CCmaps Introduction
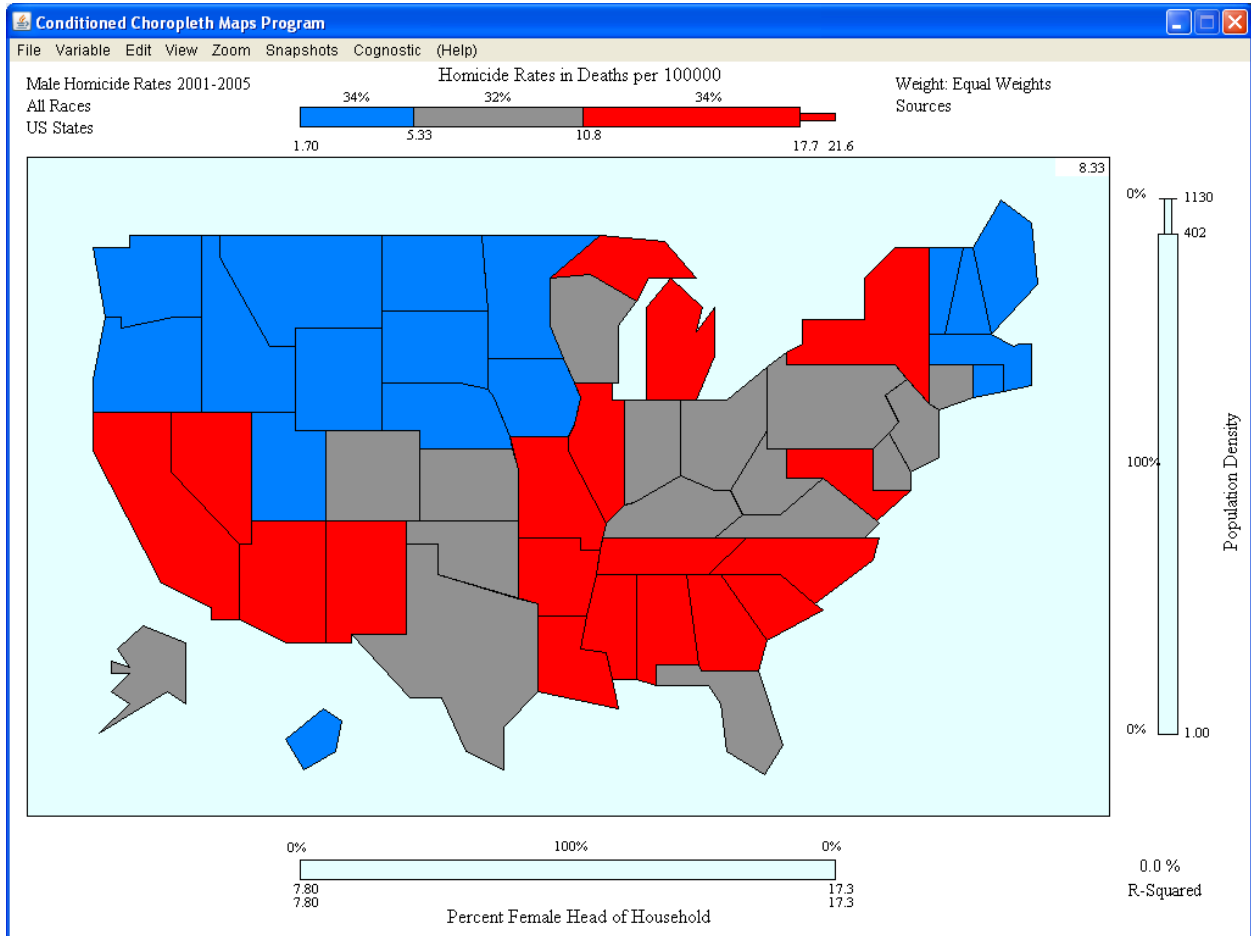## Conditioned Choropleth Maps and Visual Analytics
Draft April 14, 2010


Page   Topic

An Introduction to CCmaps
Conditioned Choropleth Maps and Visual Analytics

1. Introduction

CCmaps is free dynamic Java application for producing conditioned choropleth maps in integrated visual analytic environment. This primary purpose of this paper is to present a few definitions of concepts and introduce the CCmaps features. Other documents create the source of ideas, the designers, implementers and future directions.

2. Showing maps and representing three variables using dynamic sliders.

The choropleth maps encode a value for each region using color or shading. For example the regions may be US states and the values may be male homicide rates for 2001 to 2005. Classed choropleth maps partition the range of values into a few intervals and associate a distinct color with each interval.   The CCmaps example below is a 3-class example.
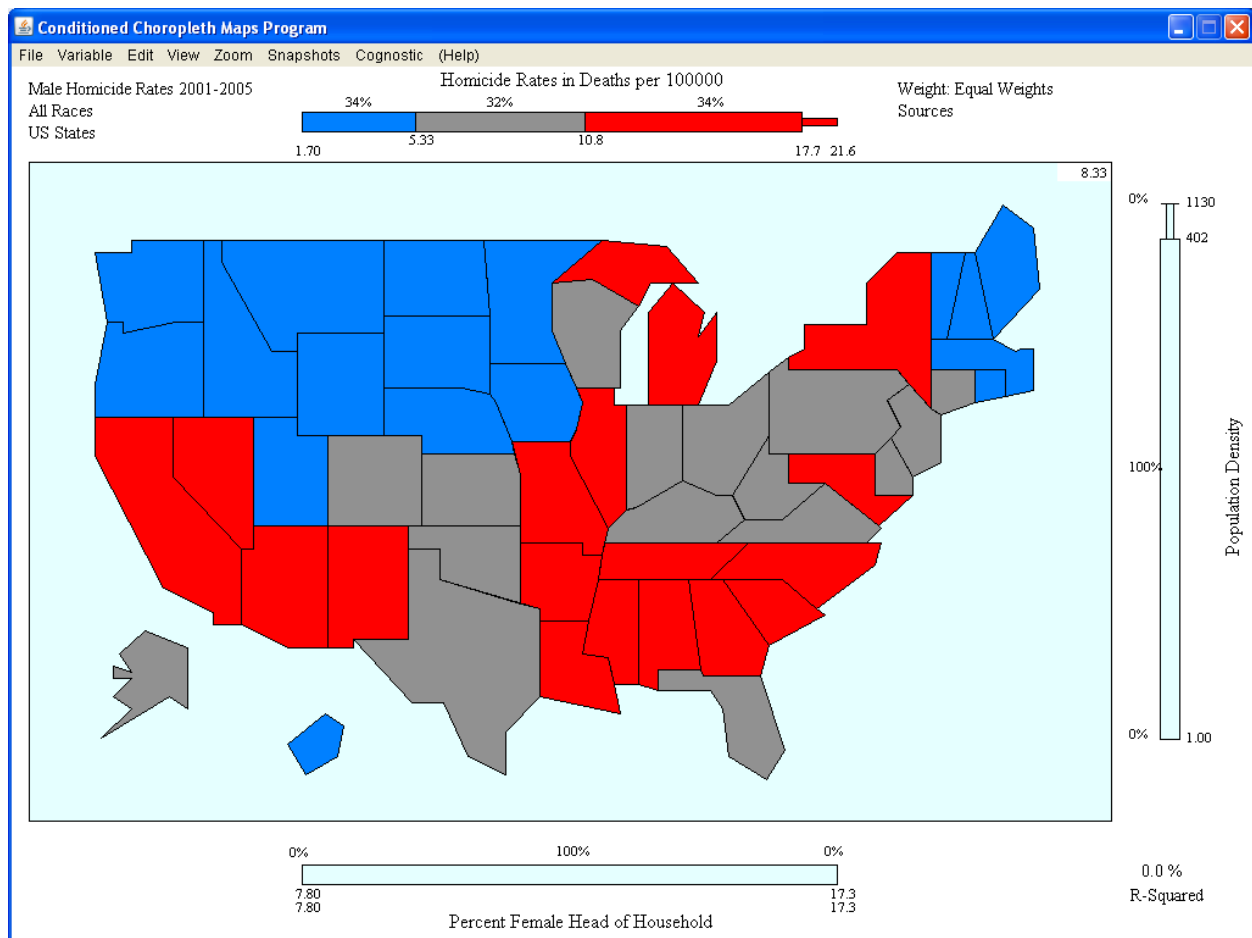
In CCmaps the color bar at the top is dynamic 3-class slider. The analyst uses this to define low, middle and high intervals by dragging the two interior vertical lines. Above, states with a value in the low interval from 1.70 to 5.33 appear blue in the map. States with values in the middle interval and high intervals appear gray and red respectively.   The analyst can change class thresholds and, using the Edit menu, the color associated with each interval.

The region color indicates class membership for one variable. CCmaps represents the region class membership for two more variables by its location in a two way grid of maps.   In the example below the analyst has associated the variable Percent Female Head House with the bottom slider and defined three classes, low, middle and high by using a mouse to click and drag the two interior thresholds, one at a time.
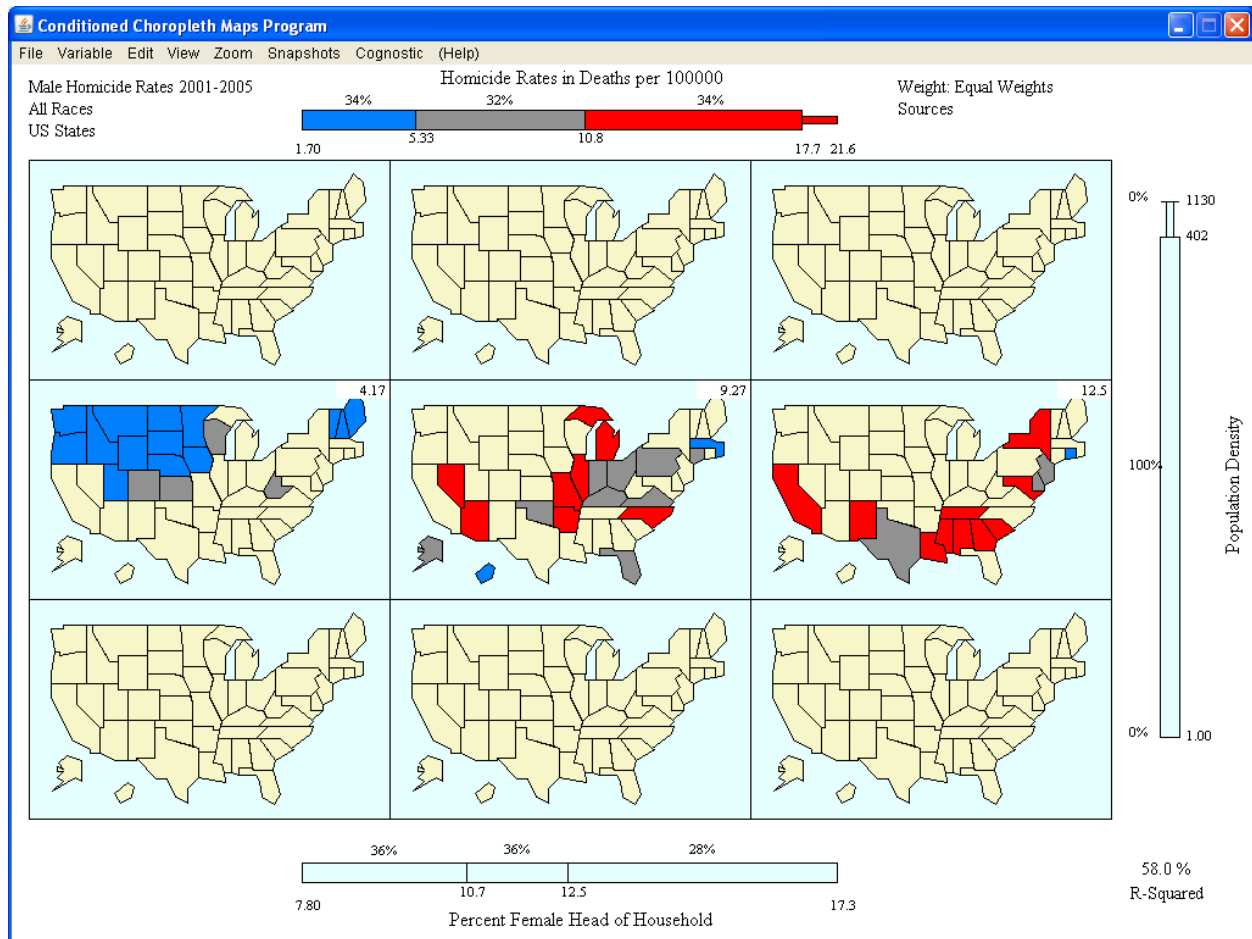
The regions with values in the low value interval appear in the left column.   Regions in values middle and high intervals appear in the center and right columns respectively.   We can see many blue states on the left, about an equal number of gray and red states in the middle, the number of red states dominates on the right.   The color and panel position makes the association of Homicide Rates and Percent Female Head of Household is obvious.

As you move a slider threshold, the states changing class memberships should update in a few milliseconds.   The response to slider change seems to occur instantly on most computers.
In the example below the analyst has associated a third variable, Population Density, with the right slider. The analyst has adjusts the slider to have the same percent of states in each of low, middle and high intervals.
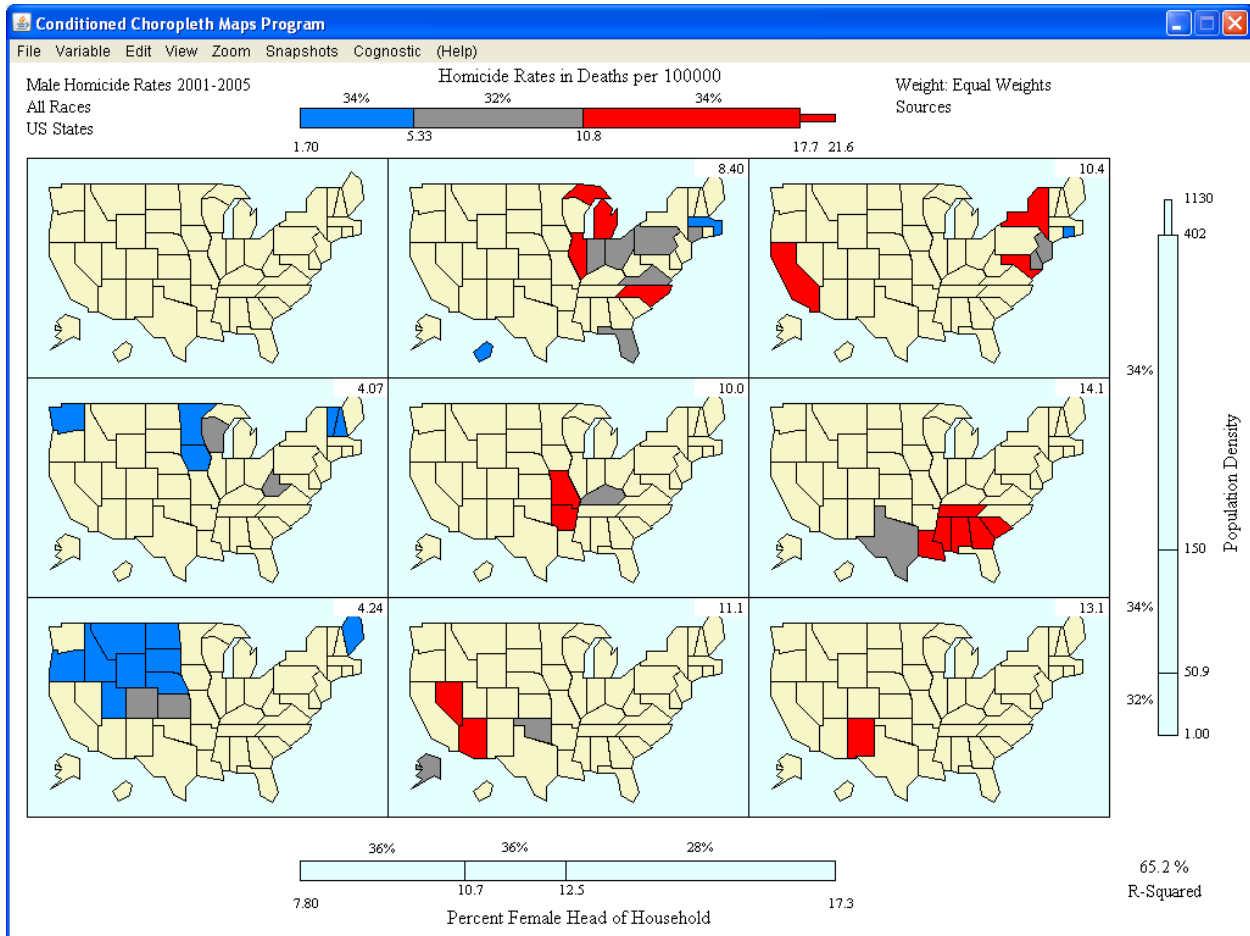

We can new see blue (low homicide rate) states in the lower left panel. We can see clusters of states such as in the middle row and right column panel.   The clustering of states (same panel and same color) indicates that neighboring states tend to have similar values for all three variables. If all the states appeared in panels along a diagonal and the colors changed from panel to panel, this would suggest highly correlated variables.

You can also adjust sliders to isolate the states with extreme values for any of the three variables.

3. Visual Analytics and Statistical Feedback

Visual analytics is the science of analytics reasoning facilitated by interactive visual interfaces. Visual analytics goals include the discovery of new patterns, hypothesis generation, confirmation of assumed patterns, and promotion of quantitative reasoning. Of particular importance here, the CCmaps interface provides quantitative feedback support analytic reasoning.

Percents appear above the class intervals in the top and bottom sliders and on the left of the class intervals in the right slider. If no variable is assign as a region weight, the weights are equal and the percents indicate the percent for regions in each of the occupied classes intervals. When a weight variable is selected, such as region population, the percents indicate the percent of the population associated with occupied class interval.
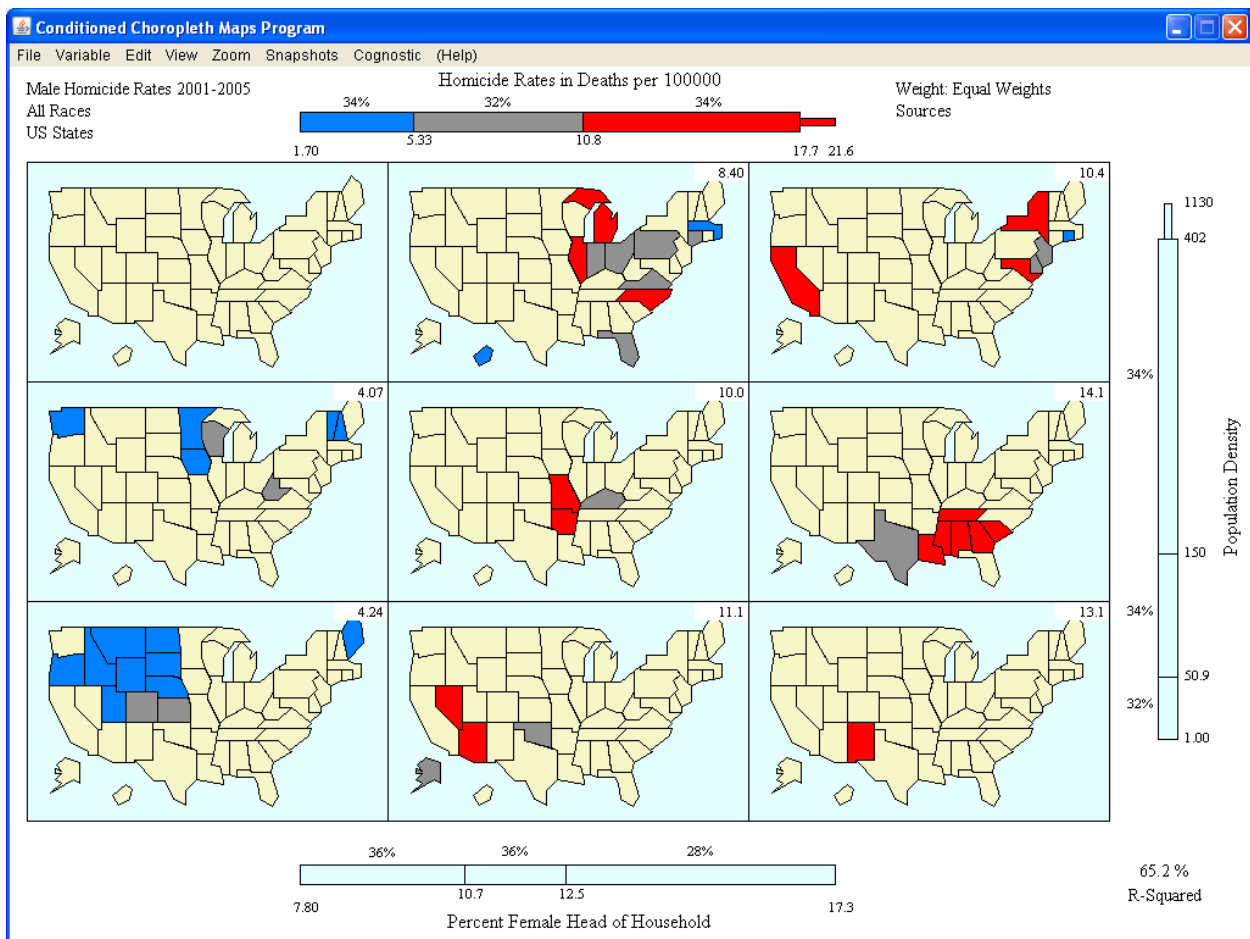

The study variable is typically associated with the color slider. The mean of the color slider variable for regions in each panel appears in the upper right of each occupied panel. We can observe how the means change across rows or columns or both. Above, the mean values increase from left to right as we read across each row. The panel means provide terse,

5

understandable and repeatable summary for each panel.   The impression base on the composite of region colors in panel can vary from person to person.

## 4. Panel Enlargement, Mouseovers, and Zooming

Mouse clicking in a panel causes it to be enlarged. The enlarge view provide access to the mouseover option. If the mouseover item is checked in the View Menu, moving the mouse over a region while cause its slider variable values to appear. The mouse was over Missouri when the screen show was taken. (The mouse was does not appear in screen shots.)
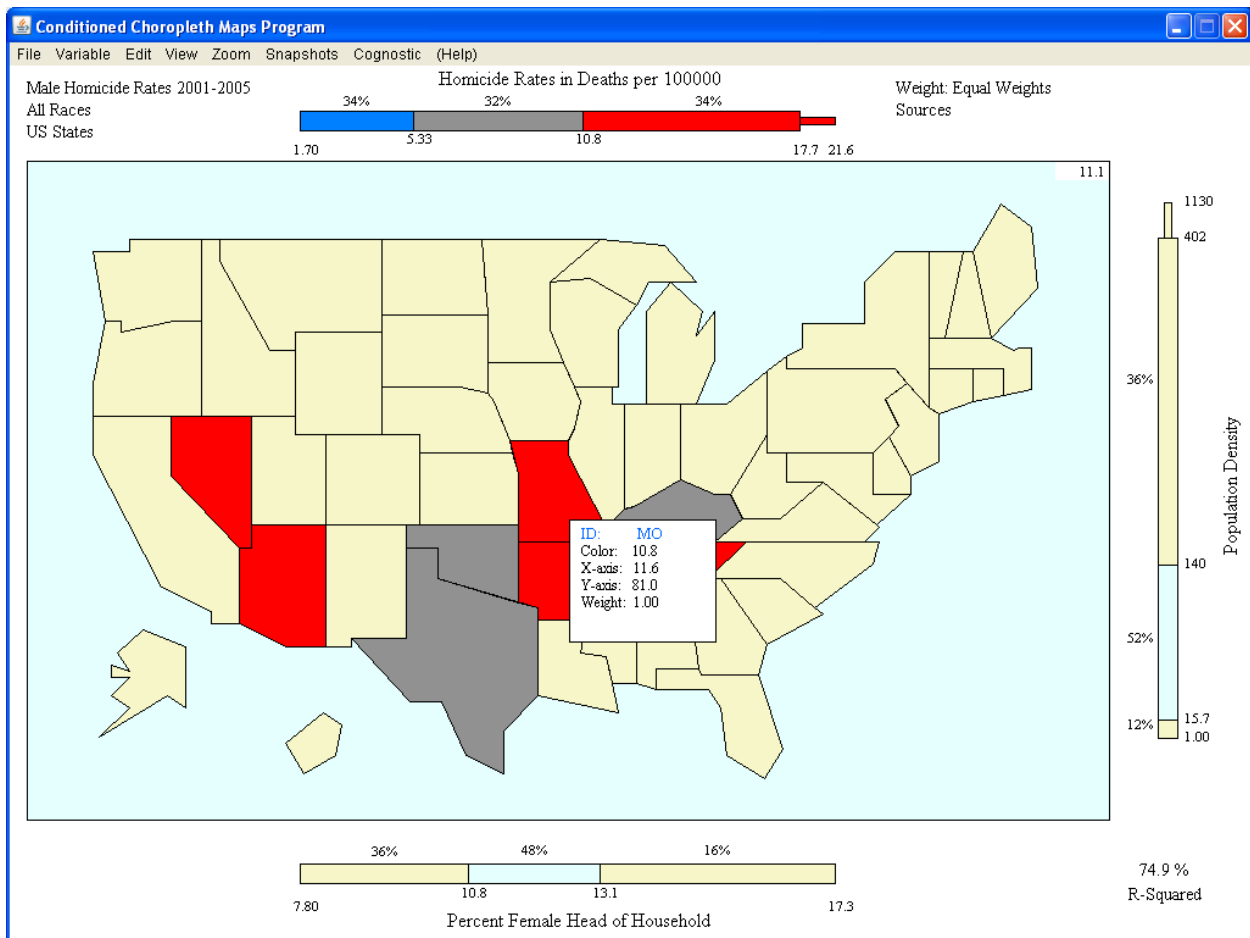

The blue intervals in the bottom and right slider bars indicate which panel in the 3 x 3 grid of panels was enlarged. Above, the blue intervals indicated this view is an enlargement of the middle column and middle row panel.

 Clicking in the enlarged map panel will return to the 3 x 3 map panel view.   The same click to enlarge and click to return procedure applies conditioned qqplot and scatterplot views.

The Zoom menu also works from the enlarged view.   There is a Zoom off/on toggle and a Zoom history list. With the Zoom on and the mouse can define a rubber band rectangle indicating the area to enlarge.   Actually two rubber band boxes appear.   One corresponds to the actual area enlarged that maintains the aspect ratio of the map. The zooming can be iterated. The Zoom history list enables refreshing to the initial state and returning to an earlier zoomed state.
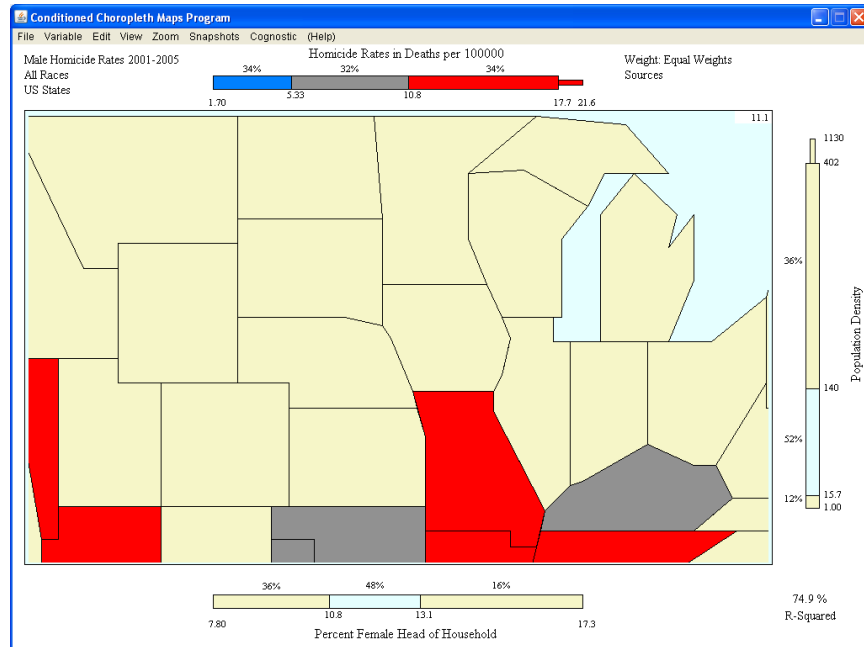
With the zoom turned off, clicking the panel returns to the 3 x 3 view. This is show below. Ideally there would be an option to update all the sliders to the excluded regions (totally or partly out of view.)   This has not been implemented.
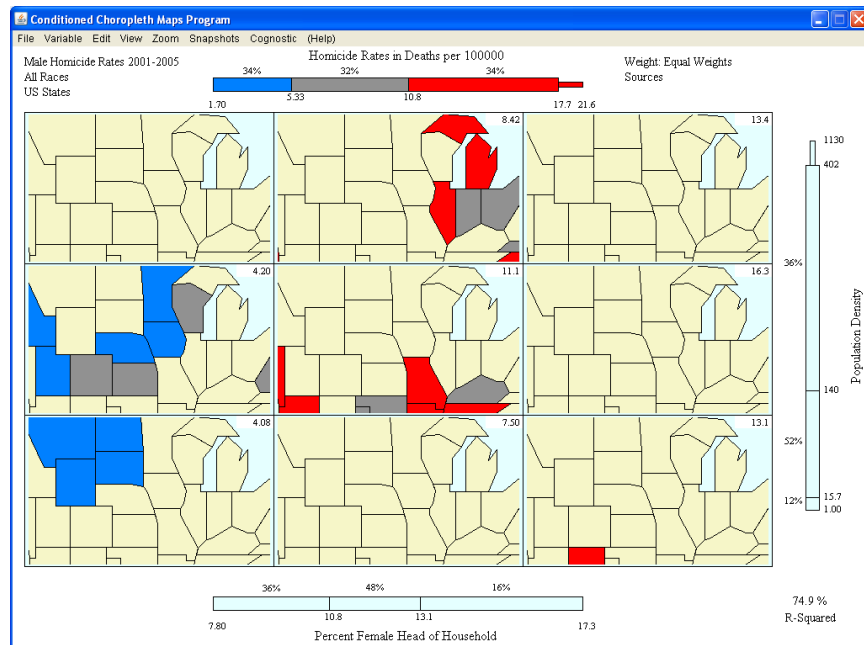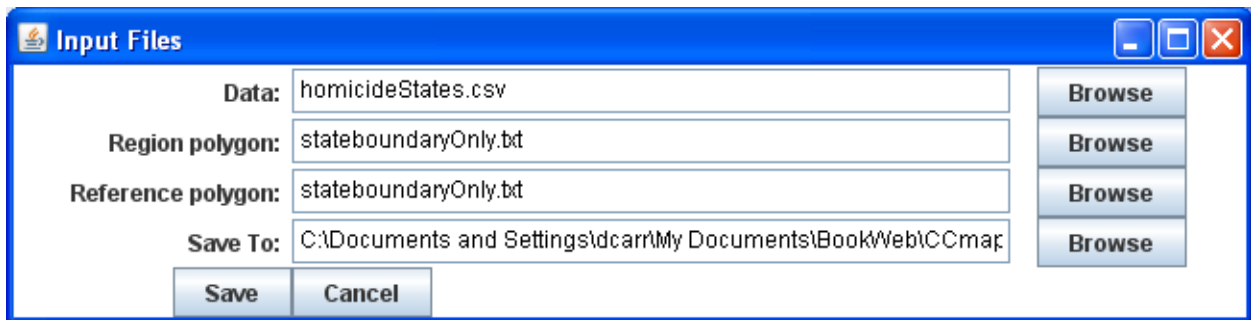
5. The File Menu

The has 4 item, new, open, save and exit.

5.1 New

This option prompts for four files names

The **data file**

This can be a comma or white space delimited file. The first row must be variables names. The first column must the region id that is used to link the region id in the region polygon file. This column is read as a character string variable.

If this column is white space delimited id's with white space must be surround by quotation marks. For example state names might ids and the link to the polygon ids. "New York" would need to have quotes in a white space delimited field

Currently all of the other columns must be numeric. The exception is that NA is interpreted as the missing data code in a white space delimited file. CCmaps used case wise deletion for the variables attached to the sliders. The regions may or may not be used dependent on the variables selected

Polygon files.

There may be two files. If the focus is US counties, that would the county polygons would be region files, and state polygons would be a reasonable reference. The region file can be entered in both places if there is no reference region files.

The current format is simple, the key word ID is follow the region name or id that links to the data file.   This followed by x and y coordinates for one polygon.   The same region can have more than one polygon. The files can be comma or white space delimited.

ID WA
1 89
7 89
7 91
18.5 91
18.5 77.5
14 77.5
5.5 75.5
5.5 77.5
3 77.5
ID MT
21.5 91
47.5 91
47.5 75.5
33.5 75.5
33.5 72.5
29.5 72.5
21.5 87.5
…

CCmaps graphics to do not properly has complex polygons such as lakes. For example Fairfax City is inside of Fairfax county. Fairfax City not plot after Fairfax County and be hidden.

Yes CCmaps really could use a shape file reader or more general boundary files capability.

The Save To line:

Enter the file path and name for control file.   This is a readable text file that CCmaps writes out. It initially contains primarily the file names just entered. Later it contains information you provide such as better variable labels can colors selection.

5.2 Open

The prompts for a CCmaps control file.   I prefix my control files with "ap_" for convenience. I often have data files and boundary files in the same folder and "ap_" usually puts the contol files on top of alphabetically sorted files.

5.3 Save

This save a control file with current slider setting, color setting, variables labels, top text annotation, etc. I often have several control files for the same data set as a make progressive refinement, or want to keep track of particular view.   Another way to keep track of annotated views used the Snapshots menus.

5.4 Exit   Self explanatory

6. Variable Menu

There are four items: picker, labels, tails, and scales.   The variable picker is below. Use this to select the variables to use on the top, bottom and right slider and the region weight variable if there is one.   Equal weights is the last line.

Variables Selection

Col... No.	Variable Name	MouseOver Display	Color Control: Top Slider	Column Control: Bottom Slider	Row Control: Right Slider	Region Weight
1	Region ID	◉				
2	allRate	○	◉	○	○	○
3	popDensity	○	○	○	◉	○
4	pctUrban	○	○	○	○	○
5	pctFemHH	○	○	◉	○	○
6	medinchh	○	○	○	○	○
7	medincfam	○	○	○	○	○
8	percapInc	○	○	○	○	○
9	medincMworker	○	○	○	○	○
10	medincFworker	○	○	○	○	○
11	povAll	○	○	○	○	○
12	povChildren	○	○	○	○	○
13	povGT65	○	○	○	○	○
	Equal Weights					◉

| Col... No.	Variable Name	Mouseover Display	Color Control: Top Slider	Column Control: Bottom Slider	Row Control: Right Slider	Region Weight
1	Region ID	◉	○	○	○	○
2	allRate	○	◉	○	○	○
5	pctFemHH	○	○	◉	○	○
3	popDensity	○	○	○	◉	○
	Equal Weights	○	○	○	○	◉

Cancel   OK   Show Summary by Variable ▼

The Variable Label window below allows entry of better slider labels than the variable names from the data file that are used by default.

The Tails window will below override the CCmaps adjacent value calculation. When the adjacent values are not the minimum and maximum values tails will topically appear as part of the slider and the scale will become piecewise linear. If you don't want tails to appear enter the minimum and maximum values. Sometimes the adjacent values calculation sees wrong or is seem there should be a tail. This provides explicit control to deal with these infrequent occurrences. [Revision should explain Adjacent values]

The Scales window.   Sometimes the appearance of numbers is improved when the variables values are scale by a power of 10 or other factor. This used to work but appear to be broken currently.
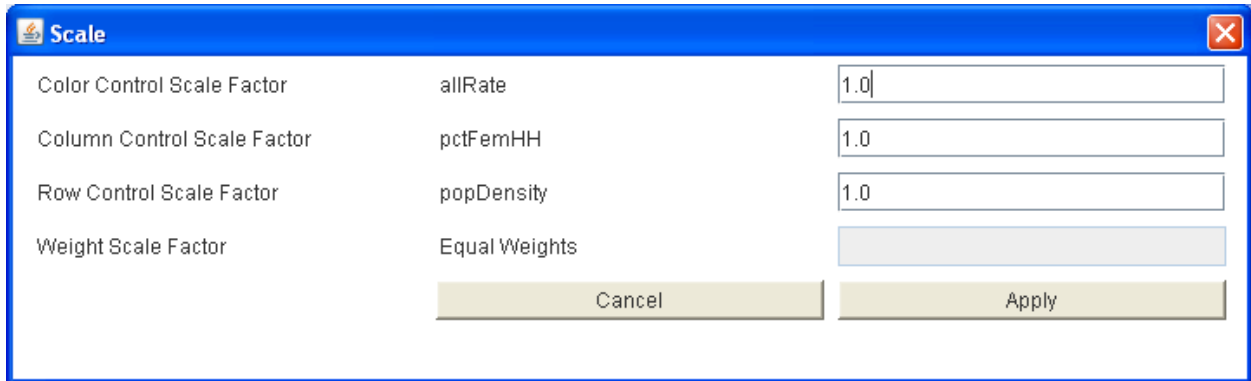
7. **The Edit Menu**

This has four items regions colors, region outline, top text, and map bounds.
As suggest the next page, you can change most of the colors in CCmaps.   Currently is seems best start with the edit menu again to change each color. Remember to save the project file before exiting keep the edits.

The region line color can also be changed similarly.   The thickness change was fully implemented but can be change by hand editing the control file.
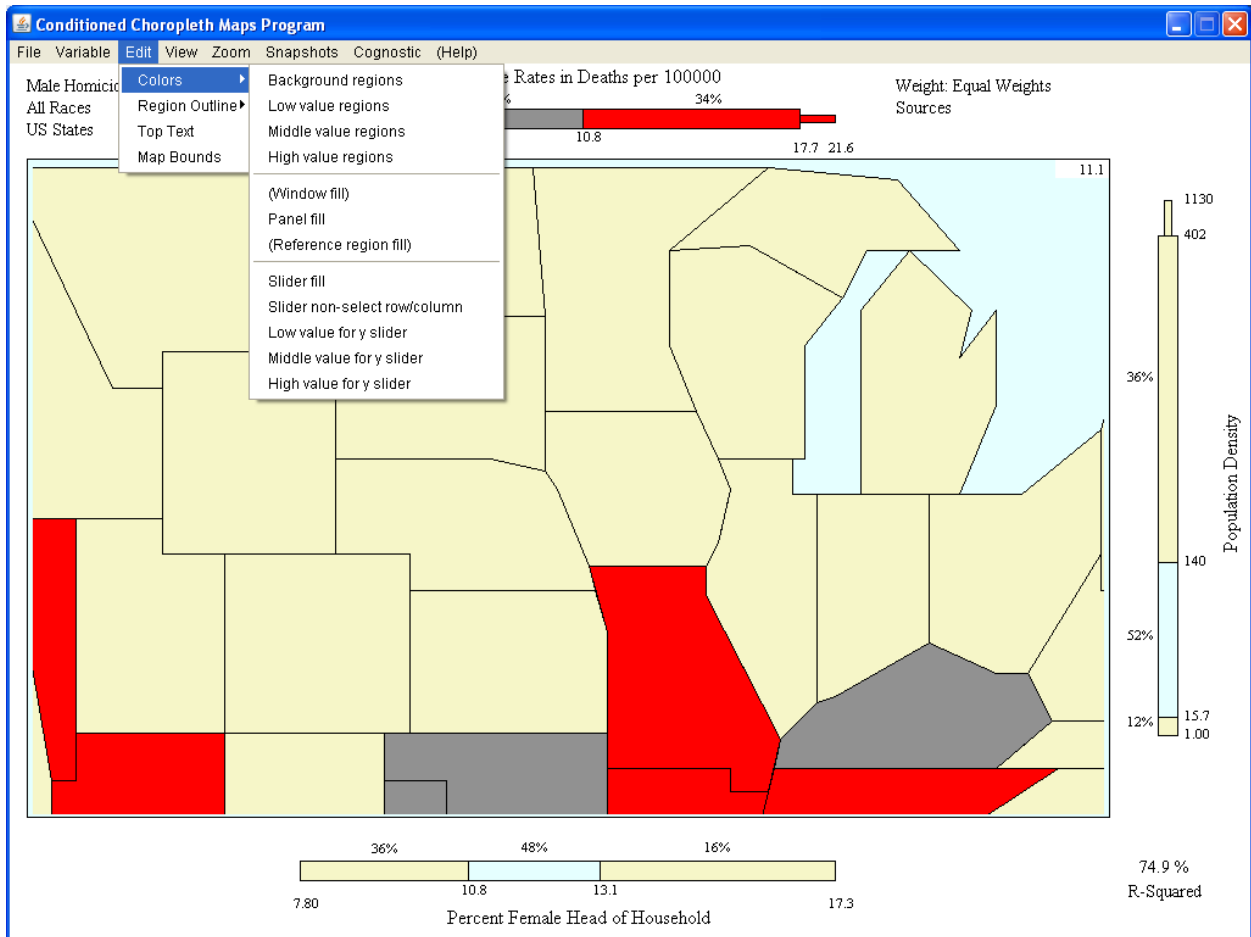
The low value region was select above.
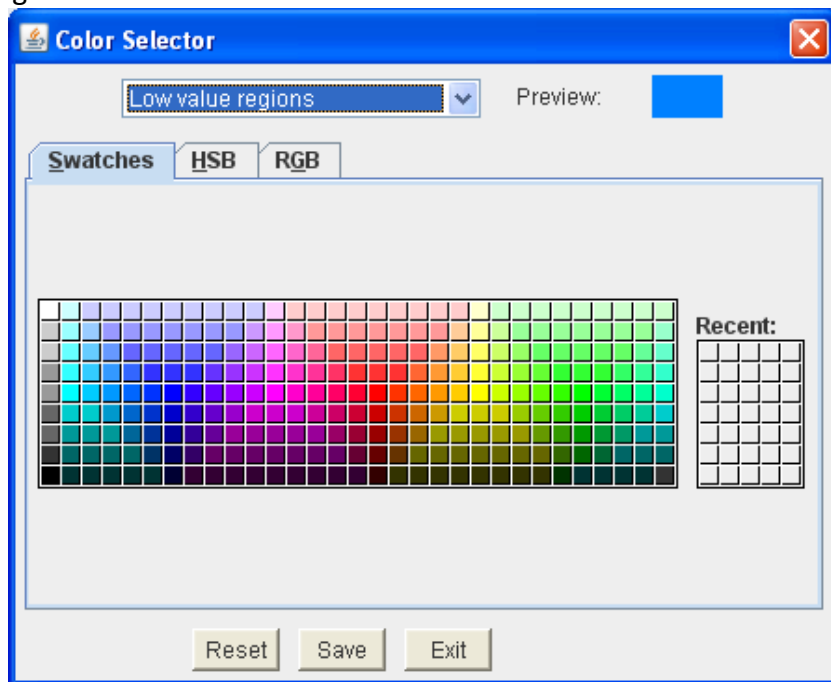
The top text annotation window.

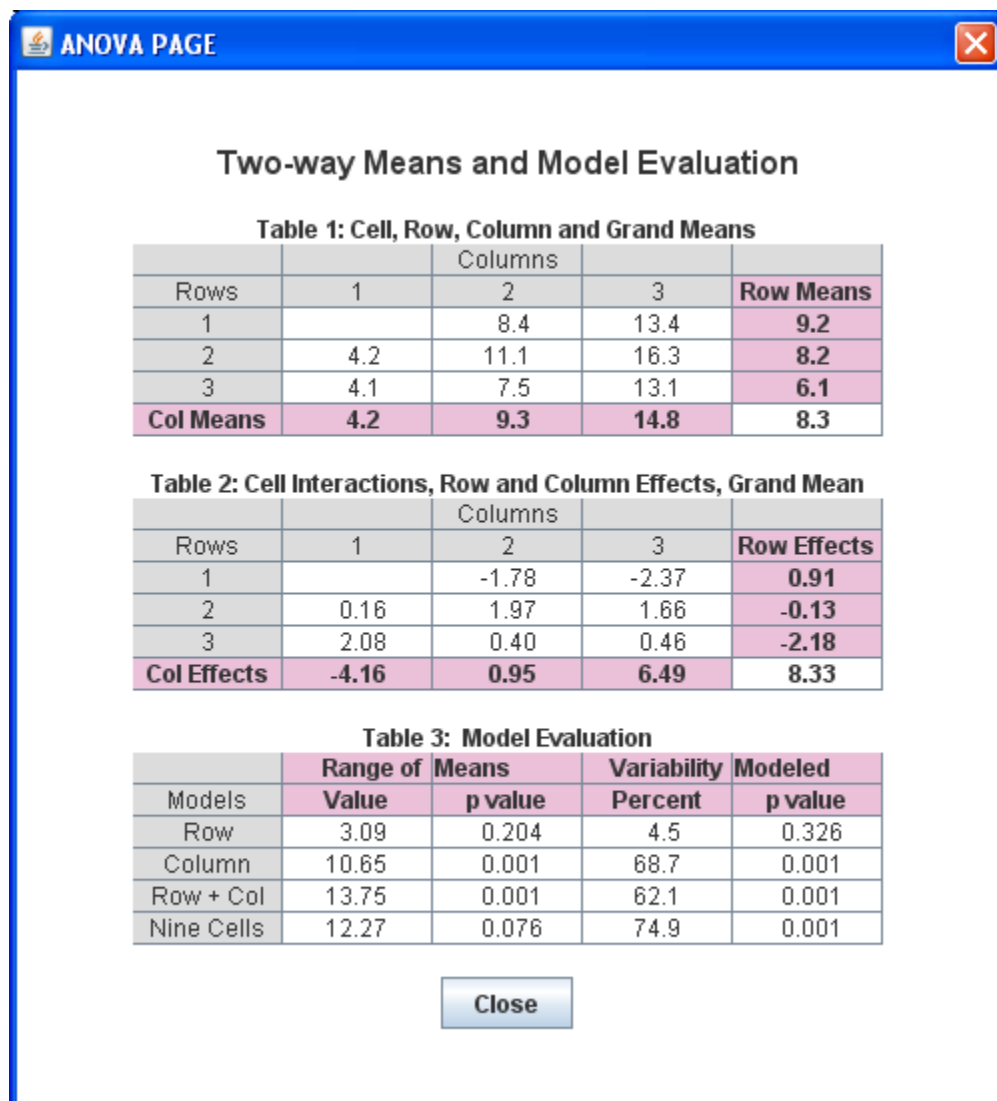| Top Texts	
Left Label 1	Male Homicide Rates 2001-2005
Left Label 2	All Races
Left Label 3	US States
Right Label 1	Weight: Equal Weights
Right Label 2	Sources
Right Label 3	
Cancel	Apply

8. The View Menu

The four main view are maps, table of mean, conditoined qqplots and conditioned scatterplot. There are also toggles to show or high region outlines and mouseovers.
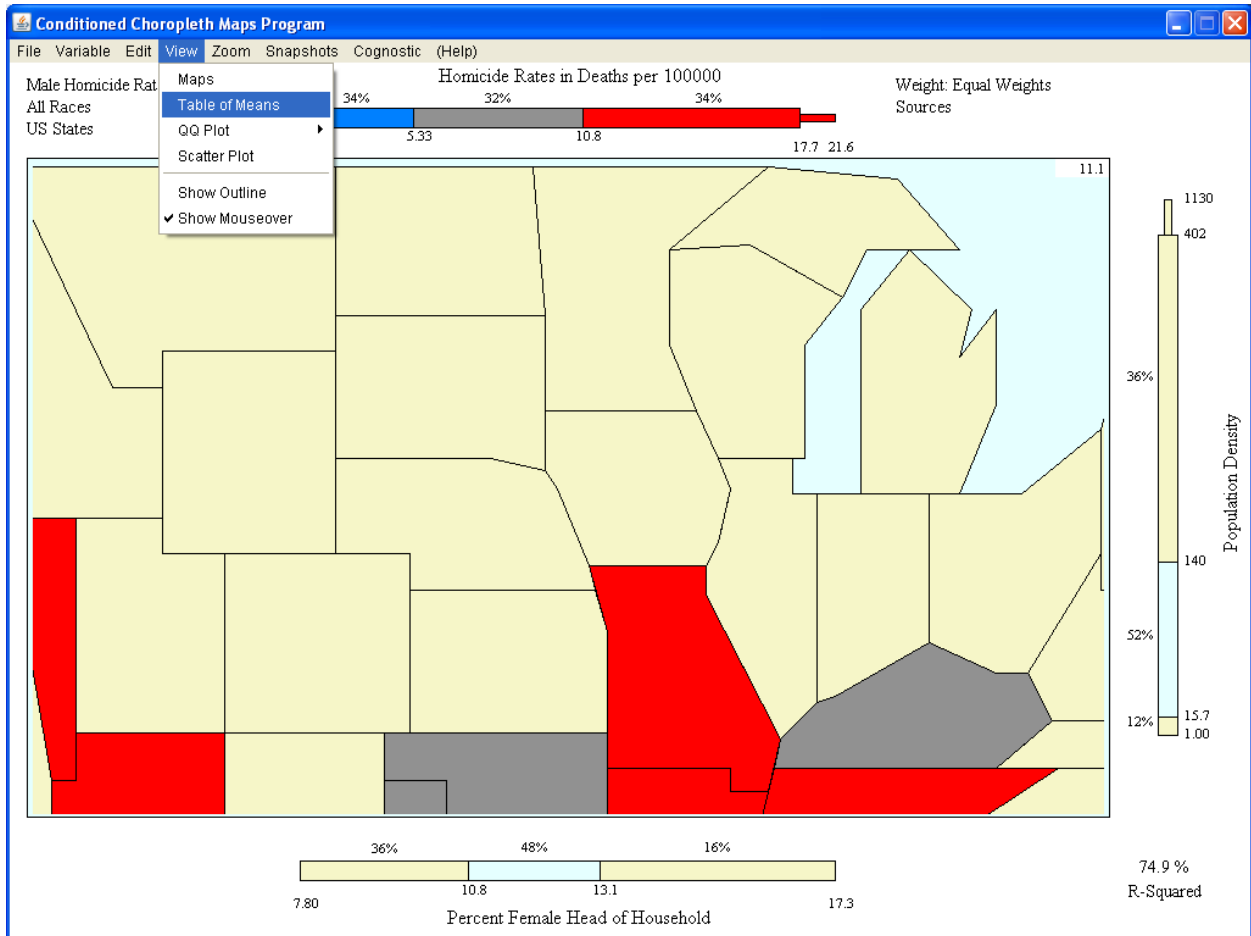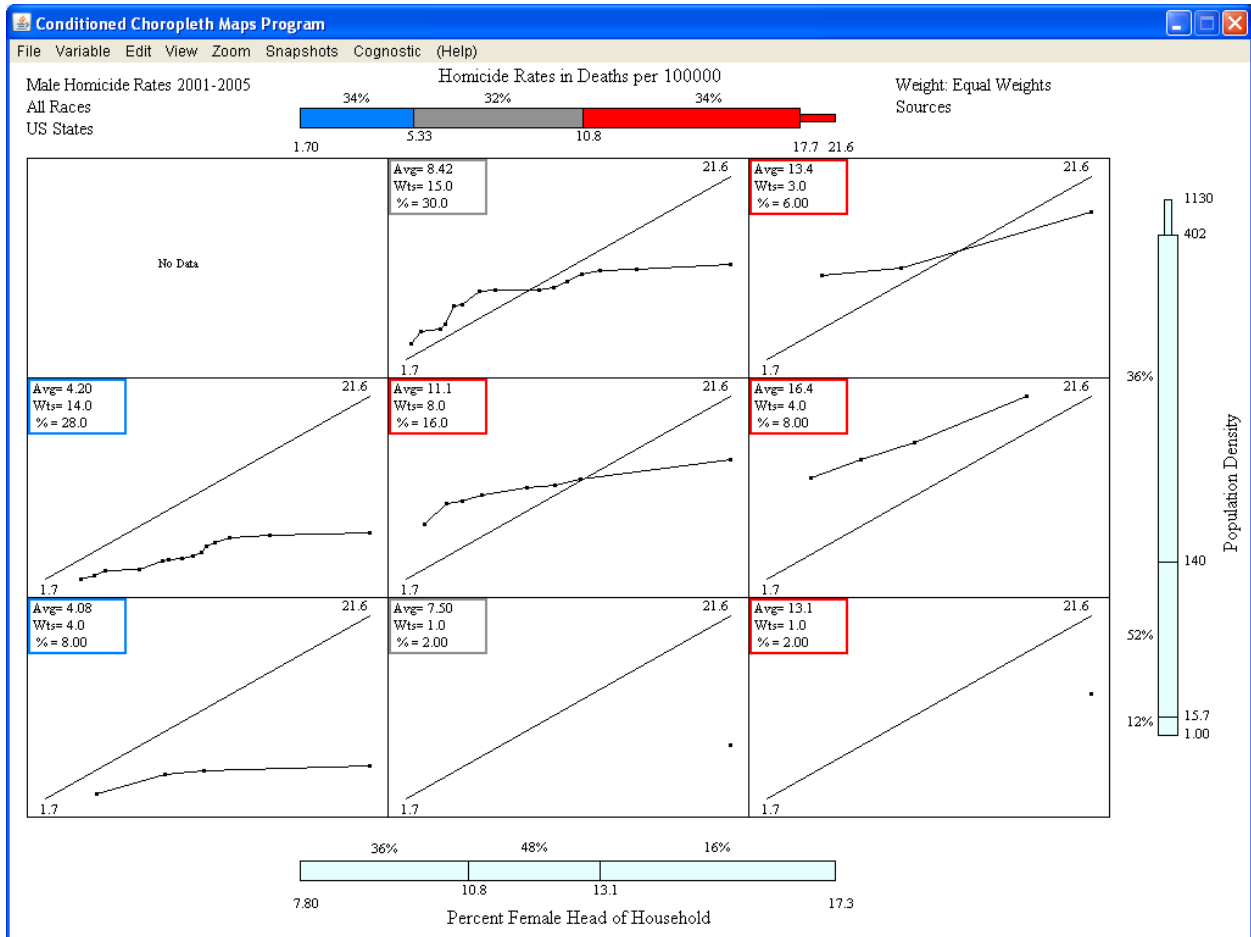
**Two-Way Mean Table, Row and Column Effects, Permutation Tests**

Example Permutation tests break the association with the conditioning variables. **Warning:** this does not account for searching to find a good fit. The proper implementation would search for a good fit using the same algorithm for each randomization.   Also the current version only include 999 permutations. This issue is basically speed and today computer's may enable address both limitations.

ANOVA PAGE

Two-way Means and Model Evaluation

Table 1: Cell, Row, Column and Grand Means

Rows	1	Columns 2	3	Row Means
1		8.4	13.4	9.2
2	4.2	11.1	16.3	8.2
3	4.1	7.5	13.1	6.1
Col Means	4.2	9.3	14.8	8.3

Table 2: Cell Interactions, Row and Column Effects, Grand Mean

Rows	1	Columns 2	3	Row Effects
1		-1.78	-2.37	0.91
2	0.16	1.97	1.66	-0.13
3	2.08	0.40	0.46	-2.18
Col Effects	-4.16	0.95	6.49	8.33

Table 3: Model Evaluation

Models	Range of Means Value	p value	Variability Percent	Modeled p value
Row	3.09	0.204	4.5	0.326
Column	10.65	0.001	68.7	0.001
Row + Col	13.75	0.001	62.1	0.001
Nine Cells	12.27	0.076	74.9	0.001

Close

**Dynamically Conditioned QQplots. Panels can be enlarged.**

Conditioned smoothed scatterplot plots. Panels can be enlarged.
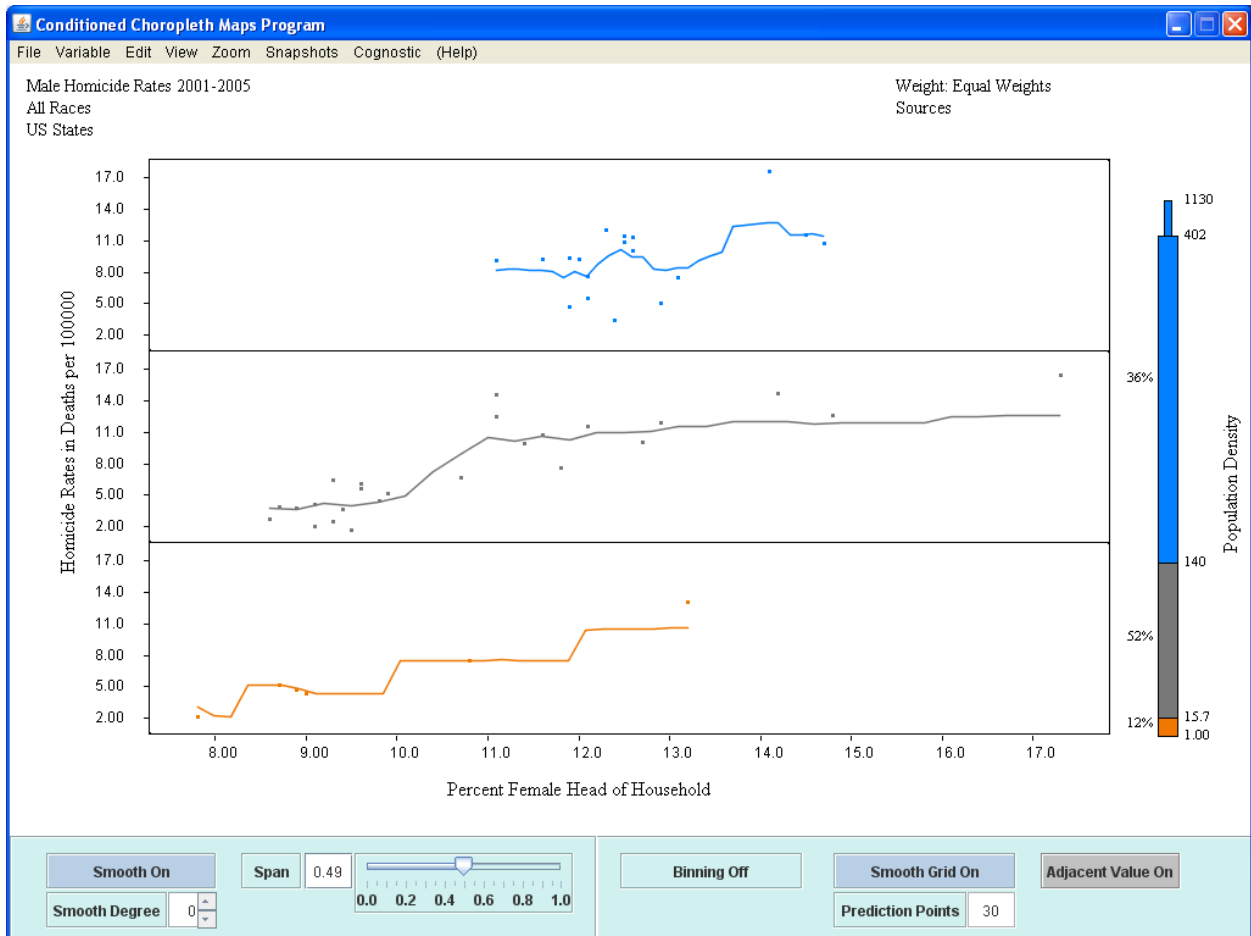
9. The Snapshot Menu.

Snapshots save the state of the program in a special binary file along with annotation. This allow a speaker to jump to different views and proceed immediately with sliders. A student implementation of voice recording (and music) has not be yet included in this version of CCmaps.
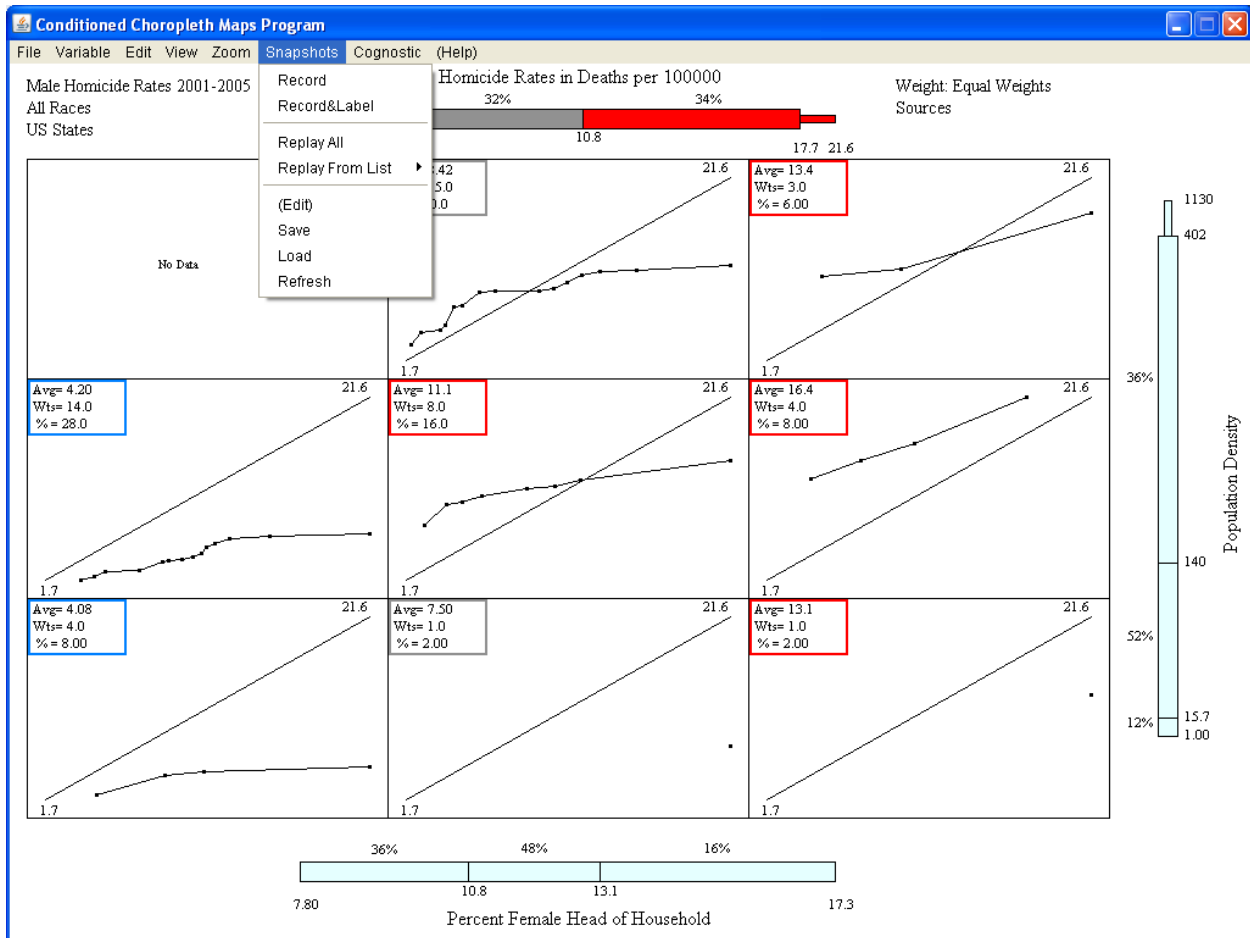
10. The Cognostics Menu

This performs a grid search for high R-Squared Slider Setting. The little window below
Allow selection of the top ten settings.

The grid search in a partial search.   Slight adjustments of the sliders and lead to better values.
A second very local search could be implemented to find best of local values.   There still could
more distance setting with even better values.
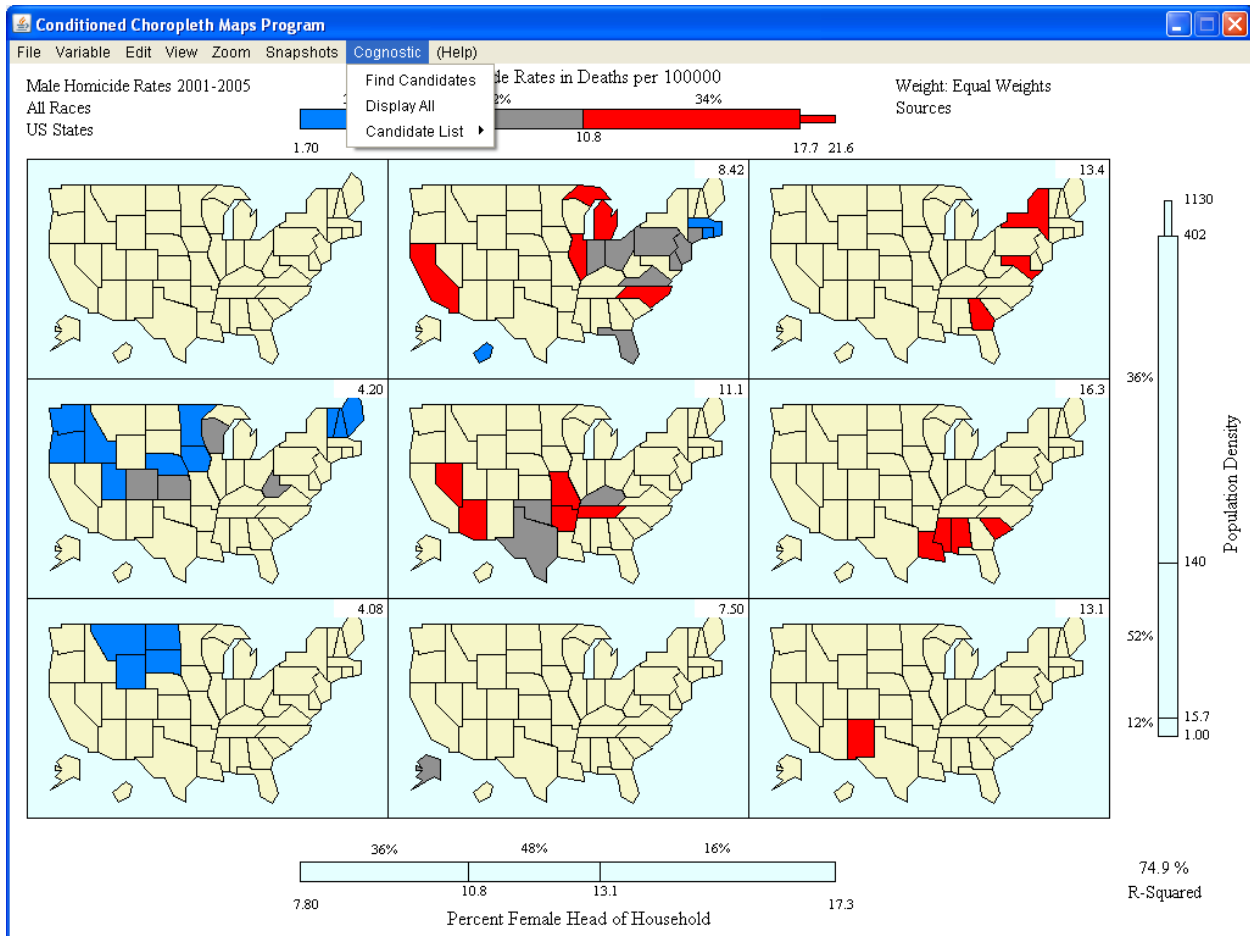
The Display all option brings up this window to try the top 10 items.