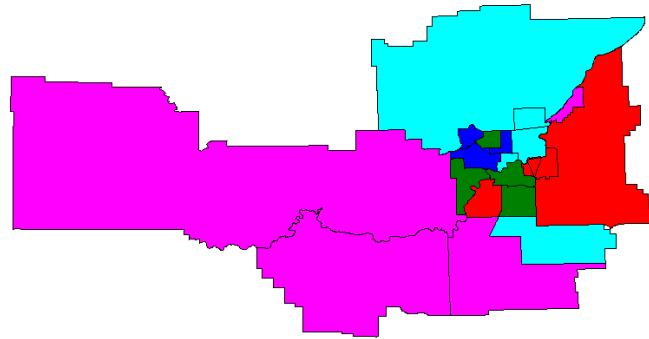
error=poisson, link=log, maxit=5, converge=0.0001, offset=logexpected

);
```

Smoothed Regional Estimates

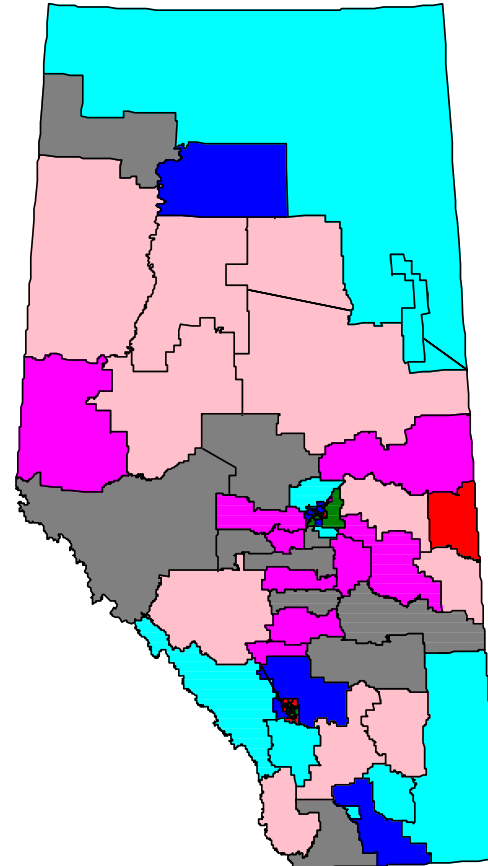
all subRHAs

Edmonton subRHAs



Predicted

■	-0.7623766 - -0.5655023	■	-0.3529624 - -0.2418017
■	-0.1643173 - -0.0436929	■	-0.0218166 - 0.1896029
■	0.3334788 - 0.5781629		



Predicted

■	-1.02418958 - -0.64156873	■	-0.61490266 - -0.35296235
■	-0.33661687 - 0.00849744	■	0.04775726 - 0.22625037
■	0.22668833 - 0.50576691	■	0.51166296 - 0.58162170
■	0.58926720 - 0.85726784		

I have a lot to learn.