Interactive Linked Micromap Plots And Dynamically Conditioned Choropleth Maps

By Daniel B. Carr[1,2], Jim Chen[1], Sue Bell[2], Linda Pickle[2]
Affiliations: George Mason University[1], National Cancer Institute[2]

Extended Abstract:
This paper describes two recently developed templates for displaying geospatially-indexed estimates: linked micromap (LM) plots and conditioned choropleth (CC) maps. Two common goals in developing these templates were to integrate more statistical information in a display than a traditional choropleth map and to provide for more rapid assessment of statistical and spatial patterns than would be provided by a table. The particular layout and integration of information makes these templates distinct from previous graphical templates.

The primary purpose of this paper is to present recent extensions of the two templates and partial results from ongoing usability assessment. Much of the recent research and Java implementation has centered at the National Cancer Institute (NCI). The particular goal there is to develop, evaluate, improve and deploy methodology for communicating State Cancer Profiles to state epidemiologists and other public health professionals. More broadly, the research is a part of NSF-funded digital government research to develop quality graphics for federal statistical summaries. The templates and extensions are relevant to communication and hypothesis generation efforts of numerous government agencies.

The most current work at NCI on LM plots is in progress and not yet approved for public release as of the writing of this abstract. A Java applet showing trial interactive extensions to LM plots is available http://www.netgraphi.com/cancer4/index.html. (Contact jchen@cs.gmu.edu if the site is down.) The displays are of test data, not of official cancer statistics. The full paper will address many changes emerging from ongoing usability evaluations. CCmaps is a java application and available as shareware from www.galaxy.gmu/~dcarr/ccmaps.

Since the two templates are little known, this abstract tersely describes the basic elements of the two templates as illustrated in Figures 1 and 2. For those wanting more description, a series of papers [1,2,3,4,5] describes the LM plot template and/or illustrates different applications, design variations, and uses. The only published description of CCmaps is in [4].

The primary purpose of LM plots is the communication of statistical summaries but [3] describes its application in data mining. Key features of the LM plot template are present in Figure 1. There are four columns of panels. The left column contains micromaps, the second contains names, the third and fourth contain statistical panels. LM plots have three types of panels (micromaps, names and statistical panels) that can take various forms. For example a micromap can be any “spatial” representation from a human body caricature to a communication network.

Additional key features of LM plots are sorting, perceptual grouping, and linking of multivariate descriptors. The study units in Figure 1 are states. These are sorted by bronchus and lung cancer mortality rates that appear in the third column. After sorting, states are partitioned into small perceptual groups. States are grouped into fives (with one exception) with groups represented as being above, equal to, or below the median. Distinct hues distinguish the five states in each group. The same hue links a state name, its representation in the micromap, and its estimates in statistical panels. Vertical position also links the name and estimates in most LM plots.

The template includes different kinds of statistical panels such as time series and scatterplots. In Figure 1 the statistical panels are dot plots with confidence bounds and reference lines. In contrast to the choropleth maps, confidence bounds are shown and estimates are displayed with high perceptual accuracy of extraction by using position along a scale. Lines at the edge of the green regions indicate the U.S. Healthy People 2010 targets. The black line is the U.S. reference value.

NCI research is evaluating many interactive options. The more obvious options include selecting different data and sorting (triangle icons appear above the columns.) Additional features include mouseovers and linked blinking, color selection, a fixed header scroll, enlargement for micromaps, and drill down to see the counties of the selected state.

The purpose of CCmaps is to help researchers generate sharper hypotheses about observed spatial patterns. The particular context considered here is mortality mapping. Before conjecturing about the spatial patterns of mortality rates, it is important to produce maps that control for suspected risk factors. In epidemiology, common practice makes separate plots by race and sex to control for differences in study unit populations. Also, an age-adjusted map or a set of age-specific maps controls for study unit differences in age distribution. However, it is uncommon to see maps where efforts have been taken to control for risk factors. Sophisticated regression models provide the best means of controlling the variation due to risk factors. CCmaps provides more rudimentary but widely accessible control by partitioning study units into more homogeneous groups based on two risk factors.

As Figure 2 suggests, CCmaps supports dynamic partitioning of study units into a 3 x 3 layout of maps using newly developed partitioning sliders. The study units in Figure 2 are health service areas (HSAs), counties or aggregates of counties based on where people get their hospital care. HSAs highlighted in a panel have risk variables satisfying row and column constraints. The slider at the bottom partitions HSAs into columns based on precipitation. The slider at the right partitions HSAs into rows based on percent of households below the poverty level. The slider at the top partitions HSAs into three color classes (shown as blue, gray and red) based on lung cancer mortality rate. Hypotheses can be about spatial patterns within panels or differences among panels. Note the red in the top right panel that highlights HSAs with high precipitation and high poverty. One hypothesis might be that Southeastern HSAs have higher cigarette smoking rates. However the strong association with precipitation warrants deeper consideration. The paper will touch on issues of data availability and confounding variables.

Current features of CCmaps include statistical annotation and plots. Sliders show the percent of **people** in each class. The population weighted mean rate for HSA highlighted in each panel appears at the top right. A 3 x 3 layout of dynamic QQplots (not shown) facilitates comparing distributions. The full paper will describe numerous extensions.

[1] DB Carr and SM Pierson. "Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps," *Statistical Computing & Graphics Newsletter*, 1996; **7**(3): 16-23.

[2] DB Carr, AR Olsen, JP Courbois, SM. Pierson, and DA Carr. "Linked Micromap Plots: Named and Described," *Statistical Computing & Graphics Newsletter*, 1998; **9**(1): 24-32.

[3] DB Carr, AR Olsen SM Pierson, and JP Courbois. *Using linked micromap plots to characterize Omernik ecoregions. Data Mining and Knowledge Discovery* 2000; 4:43-67.

[4] DB Carr, JF Wallin, and DA Carr. “Two New Templates for Epidemiology Applications. Linked Micromap Plots and Conditioned Choropleth Maps,” *Statistics in Medicine* 2000; 19:2521-2538.

[5] DB Carr. Designing Linked Micromap Plots for States with many Counties,” *Statistics in Medicine* 2001; 20:1331-1339.

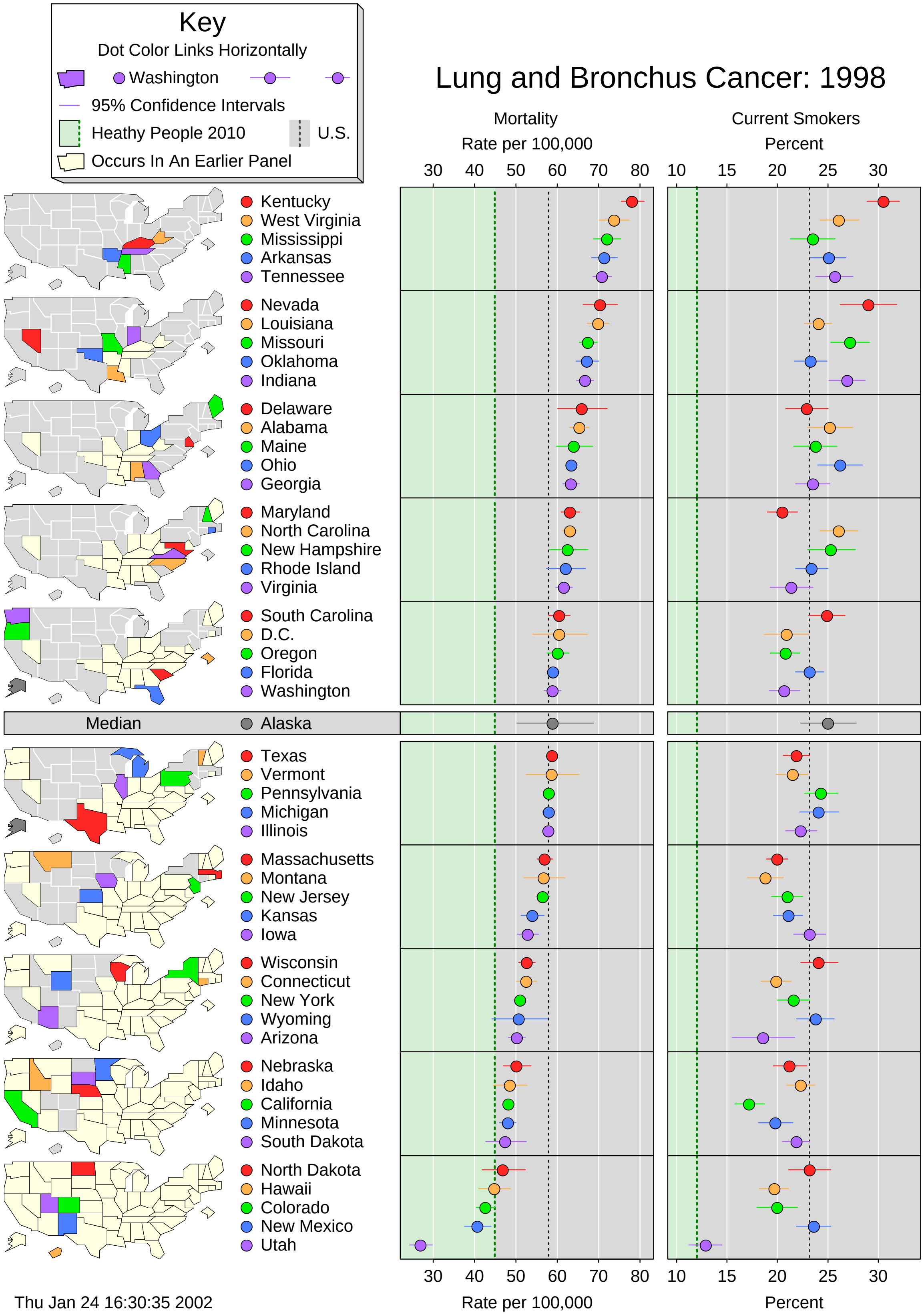

Figure 1. A Linked Micromap Plot

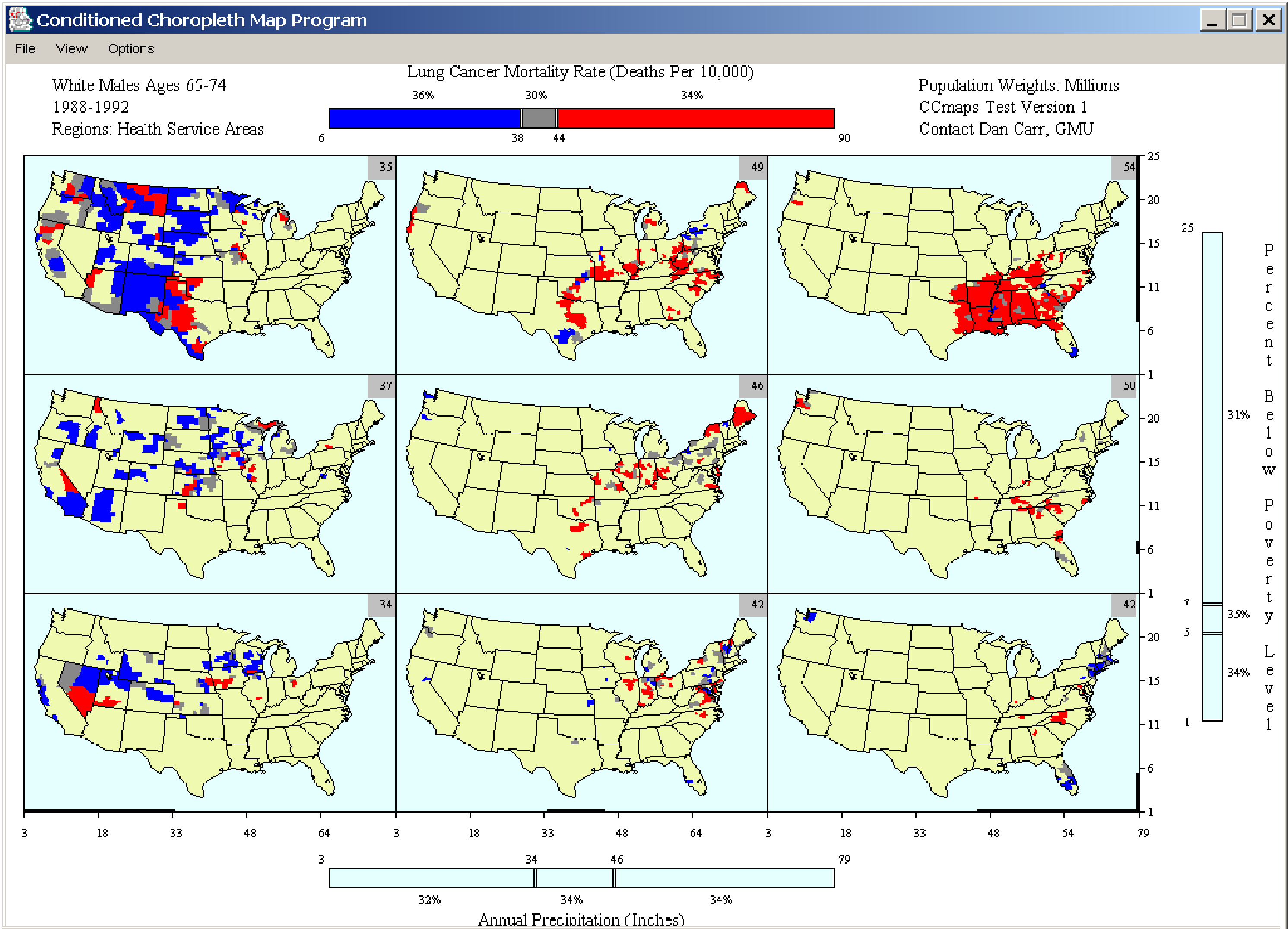

Figure 2. A Conditioned Choropleth Map