

hvpcp.apr user's guide: set up and tour

by Rob Edsall

HVPCP (HealthVis-ParallelCoordinatePlot) is a visualization environment that serves as a follow-up to HealthVis (produced by Dan Haug and Alan MacEachren at Penn State) and HealthVis II (which included a parallel coordinate representation, developed by Rob Edsall), adding the important ability of examining health and related data on three scales, rather than just one. Those three scales are county-level, state-level, and Health-Service-Area-level. Also included as added features are easier-to-use GUI tools, with the most frequently used operations now available in dialog boxes near the appropriate display (rather than menus or in the tool or button bar of the overall GIS interface).

Requirements:

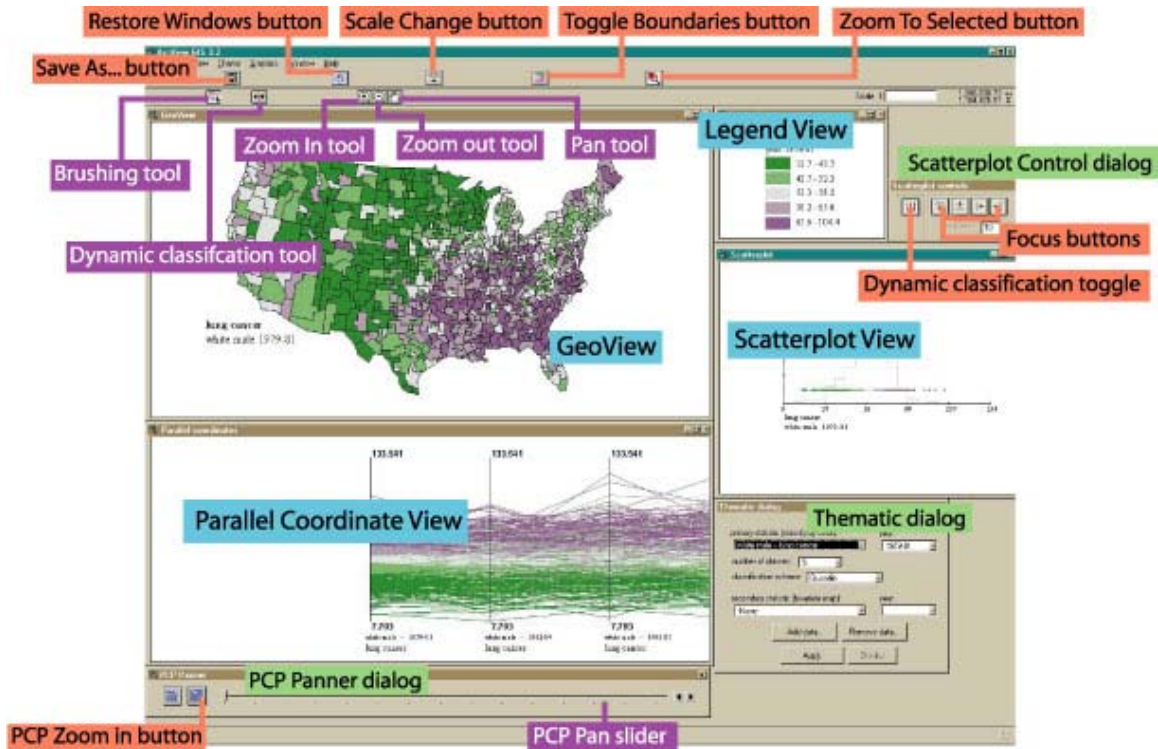
- ArcView 3.x
- 1280 x 1024 pixel 16-bit display (or better; other settings work, but the system was designed for this resolution)
- about 5MB of hard disk space accessible on the root directory of the c: drive

Setting up

Upon downloading zip file, unzip the file to root directory of the c: drive of the PC you are working on (such that you create a c:\hvpcp\ directory on your hard drive). Doing this (as opposed to copying it to some other directory or drive) will allow you to bypass time-consuming set-up steps later.

Once you've copied the health folder to your hard drive, open ArcView as normal. When asked, select "Open an existing project," and click OK. Navigate to the c:\hvpcp\ directory and open hvpcp.apr. In a few seconds, an interface like the one below will appear.

Note: It is recommended that you immediately save this open project under a different name (using File/Save Project As... or the Save As... button). If you simply Save the project (using the File/Save Project, for example), it will overwrite hvpcp.apr with any modifications you've made to it during your work session.



A quick tour

Loading sample data

As shipped, HVPCP has a sample data set (lung cancer rates by three-year averages for health service areas) pre-loaded. In this section, you will be introduced to the process of loading attribute data into HVPCP.

First, select a scale of analysis (county, state, or HSA) by clicking on the scale change button. For this tour, select Health Service Area and click Apply.

In the Thematic dialog, click Add data... Navigate to the c:\hvpcp\support\data directory. Since you have chosen HSA as your scale of analysis, only those data files whose names end in hsa will be visible. You can load files of type .dbf, .txt, or INFO (by selecting a List files of type pull-down menu option). Select one of the files, taking note of the name.

You may be prompted to create a “header file” for this sample data set.

Creating header files. The header file is metadata (data about the data) that is stored and accessed separately from the data set you are loading. In some cases for the sample data, a header file already exists. In other cases, you will have to create the header file from scratch. If the header file already exists, you will be asked to inspect the existing header file information and confirm that it is correct. If you find that it is not correct, you will be asked to create a new header file from scratch.

Creating a header file requires some degree of knowledge about the data set you are loading. It is assumed that you know that the real-world meaning of the column headings in the table (e.g. that the column Prate1 represents “prostate cancer, white males, 1979-81.”) The names you select in the creation of the header files will be those shown in the representations (the map, scatterplot axis labels, etc.).

To create a header file, you are first prompted to select a join field (usually the unique identifier field, like hsa, county_fips, etc.). Once you’ve selected a join field, you will be prompted to enter information about each of the columns of data (also known as fields) in the data set. Enter information only on those fields that are relevant to your research (in many cases, for example, the first several columns are indexes; the relevant data – the statistics themselves – are not found until the third or fourth column). If you are not interested in making available the field shown, click Cancel for the Metadata dialog. When you come to relevant attribute information, you are asked to enter the demographic group (e.g. “white males”), statistic (“prostate cancer”), and year of observation (“1979-81”). Once each text box has a value (if irrelevant, type in a dash or hyphen – don’t leave it blank), click OK to proceed to the next field.

After all the metadata has been entered, you will be prompted to check your work for errors. You may choose to recreate the header file if there are errors, or deem the metadata correct, and the data will load normally, creating a scatter plot and map of the first column of data in the newly loaded data set, and parallel coordinate plot (see description below), the axes of which are, by default, ordered as found in the data table.

Loading your own data

ArcView (and thus HVPCP) accepts attribute files (like tables of health statistics or covariates) if organized into observations by county, HSA, or state. The files must be comma- or tab-delimited .dbf, INFO, or .txt files. Each column must represent a variable (like cancer rates for a certain cancer for specific year) and each row must represent an enumeration unit (like county, HSA, or state). Each enumeration unit must be identified with a unique identifier (like fips code or HSA id number). Each column needs to have a string header (like Prate1, or ID, etc) *without spaces*. Example acceptable data file formats can be found by examining the data files in c:/health/support/data.

Naming convention. To be “visible” to HVPCP, the name of each data file must end in either “county,” “hsa,” or “state.” For example, a data file with lung cancer rates for black males might be named bmlunghsa.txt (with the hsa tail of the name a required element). If necessary, rename files you plan to import using this convention.

To load a file, click the Add Data... button on the Thematic Dialog box, and proceed as described above.

The four views

The GeoView. Designed to allow the examination of spatial relationships, the GeoView is a choropleth map with enumeration units of any of the three scales (county, state, HSA). The units are colored according to a customizable classification (described below): at this stage in development, the color scheme is limited to a purple-to-green hue transition scheme for a univariate (one-variable) map, or a four-class blue-gray saturation-value progression scheme for a bivariate (two-variable) map. The GeoView can be modified by:

- **Zooming:** using the zoom in/out tools on the main GUI tool bar, the GeoView can be magnified in or out by either (a) clicking on a point of interest with the tool (zooming in to 125% or out to 50% and centering the view on the point clicked), or (b) click-dragging a box around a region of interest.
- **Toggling the boundaries:** using the boundary toggle button, thin black boundaries outlining the enumeration units on the GeoView can be turned on or off.
- **Restoring the original view:** using the Restore Windows button, the GeoView (and all other Views) can be restored to their original zoom extents and positions on the screen.
- **Adjusting mapped attributes:** using the Thematic Dialog, you can adjust the fields mapped in the GeoView. The various demographic-statistic combinations (as reported by you in the header file you created above) are available, along with their relevant dates. You can adjust the classification schemes (those automated are Quantile, Equal Interval, and Natural Breaks; you can adjust the classifications manually with the scatterplot controls, described in the scatterplot section below) and number of classes using the appropriate Thematic dialog pull-down menus. You can create a four-class bivariate map by selecting a second demographic-statistic-year combination as a secondary statistic. After each adjustment, hit Apply, and the adjustments will be reflected in the updated Views (all views may be updated).

The Scatterplot. The scatterplot shows the range and distribution of one variable and the range, distribution, and relationship between two variables. The points of the scatterplot each correspond to an enumeration unit on the map, and are colored to correspond to the color scheme of the map. You can examine this correspondence by using the **brushing** feature of HVPCP.

- **Brushing the Views.** Select the brushing tool. Move your cursor over the map and click-drag a small region. All of the enumeration units on the map that intersect the box you've drawn are highlighted in red. In addition, all of those observations are also highlighted in their respective locations in *attribute space* (the "space" of the scatterplot; as opposed to physical space, like the north-south-east-west of the GeoView) are also highlighted. The corresponding observations are also highlighted in the Parallel Coordinate Plot View (see below).

In addition, you may brush, in a similar manner, the points of the scatterplot (or, as described below, the lines of the parallel coordinate plot).

Also supported is the "click" type of brushing, by which you may select one observation by clicking (and not dragging) on one observation or enumeration unit.

Also in the scatterplot is a faint histogram representation, to reinforce the visualization of the variable distribution in attribute space.

The scatterplot can be adjusted in similar ways to those described above for the GeoView (zooming, restoring the view; adjusting mapped attributes). In addition, the scatterplot can be used to interactively classify the observations.

- **Zooming.** Zooming into the scatterplot can help examine details of the distributions. Upon zooming, the value represented by the left and right side (and top and bottom, in a bivariate representation) of the current display extent is labeled. The axis labels and values "follow" you in as you zoom, giving you orientation in the attribute space.

- **Dynamic classification.** In a univariate representation (where -None- is showing as a secondary statistic), the classification of a scatterplot may be made dynamic by pressing the dynamic classify toggle. Upon doing this, vertical black line(s) appear in the scatterplot in the locations of the class breaks. In addition, the dynamic classify tool (on the main GUI tool bar) is automatically selected. By moving the cursor *directly* over one of these lines and clicking the left mouse button, you can select that class break line and move it to a different place in the distribution (perhaps to an apparent natural break, or to a value of particular significance, like a national mean). Upon release of the class break line, the scatterplot, map, PCP, and legend will all update to reflect the change. In a bivariate representation, clicking the dynamic classify toggle allows you to adjust the class break for either variable (as above).

In addition, HVPCP allows **focusing** by percentile. In a univariate two-class map, the class break can be moved by clicking the focus left/right buttons. On a bivariate map, either variable may be focused by clicking on the focus left/right buttons (for that field on the x-axis) or the focus up/down buttons (for the field on the y-axis). By clicking a focus button in the scatterplot control dialog, you can automatically move the class break to a selected percentile position (i.e. such that, say, the lowest 65% of the observations in that variable are colored with a light value). The percentile value is shown in the Legend View. You can alter the focus step such that with each click of the focus button, the class break line moves by a customized percentile amount. In so doing, you would be able, for example, to examine only those observations that are in the top 5% of both variables.

The Parallel Coordinate Plot View. The parallel coordinate plot (PCP) shows each observation as a series of unbroken line segments, passing through each of a series of parallel (vertical) axes at a point which represents that observation's relative place among the range of values for the variable represented by that axis. The variables are labeled along the bottom of the view, beneath each axis. Each (semi-horizontal) line, representing one observation, is colored using the same method as that employed in the other Views. For example, a dark purple line in the PCP corresponds to a dark purple enumeration unit in the GeoView, and a dark purple point in the scatterplot.

Like the scatterplot, the use of the zoom tool in the PCP results in the labeling of the values along an axis at the top and bottom of the display. Using the resize window button will reset the display extent and positioning of the PCP View as well as all of the other Views.

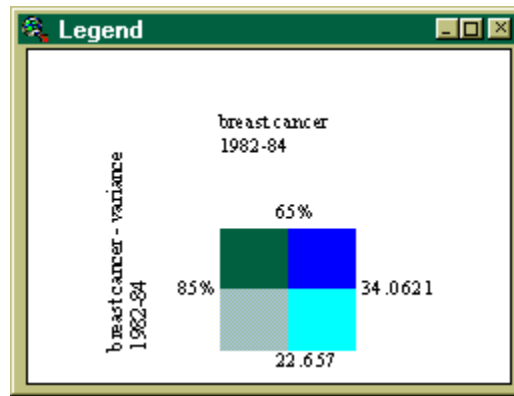
The PCP can be manipulated in four unique ways:

- **Panning along the observations.** The extent of the PCP can be seen by sliding the slider bar known as the PCP panner all the way to the right and left. You will see the range of demographic-statistic-times (all of the combinations possible in the data sets loaded and available to select in the Thematic dialog) represented as black vertical axes. You can also use the stepper tool (the left and right arrows on the right end of the panner) to fine-tune your panning.
- **Zooming out.** You can zoom out to the default extent of the PCP without altering the other Views (which happens when you use the resize window button) by clicking the PCP Zoom In button.

- Reordering the axes.** An axis may be moved a small distance at a time. By doing so, you will see different relationships and can compare variables of interest. Select the dynamic classify toggle on the Scatterplot Control Dialog. Upon moving the cursor over the PCP view, line the crosshairs directly over the axis you'd like to move. Click, and the axis should be selected, indicated by black boxes on its top and bottom. Click the axis again, and drag (with the mouse button down) the axis to another location in the plot, and release¹. Dragging it between two existing axes will reorder the axes, inserting the moved axis between the destination axes.

The Legend. The fourth View is a straightforward legend representation of the colors used on the other three Views. Though the same set of tools (panning, zooming, brushing etc.) are usable on the Legend View, such interaction tends to confuse rather than facilitate any analysis and is not recommended.

The bivariate legend is perhaps not as intuitive as the univariate version. In the example below, the four colors in the map, scatterplot, and PCP are represented in the Legend view with the following interpretation:



The two variables represented are “breast cancer, 1982-84” and “breast cancer variance, 1982-84.” On the three other views, all observations (enumeration units) with breast cancer rates of over 22.657 are colored either royal (saturated dark) blue (upper right) or bright (saturated light) blue (lower right). The rate 22.657 represents the 65th percentile for the years 1982-84; 65% of the observations are below this value, while 35% are above this value. The secondary statistic, represented on the vertical, is “breast cancer variance, 1982-84.” The class break line for this statistic is 34.0621. All observations above this value are colored either dark blue or dark blue-gray. This class break value corresponds to the 85th percentile for that statistic. To summarize:

¹ At this time, the axis may only be moved within the display extent of the window. It cannot be dragged “off” the window; though it may seem intuitive to pan the PCP while moving the axis, that is impossible. You will need to drop the axis, pan the PCP (using the PCP Pan Slider), reselect the axis, and move it step-by-step to the desired location.

Those observations colored...	Have relatively _____ breast cancer rates, and ...	Have relatively _____ breast cancer variances.
light gray	low	low
light (bright) blue	high	low
dark gray	low	high
dark (royal) blue	high	high

Error messages

Great efforts were made to make the environment bug-free. Unfortunately, error messages and bugs may occasionally appear, and thus it is highly recommended that you save your work often. If an error message appears, click OK; it is possible that the error will not influence the representations or the operation of the system. In some cases, the piece of code that contains the error will be opened and displayed on top of the other windows. Close this code by simply clicking the x button in the upper right. In other cases, an error will appear because the interaction you attempted was performed too quickly (“A.Script is already in the process of being executed” is a typical error message here). In that case, click OK and perform your interactions more slowly. But in all cases, when the environment has performed as you intended, save your work so that the effect of any bugs or glitches will be minimized.

Contact the developer

Rob Edsall is an Assistant Professor of Geography at Arizona State University. The HVPCP system was completed as part of his doctoral research, and he can help with any questions or problems you might have. He would also be pleased to receive and feedback or constructive criticism about the design or operation of the system. Feel free to contact him through e-mail, regular mail, or telephone.

Rob Edsall
robedsall@asu.edu
 Department of Geography
 Arizona State University
 Tempe AZ 85287-0104
 (480) 865-5023