

Conditional Maps in Arcview*

J. Blair Christian
Dept of Statistics, MS-138
Rice University
Houston, TX 77005-1892
blairc@stat.rice.edu
713-348-2695

David W. Scott
Dept of Statistics, MS-138
Rice University
Houston, TX 77005-1892
scottdw@stat.rice.edu
713-348-6037

Abstract

In an accompanying paper, we describe an implementation of a smoothed conditional mapping system, which we have implemented in arcview. Due to space limitations in that paper, we included only a small fraction of the possible cancer-related graphs based upon the SEER collection. In this demonstration, we will make available our arcview software add-ons, and demonstrate its performance with real NIH-NCI data. We will compare the results to the more conventional color-matrix choropleth displays (also in arcview).

1 Introduction

In practice, visualization of multivariate spatial data plays an important role in the effective dissemination of government data to both researchers and the general public. This demonstration presents a graphical user interface implemented in Arcview that creates smooth, conditional maps as described by Christian and Scott (2002). The demonstration will present data from the SEER (1998) cancer database, and allows the user to choose parameters for the variables of interest. These conditional maps are compared with conditional choropleth maps for the SEER breast cancer mortality and mammogram rates by county.

2 Using the Graphical Interface

The graphical user interface, known as a “dialog” in Arcview, is shown in Figure 1. The user chooses a “theme” from the current view. The theme must be an arcview shapefile with points as the shape field, and at least two variables of interest. The data in our examples was created by taking cancer data by county (polygons), and converting the shapefile to one where the centroids of the polygons are used as the points in the shape field, and our variables of interest are breast cancer rates (“Brsmooth”) and tercile of mammogram screening (“Tercile”) for each county.

*Research was supported in part by the National Science Foundation grants NSF EIA-9983459 (digital government) and DMS 99-71797 (nonparametric methodology). The authors would like to thank Linda Pickle and Sue Bell for their collaboration and for providing the data, as well as DG collaborators Drs. Carr, MacEachren, and Brewer.

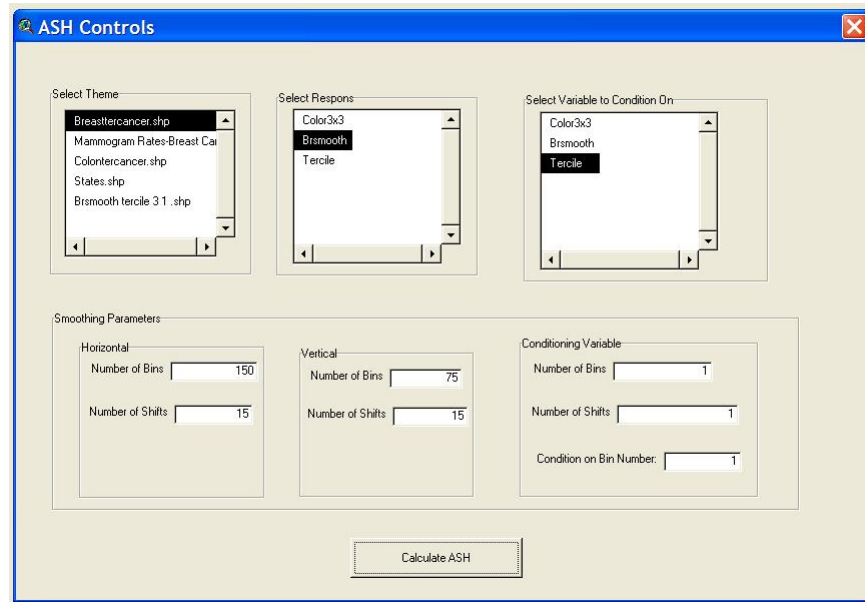


Figure 1: The Arcview dialog window that controls the smooth conditional maps.

After choosing a theme, the user then chooses the variable of interest in the “Select Response” box, which is cancer mortality in our case. Next the user chooses the variable that she wishes to condition on; this variable is the cancer screening rate in our example.

Finally, the user chooses the bin widths and smoothing parameters. For data over the entire United States, spatially square bins seem to produce more appealing pictures, so typically the horizontal smoothing parameter is chosen to be twice the size of the vertical smoothing parameter in order to preserve this geographically square bin shape. Smoothing parameters are also typically chosen to be 10% or less of the number of horizontal bins.

3 Examples

We briefly examine the spatial distribution of breast cancer rates as a function of the percentage of adults having a mammogram during a two year period; see Pickle and Su (2002) and SEER (1998). In Figure 2, we begin by displaying the original data, choropleth maps of both the breast cancer rates and the mammogram rates. Below each of those are smoothed maps of the rates.

Next in Figure 3, we present the conditional maps, a 3×3 conditional choropleth map and three slices from an ASH conditional regression map. Clearly, the fine detail in the conditional regression maps is more easily comprehended.

Please note that the files needed to make conditional maps in Arcview are available via ftp at <ftp://ftp.stat.rice.edu/pub/blairc/GIS>. Also note that the 9-class sequential legend was created using Brewer’s (1999) recommendations via the Color Brewer, available at <http://www.colorbrewer.org>.

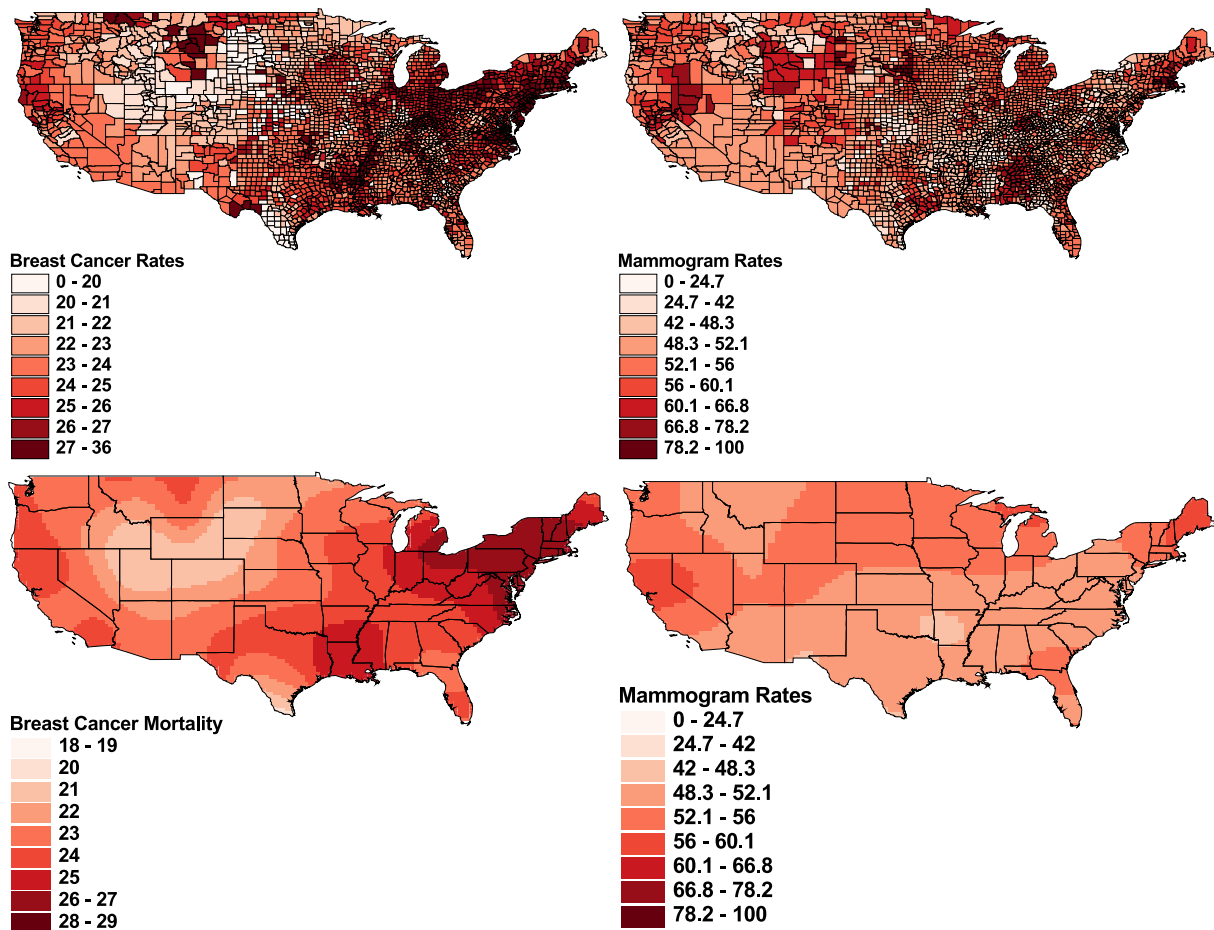


Figure 2: Breast Cancer Rates (left), and Mammogram rates (right).

References

- Brewer, Cynthia A. (1999), "Color Use Guidelines for Data Representation," Proceedings of the Section on Statistical Graphics, American Statistical Association, Baltimore, pp. 55-60.
- Pickle, L. W. and Su (2002), "Within-state geographic patterns of health insurance coverage and health risk factors in the United States," *Am J Prev Med* 22(2):75-83.
- Scott, D.W. and Christian, J.B. (2002), "A Theory and Implementation of Smooth Conditional Maps," *Proceedings of the Second National Conference on Digital Government*, E Hovy, Ed., Digital Government Research Center, Los Angeles.
- SEER (Surveillance, Epidemiology, and End Results) Program Public-Use Data (1973-1998), National Cancer Institute, DCCPS, Surveillance Research Program.

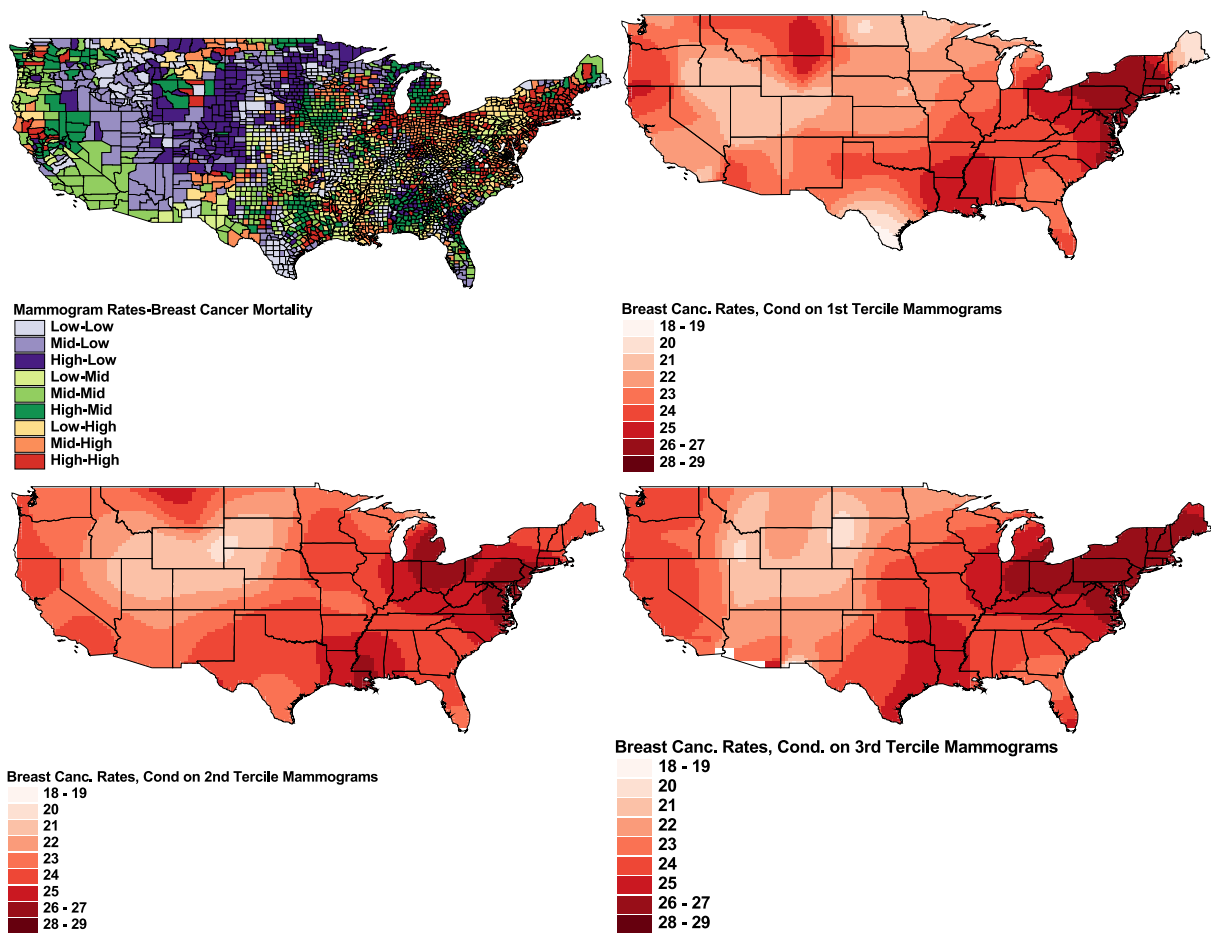


Figure 3: Breast cancer rates versus mammogram rates. One choropleth and 3 ASH estimates.