The State Cancer Profiles Web Site and Extensions of Linked Micromap Plots and Conditioned Choropleth Map Plots

Daniel B. Carr[1, 2], Sue Bell[2], Linda Pickle[2], Yuguang Zhang[1], and Yaru Li[1]
Affiliations: George Mason University[1], National Cancer Institute[2]
Email: dcarr@gmu.edu, bellsu@mail.nih.gov, picklel@mail.nih.gov

Abstract

The National Cancer Institute's State Cancer Profiles web site that was launched in mid-April put the power of new graphical tools into the hands of cancer control planners and the public. This paper emphasizes advances in two graphics templates, one heavily used in the web site and one planned for future use. These templates for showing geospatially-indexed statistics address needs that are less well addressed by more conventional maps, such as the representation of estimate uncertainty. The web site and templates both evolved through extensive usability testing of targeted users. The results described in this paper signal highly visible acceptance of the new interactive and dynamic forms of these templates.

1. Introduction

This paper describes a new Web site that combines the latest research in data visualization sponsored by the National Science Foundation's Digital Government Quality Graphics initiative with features to ensure accessibility by the visually impaired. After providing a brief background on the Web site, the focus turns to the recent development of Linked Micromap (LM) plots, an integral component of the new Web site. Then we describe advances in the dynamically conditioned choropleth (CC) maps, part of future site development plans. The paper closes with comments about the anticipated transfer of this graphics software to other federal agencies who are participants in the data visualization research and indicates the immediate availability of conditioned choropleth maps as shareware.

2. The State Cancer Profiles Web Site

The State Cancer Profiles Web site provides a "quick stop" for cancer related statistics for planners, policymakers, and epidemiologists. It was developed by the National Cancer Institute (NCI) in collaboration with the Centers for Disease Control and Prevention (CDC) and is an integral component of NCI's Cancer Control PLANET, a Web portal that links to resources for comprehensive cancer control. The Web site provides national, state, and county views of cancer statistics collected and analyzed in support of annual federal reports. Focus is on eight major cancer sites for which there is evidence of the potential to improve outcomes either by prevention or by screening and treatment.

Cancer statistics include mortality and incidence counts, rates, and trends by sex and race/ethnicity. Recent incidence data are available for cancer registries participating in CDC's National Program of Cancer Registries (NPCR) that met selected eligibility criteria. Both historical and recent incidence data are available for cancer registries participating in NCI's Surveillance, Epidemiology and End Results (SEER) program. Prevalence of risk factors and screening, health-related goals for the next decade (NCHS 2000), and demographics complete the profiles. The interactive graphic capabilities allow a user to quickly explore patterns and potential disparities. For example, the user can easily compare graphs of national or state trends for Whites, Blacks, Hispanics, Asian Pacific Islanders, and American Indian Alaskan Natives. LM plots provide the primary graphic template for users to explore spatial relationships among the latest rates, percents, and counts for cancer statistics, demographics, risk factors, and screening (http://www.statecancerprofiles.cancer.gov/micromaps/).

3. Linked Micromap Plots

Linked micromap (LM) plots follow a template that has emerged during the last decade. The basic template has been described (Carr et al. 1998, Carr et al. 2000, Carr, Wallin, and Carr 2000, and Carr

2001) and its wide applicability noted. The key features of LM plots include panels of region names, statistical graphics, and small maps called micromaps. As indicated in Figure 1, a screen shot from the Web site, a common color links a state name with statistical estimates of cancer rates, risk factors or demographics and the region polygon on the map. The template allows for an accurate representation of rate uncertainty and reference values such as Healthy People 2010 objectives (NCHS 2000) shown as the green area. The link between statistical values and geography makes LM plots very useful for revealing and communicating spatial patterns in the data. Interactive extensions of LM plots (Wang et al. 2002 and Carr 2002) are now in use. For example, the triangles in Figure 1 provide for sorting the entire graphic using one of the display columns.

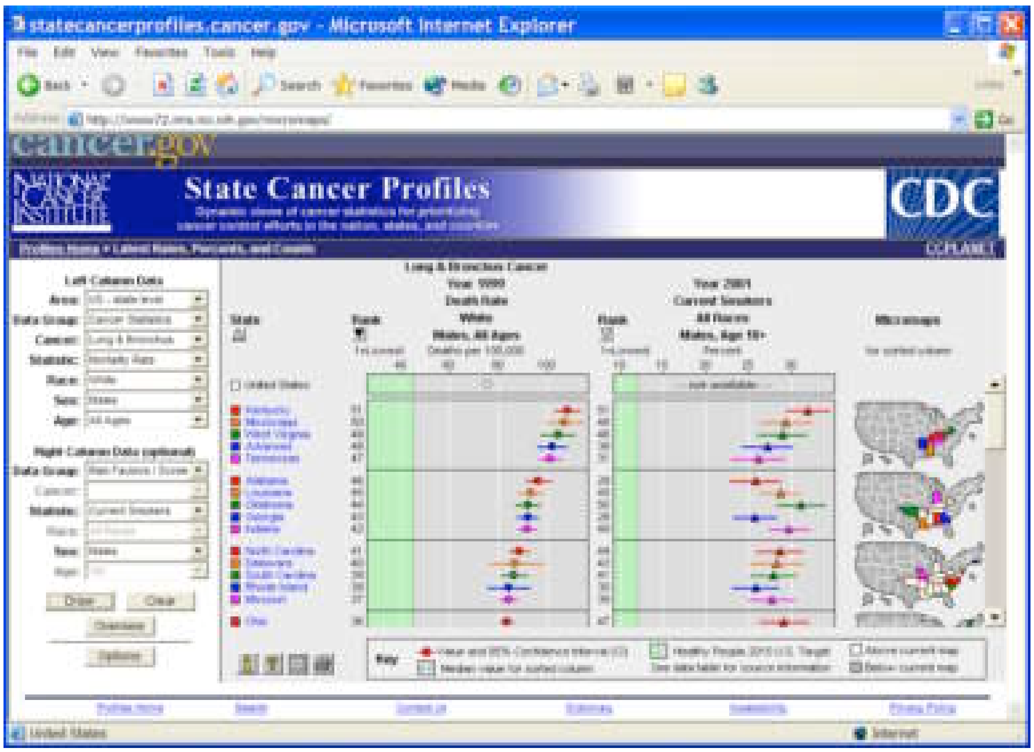

Figure 1. Link Micromap Plots. Only three full rows of panels show.

While some design changes (e.g., variable selection and labeling) were specific to the NCI implementation, the major design challenge arose from taking LM plots from a portrait print image to a landscape screen image. The change in orientation significantly affected the number of rows that could be displayed. For example, on the screen it was not possible to legibly display all states, much less all 254 counties in Texas. The requirement for a standard banner and logos for branding exacerbated the problem by further reducing the vertical screen real estate available for data display. Laboratory-based usability testing immediately drew attention to problems with the initial design in which both the browser window and the LM plot pane had vertical scroll bars. Users were drawn to the browser window's scroll bar and became confused and frustrated looking for states not shown. In the current design, there is no scroll bar for the browser window if the screen area setting is for 1024 by 768 pixels or more. In the next round of usability testing, users intuitively scrolled down to find the additional states. However, observation of users at their desktops identified a key group of users who either were limited to or preferred an 800 by 600 screen area setting. Challenged with meeting the needs of this subgroup but not wanting to penalize users with higher screen area displays, a further re-design minimized the need for horizontal and vertical scrolling by ensuring that the graphical display fit in an 800 by 600 screen area.

Scenarios were tested where users entering with lower screen area settings were provided with a pop-up window to advise them of the option of changing their screen area setting.

The limitation on the vertical display area and screen resolution motivated the development of an analyst-centric map and plot overview. This is accessible via an Overview button in LM plot view. The overview appears in a pop up window, where analysts can more easily compare how their state (county) ranks with respect to other states (counties).

The quantitative background of the targeted users of the State Cancer Profiles Web site is higher than the general public but is lower than researchers who have had formal statistical training. We added lines indicating a 95% confidence interval to the points representing rate estimates in order to communicate the uncertainty in rate estimates. Because the addition of these lines increases the data range and thereby reduces the screen resolution available for representing the point estimates, LM plots provides an option to suppress the confidence lines. We also added a rank statistic beside each graphics panel to accommodate the desire of users to make comparisons based on relative rank. Usability tests confirmed our conjecture that many targeted users would expect a rank of 1 to indicate the "best" rather than the lowest value as in a traditional statistics setting. This can lead to communication problems. For mortality rates a high rate is "bad" while for screening a high rate is "good". An increase in incidence rates might be considered good in the context of improved screening. We are careful to label rank 1 as highest or lowest and to avoid the context dependent word "best". As the confidence lines help to show, many consecutive ranks may not be statistically different. Future research work will seek to provide an additional visual indication so the user knows which ranks are statistically different.

Effective in June 2001, Section 508 of the Rehabilitation Act took effect requiring that federal agencies' electronic and information technology be accessible to people with disabilities. In January 2003, two blind users were tested on the State Cancer Profiles Web site. Although LM plots were not included in this initial test, the blind users successfully navigated the Web site using the pick lists and the data table option following the same scenarios used to test sighted users. We anticipate that LM plots will also be compliant and usable by persons with disabilities because the same approach to data selection is used and a data table option is provided.

4. Dynamically Conditioned Choropleth Maps

Conditioned choropleth (CC) maps are under consideration for future inclusion in the State Cancer Profiles Web site and were used by epidemiologists who participated in a Cancer Surveillance Institute held January 2003. As indicated in Carr, Wallin and Carr 2000 and Carr et al. 2002, the ostensible purpose of CC maps is to promote sharper hypothesis generation about patterns appearing in maps. Figure 2 introduces an example based on analyzing smoothed colorectal mortality rates (males and females combined) for US counties over the period 1995 to 1999. This is an enlarged view with the partitioning sliders set to include all areas in a single map. The pattern of red (higher rate) counties is dramatic in the northeast and along the Mississippi River. Yellow shading indicates missing values.

The sliders allow an analyst to explore the data by partitioning a single choropleth map into a two-way layout of partial maps, each conditioned on values of two other variables. The regions appearing in the partial maps are more homogeneous with respect to the conditioning variables. Figure 3 partitions the choropleth map from Figure 3 using an estimate of the prevalence of colorectal cancer screening by fecal occult blood test (FOBT) for the column slider on the bottom and using the county's socioeconomic status for the row slider on the right. The best opportunity to improve colorectal outcomes is to increase screening in areas with low screening, shown in the first column, that have high mortality rates shown in red. Additionally, areas with lower socioeconomic status are classified on the bottom row. These areas may require more financial assistance to finance outreach, education, screening, and treatment programs. Using the sliders, an analyst can tease out more or fewer areas in a particular classification of high or low.

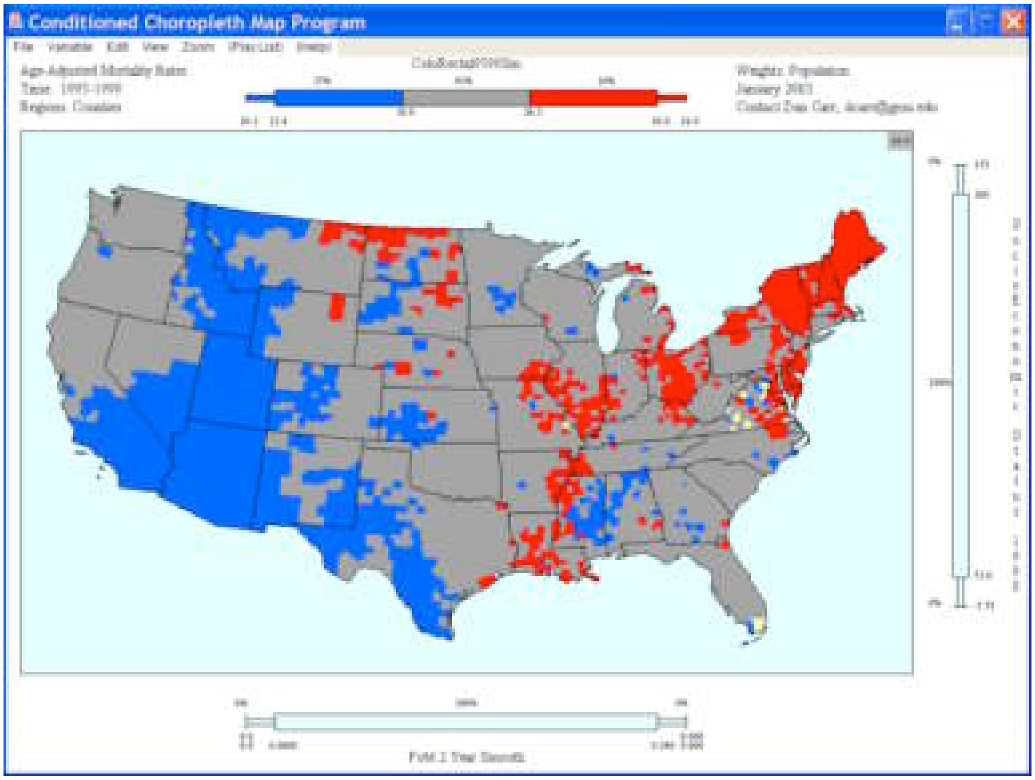

Figure 2. An Enlarged View With Sliders Set To Include All Available Regions

Recent progress on CC maps includes faster sliders, piecewise linear sliders to handle the commonly occurring outliers and skewed distributions, and variable selection from comma delimited files. The sliders are now fast enough to smoothly handle US county maps on a 2 MHz computer. Slider experience with county mortality rates and related risk variables motivated the development of a piecewise linear slider that maintains resolution in the central body of the data whether or not there are outliers. The color slider at the top of Figure 3 has thin bars at the left and right with a thicker bar in the much longer central region. The presence of thin bars for low values, high values or both, indicates that there is at least one outlier. When thin bars appear, the slider is piecewise linear. That is, the change in slider value is linear as the slider moves from pixel to pixel for each portion of the slider but the rate of change is faster for the outliers than for the main body of data, providing more resolution where there is more data.

5. Transfer to Other Agencies

Graphics research for the State Cancer Profiles Web site is part of the NSF-funded Digital Government research. Research with the US EPA, Bureau of Labor Statistics, Bureau of Transportation Statistics and the Department of Agriculture has used the templates to study biodiversity and characterize Omernik ecoregions, show unemployment statistics, display pedestrian fatalities, and show crop production statistics. The new Java graphics software is to be made available to these and other agencies associated with the graphics research. Because the Web site focuses on NCI applications and software, LM plot repurposing will involve alterations such as connecting to different databases. The development of CC maps has centered at George Mason University. This Java shareware is immediately available via the Web site www.galaxy.gmu/~dcarr/ccmaps.

In conclusion, data visualization based on quality graphics is now an integral component of the US cancer surveillance system. Cancer control planners have exciting and easy to use graphics to facilitate their analysis on where to focus limited resources to reduce the US cancer burden.

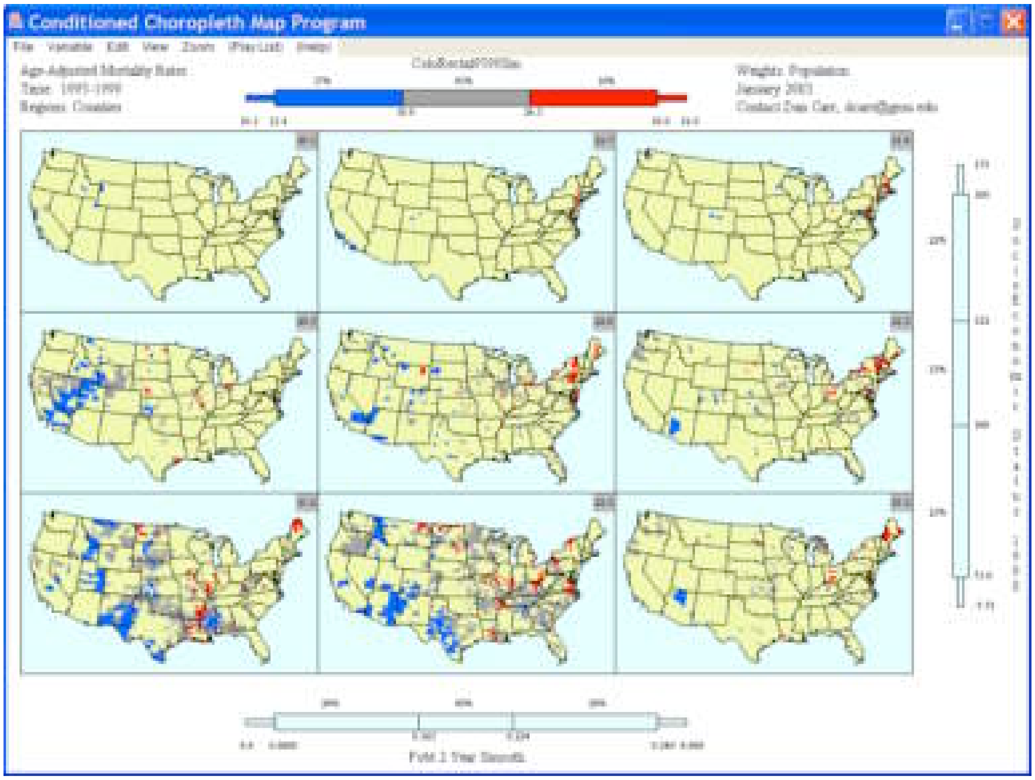

Figure 3. A 3 x 3 Partitioned View.

6. Acknowledgements

NSF Grant 9983461 supported this research. Bill Killam played a major role in the web site usability assessment. David Eyerman played a major role in the Java implementation of LM plots.

7. References

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

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

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

DB Carr. 2001. Designing Linked Micromap Plots for States with many Counties," *Statistics in Medicine* , 20:1331-1339.

NCHS. 2000. Healthy People 2010 Objectives. http://www.cdc.gov/nchs/about/otheract/hpdata2010/abouthp.htm, accessed April 11, 2003.

Wang, X., J. X. Chen, D. B. Carr, B. S. Bell, L. W. Pickle. 2002. May/June. "Geographics Statistics Visualization: Web-based Linked Micromap Plots," Ed. J. Chen. *Computing Science and Engineering*, pp. 92-94.

DB Carr, J Chen, BS Bell, L Pickle, Y Zhang. 2002. "Interactive Linked Micromap Plots and Dynamically Conditioned Choropleth Maps," *Proceeding of the Second National Conference on Digital Government*, Pub. Digital Government Research Center. pp. 61-67.