SOMVIS:
A Multivariate Mapping and Visualization Tool

User Manual
Version 2.0

Author:
Diansheng Guo

Department of Geography
University of South Carolina

Website:
www.SpatialDataMining.org

October 7, 2008

Launching SOMVIS:

To run this program, it is recommended to have the latest version of Java installed on your machine. You can visit http://java.com/en/download/index.jsp and test if you have the latest version. .

Then you simply start SOMVIS using one of the following options:

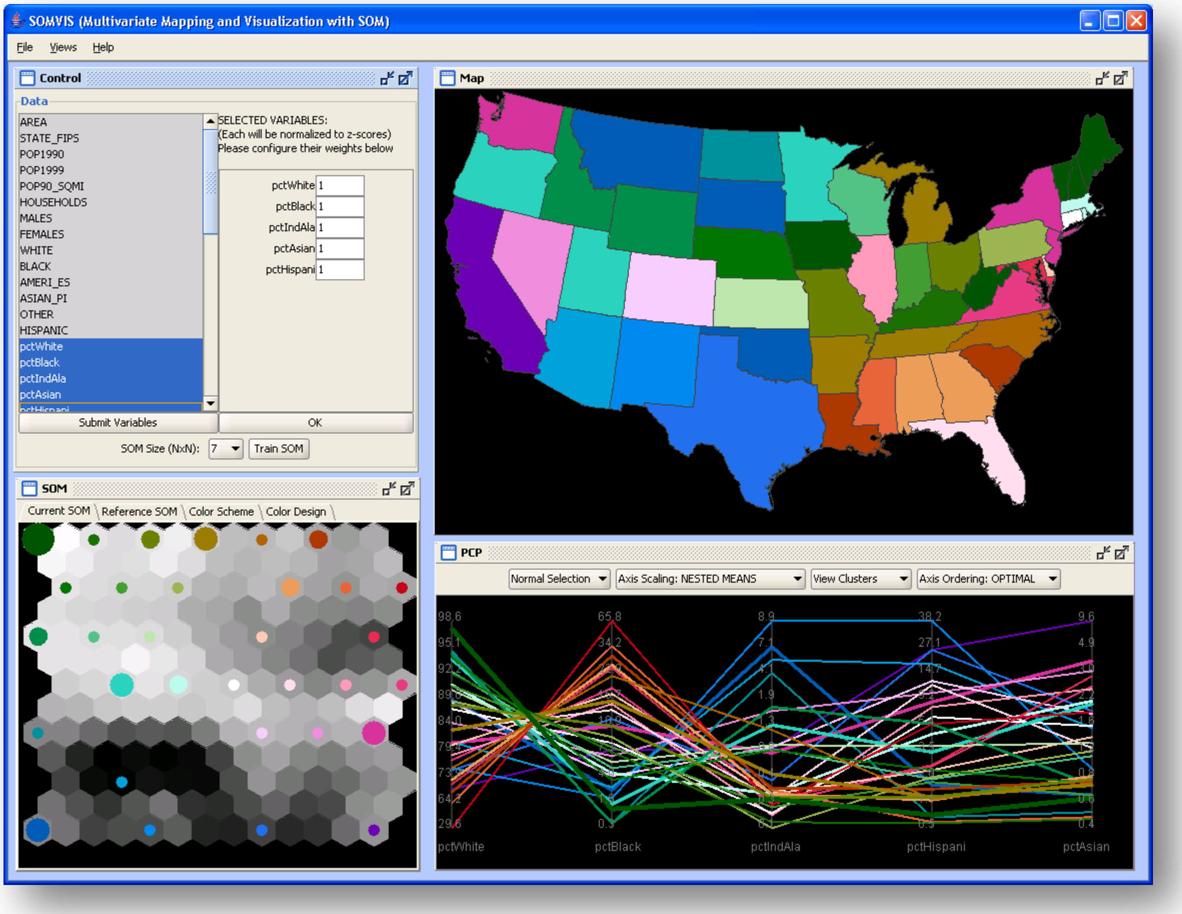
- Double click the file **somvis2.0.jar**; OR
- Right click on **somvis2.0.jar** → open with... → Java …; OR
- Execute in a command window with "**java -jar somvis2.0.jar**". This approach allows you to review output messages when unexpected error occurs.

(If for some reason, your downloading software actually saves the file as "**somvis2.0.ZIP**", you need to rename the file back to "**somvis2.0.JAR**" before running it as instructed above.)

INTRODUCTION

SOMVIS is a software package for the discovery, interpretation, and presentation of multivariate spatial patterns. The software tool integrates computational, visual, and cartographic methods together to detect and visualize **multivariate** spatial patterns. It is able to:

- perform multivariate clustering and abstraction with a Self Organizing Map (SOM);
- encode SOM result with colors derived from a two-dimensional color scheme;
- visualize the multivariate patterns with an enhanced Parallel Coordinate Plot (PCP) display, which serves as a multivariate "legend" in the integrated system;
- visualize spatial variations of multivariate patterns; and
- support human interactions to explore patterns from different perspectives.


SOMVIS (Multivariate Mapping and Visualization with SOM)

SELECTED VARIABLES:
(Each will be normalized to z-scores)
Please configure their weights below

pctWhite 1
pctBlack 1
pctIndAla 1
pctAsian 1
pctHispani 1

SOM Size (NxN): 7

Normal Selection
Axis Scaling: NESTED MEANS
View Clusters
Axis Ordering: OPTIMAL

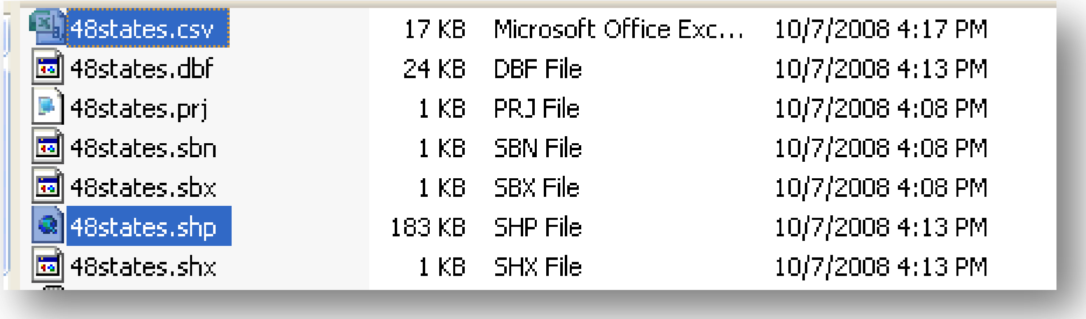
Related Publication:

Guo, D., M. Gahegan, A.M. MacEachren, and B. Zhou. "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach". Cartography and Geographic Information Science,Vol. 32, No. 2, 2005, pp. 113-132.

Guo, D., J. Chen, A. M. MacEachren, and K. Liao (2006), "A Visualization System for Spatio- Temporal and Multivariate Patterns (VIS-STAMP)", IEEE Transactions on Visualization and Computer Graphics, 12(6), pp. 1461-1474.

DATA FORMAT

The easiest way to create a data set for SOMVIS is to take a standard set of ESRI ArcGIS shape file and save its .dbf file as a .csv file. SOMVIS need both the **.shp** file and the **.csv** file. For example, following is the example data set that comes with this manual. The “48states.csv” file was directly converted from the “48states.dbf” file. The two highlighted files are needed for SOMVIS.



48states.csv	17 KB	Microsoft Office Exc...	10/7/2008 4:17 PM
48states.dbf	24 KB	DBF File	10/7/2008 4:13 PM
48states.prj	1 KB	PRJ File	10/7/2008 4:08 PM
48states.sbn	1 KB	SBN File	10/7/2008 4:08 PM
48states.sbx	1 KB	SBX File	10/7/2008 4:08 PM
48states.shp	183 KB	SHP File	10/7/2008 4:13 PM
48states.shx	1 KB	SHX File	10/7/2008 4:13 PM

LOADING DATA

File → Load Data …

Choose either the .csv file or the shape file (the other one will be automatically loaded)

Only the numerical fields (columns) will be used by SOMVIS. Other columns will be ignored.

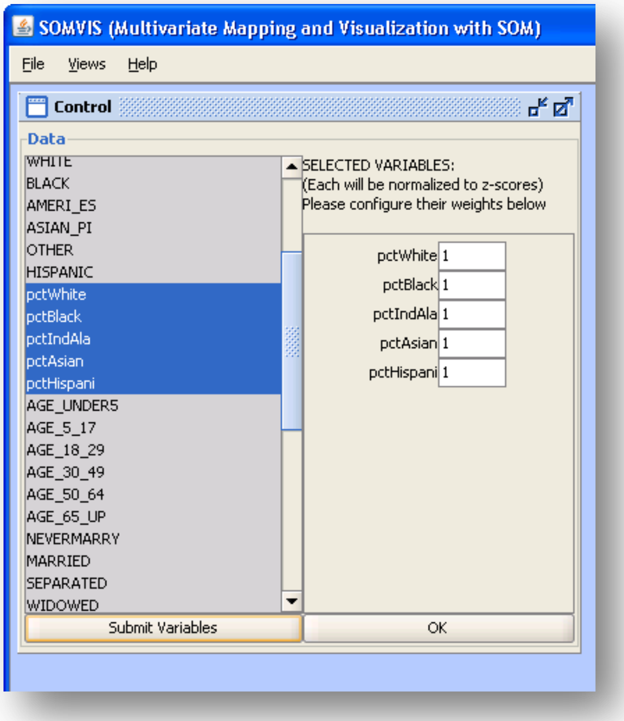
Select multiple variables from the list

Click “Submit Variables”

The values of each variables will then be normalized to Z-scores (i.e., zero mean and unit variance).

You may give each variable a different weight (which will be multiplied to the z-score).

Click “OK”

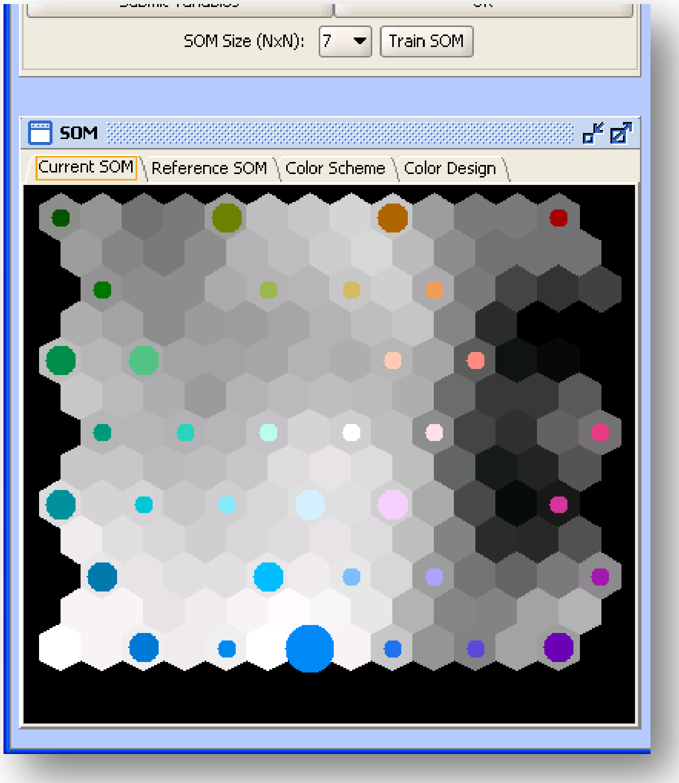


WHITE
BLACK
AMERI_ES
ASIAN_PI
OTHER
HISPANIC
pctWhite
pctBlack
pctIndAla
pctAsian
pctHispani
AGE_UNDER5
AGE_5_17
AGE_18_29
AGE_30_49
AGE_50_64
AGE_65_UP
NEVERMARRY
MARRIED
SEPARATED
WIDOWED

SELECTED VARIABLES:
(Each will be normalized to z-scores)
Please configure their weights below

pctWhite	1
pctBlack	1
pctIndAla	1
pctAsian	1
pctHispani	1

SELF-ORGANIZING MAP CLUSTEREING



SOM Size (NxN): 7

The normalized (and optionally weighted) multivariate data is processed by a Self-Organizing Map (SOM) to derive clusters of spatial objects based on their multivariate similarity. The SOM uses the Euclidean distance to assess multivariate similarity between spatial objects.

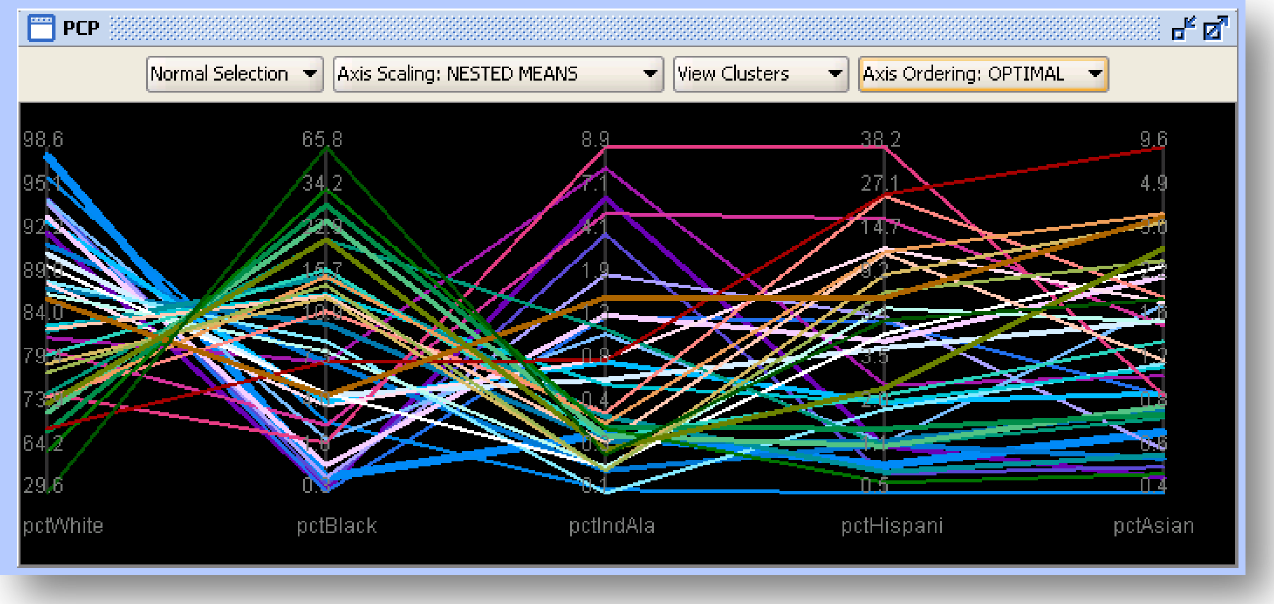
The default size for SOM is 7x7 (i.e., 49 clusters). Depending on the data set size, one many choose a smaller or larger size. Then click "Train SOM". The SOM clustering result is visualized as depicted in the snapshot to the right, which uses two different types of hexagons: (1) *node hexagons*, each of which contains a circle that is scaled to depict the number of data items in the node (cluster); and (2) *distance hexagons*, each of which is shaded to represent the multivariate dissimilarity between two neighboring nodes (i.e., two codebook vectors). This kind of graphic display is called the U-matrix. For methodological details about SOM, readers are referred to book:

Kohonen, T. (2001). Self-organizing maps, Berlin ; New York : Springer.

A data item is assigned to a node (cluster) if that node's codebook vector is the closest to the data item. A node can have more than one data item assigned or may have no data item assigned (in which case it is an empty node). The area of the circles inside node hexagons represents the number of items contained in each node. Each circle is filled with a color from a 2D color scheme so that similar clusters (i.e., nearby nodes) have similar colors. See Guo et al. 2005 for detailed explanation.

The user can rotate or flip the 2D color scheme in case a certain corner is desired to have a certain color. These functions are enabled at the "Color Scheme" tab. The user may also changed the 2D color design by going to the "Color Design" tab. See Guo et al. 2005 for detailed explanation on these functionalities.

MULTIVARIATE VISUALIZATION WITH PARALLEL COORDINATE PLOT (PCP)


PCP

Normal Selection

Axis Scaling: NESTED MEANS

View Clusters

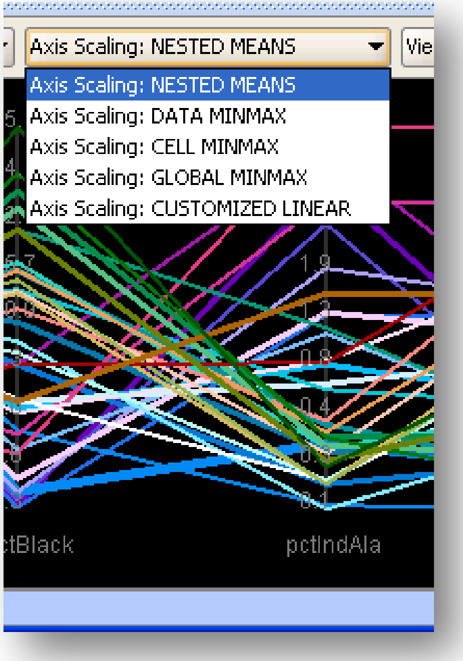
Axis Ordering: OPTIMAL

The PCP supports two detail levels:

- **View Clusters:** Each string represents a cluster with its mean vector. The thickness of each string represents the cluster size (i.e., the number of data items in the cluster).
- **View Data Items**: Each string represents a data item with its multivariate vector.
- For either choice, each cluster (or data item) has the same color as it does in the SOM view.

The PCP supports five axis scaling methods:

- **Nested-Means:** scaling on each axis using nested means and thus adjust the spacing of intervals according to data distribution. This method can alleviate the overlapping problem in PCP for skewed data distribution. Specifically, nested-means is a *non-linear scaling* method that recursively calculates a number of mean values (and sub-means) and uses these values as break points to divide each axis into equal-length segments. Therefore, nested-means scaling always puts the mean value at the center of each axis and thus makes axes defined by different units and data ranges comparable.


Axis Scaling: NESTED MEANS

Vie
Axis Scaling: NESTED MEANS
Axis Scaling: DATA MINMAX
Axis Scaling: CELL MINMAX
Axis Scaling: GLOBAL MINMAX
Axis Scaling: CUSTOMIZED LINEAR
tBlack
pctIndAla

- **Data Min-Max:** each axis is linearly scaled using its min and max values.
- **Cluster Min-Max** (or Cell Min-Max): each axis is linearly scaled using cluster centroid min and max values.
- **Global Min-Max:** this option is only useful when all the variables are comparable to each other, for example percentage values. Axes will be scaled linearly using the global min and max values (for all variables).
- **Customized Linear:** this option is only useful when all the variables are comparable to each other, for example percentage values. The user will define the min and max (same for all variables) to linearly scale each axis. In future versions, the user may be able to define the min/max differently for each axis.

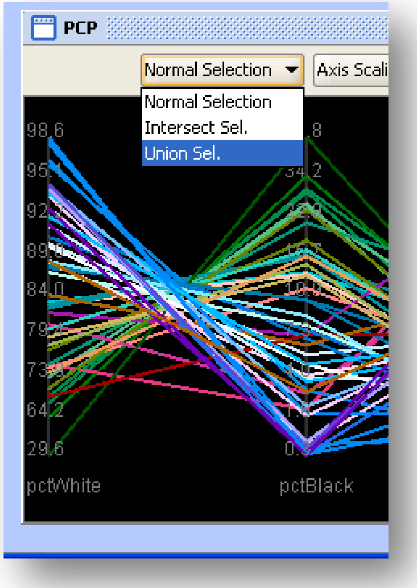
The PCP can *optimally* order dimensions

- **Optimal Ordering of Axes:** dimensions are ordered using an optimal hierarchical ordering method based on the mutual correlations among dimensions

  Guo, D. and M. Gahegan (2006). "Spatial ordering and encoding for geographic data mining and visualization." Journal of Intelligent Information Systems **27**(3): 243-266.

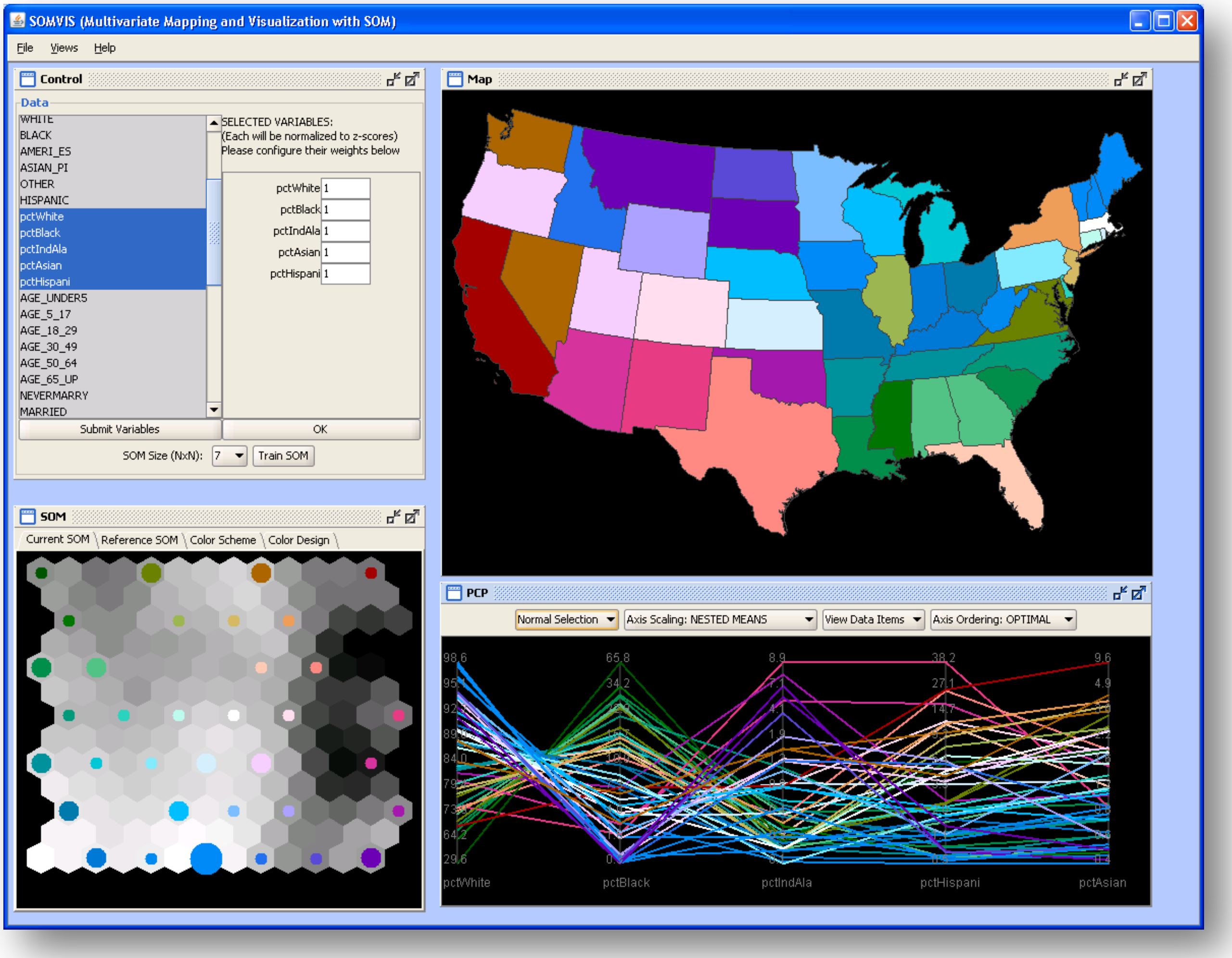
- **Original Ordering of Axes:** dimensions are in their original order as in the data file.

The PCP supports different types of selection at different levels



- The PCP at the cluster level presents a global view of the overall patterns. A user can select one or more clusters in the PCP (or in the SOM), then switch to the data item level (instead of the cluster level), and examine all the data items in the cluster(s).
- Selection can also be made at the data item level. For example, one can show data at the item level in the PCP and then select a single data item to read its exact variable values. One can also switch back to the cluster level and see which cluster the selected item belongs to. If that cluster also contains other items, its circle will become a wedge to show the partial selection.
- By selection "Intersect Sel." or "Union Sel.", the user may also combine two different selections or select within a selection.

MULTIVARIATE MAPPING



SELECTED VARIABLES:
(Each will be normalized to z-scores)
Please configure their weights below

pctWhite 1
pctBlack 1
pctIndAla 1
pctAsian 1
pctHispani 1

SOM Size (NxN): 7

Normal Selection
Axis Scaling: NESTED MEANS
View Data Items
Axis Ordering: OPTIMAL

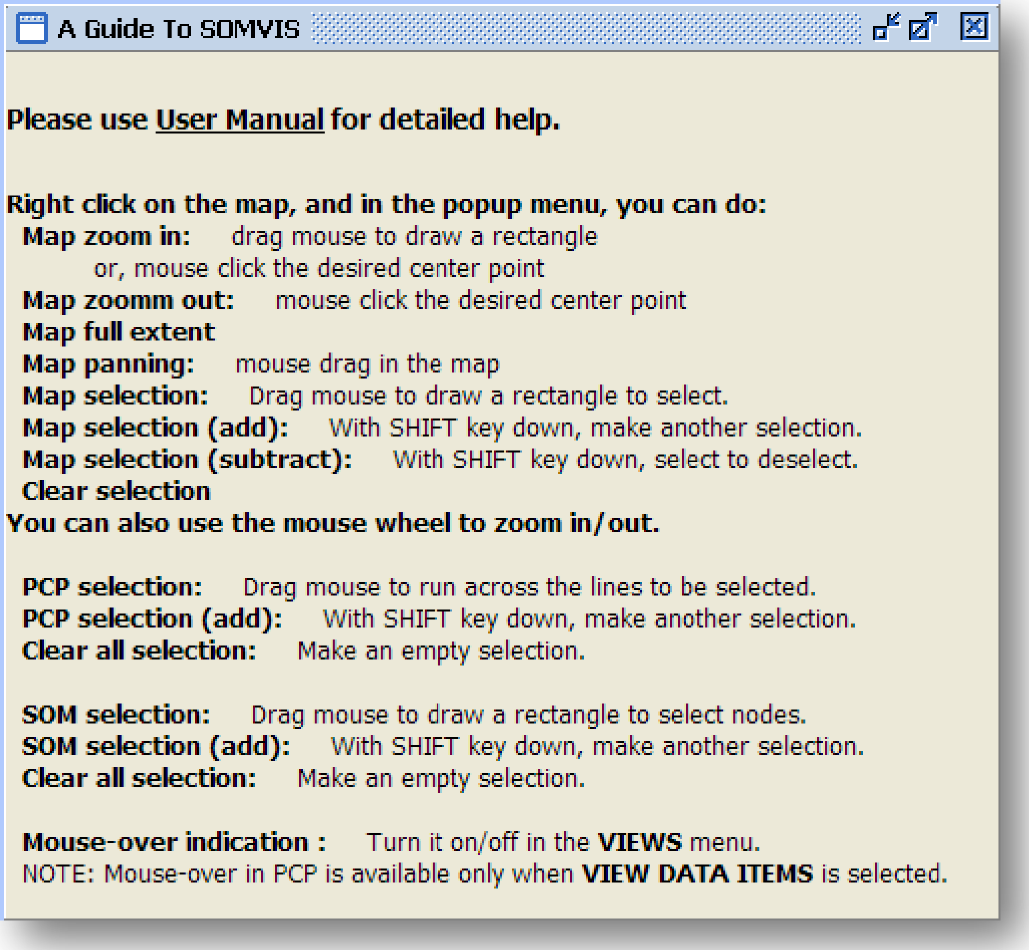
To examine multivariate spatial patterns in the geographic context, mapping is indispensable. The colored SOM result make it possible for making a "multivariate choropleth" map, where each data item (i.e., a spatial object such a U.S. state) is assigned a color based on the SOM node that contains the item. From a thematic mapping perspective, the SOM serves as a multivariate classification method and the PCP serves as the legend. The resulting map is a holistic view of the spatial distribution of multivariate patterns.

An advantage of this integrated approach is that, even without human interaction (e.g., brushing and focusing), we can still perceive a holistic view of the major patterns by looking at only three views (i.e., SOM, PCP, and MAP). The default background for the three views is black, which the user can change (**Views → Change Views' Background …**). Since colors are used for multivariate mapping, it is suggested to set the background with a grayscale tone (e.g., black, white, or gray).

INTERACTIVE USER EXPLORATION AND INTERPRETATION

In addition to constructing a holistic view of patterns, SOMVIS also supports a variety of user interactions that allow the analyst to examine patterns in detail. Each view is able to support user selections and the selection made in one view will be highlighted in other views. The user can make a selection in one component and then refine that selection in the same or another visual component by adding or subtracting new selection(s). Each component should be able to respond to selections made at different levels (i.e., data elements or clusters).

To facilitate efficient user interaction, a set of short keys are supported (see below for details). These shortcut keys can be looked up at **Help → Help Contents**.



A Guide To SOMVIS

Please use User Manual for detailed help.

Right click on the map, and in the popup menu, you can do:

Map zoom in: drag mouse to draw a rectangle

or, mouse click the desired center point

Map zoomm out: mouse click the desired center point

Map full extent

Map panning: mouse drag in the map

Map selection: Drag mouse to draw a rectangle to select.

Map selection (add): With SHIFT key down, make another selection.

Map selection (subtract): With SHIFT key down, select to deselect.

Clear selection

You can also use the mouse wheel to zoom in/out.

PCP selection: Drag mouse to run across the lines to be selected.

PCP selection (add): With SHIFT key down, make another selection.

Clear all selection: Make an empty selection.

SOM selection: Drag mouse to draw a rectangle to select nodes.

SOM selection (add): With SHIFT key down, make another selection.

Clear all selection: Make an empty selection.

Mouse-over indication : Turn it on/off in the **VIEWS** menu.

NOTE: Mouse-over in PCP is available only when **VIEW DATA ITEMS** is selected.

ACKNOWLEDGEMENTS

This software took was initially started and supported, in part, by NSF grant #9983445, NSF grant #EIA-9983451, grant # TS 1125, ATPM/CDC and grant CA95949 from the National Cancer Institute (NCI).

Thanks to Dr. Masahiro Takatsuka (masa@vislab.net) for providing the initial SOM implementation and Dr. Frank Hardisty (hardisty@sc.edu) for providing the component for automatically coordinating events and listeners.

This package includes a free java library from JGoodies (www.jgoodies.com) to improve the look and feel of user interface.