DV User Manual

Table of Contents

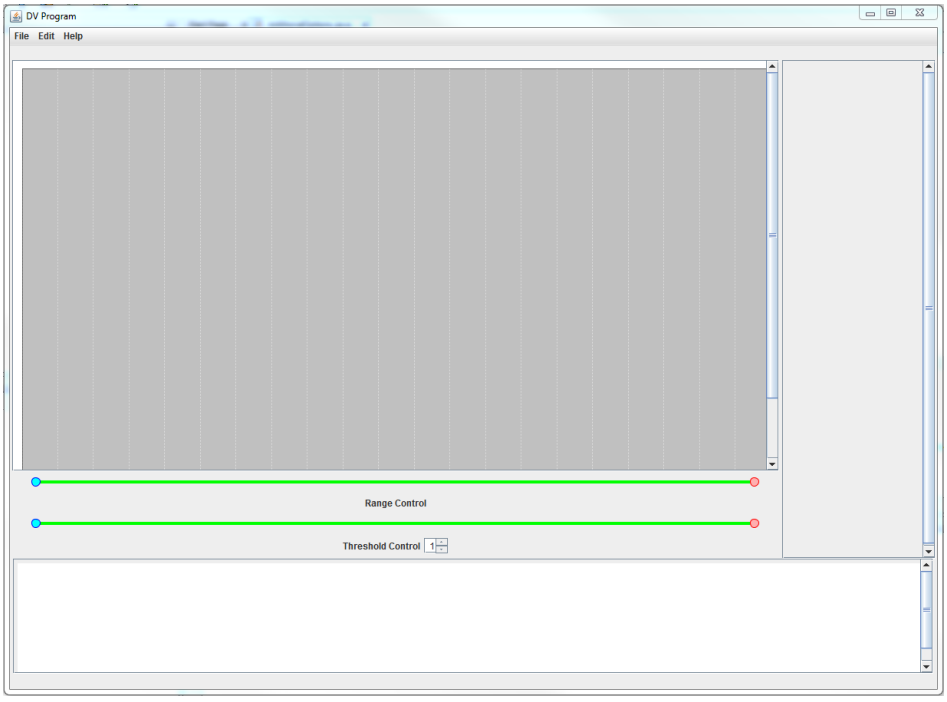
By Daniel Van Houten, Morgan Leblanc, Tyler Swan, Fawziah Alkharnda, Stephan Adams

1. Introduction

The purpose of the DV program is to render a visualization of data contained in one or more .csv files to improve data analysis. The graph is composed of many lines which represent a single data object, or row of the data set. Each line is composed of multiple vectors associated with the field's value, or the columns of the data set. This allows a user to visualize data with multiple dimensions through the methodology of the program. In addition, the user will also be able to interact with the visualization through various program functions, such as domain constraints, or field angle manipulation. Using the tools of the DV program, a user can view data of many different dimensions, make classifications of data objects, or draw other conclusions from the data.

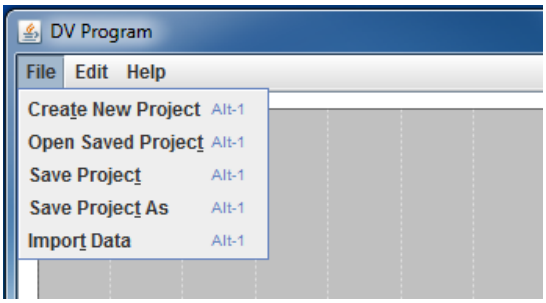
1.1 Getting Started

After the DV program has finished loading, it will display the following:

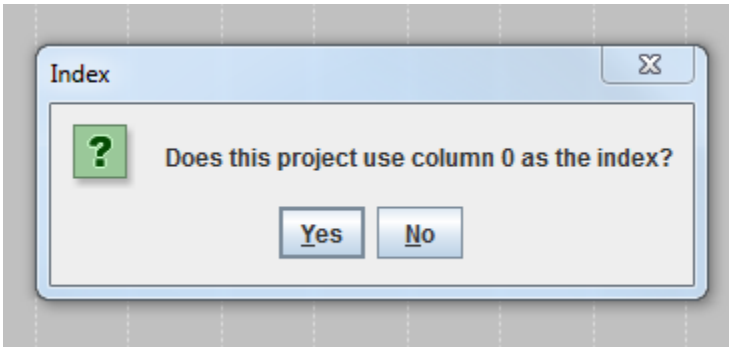


This initial screen is where a user will begin the process to create a new project, or pick up from another project. To start a new project, click on the "File" drop-down menu.

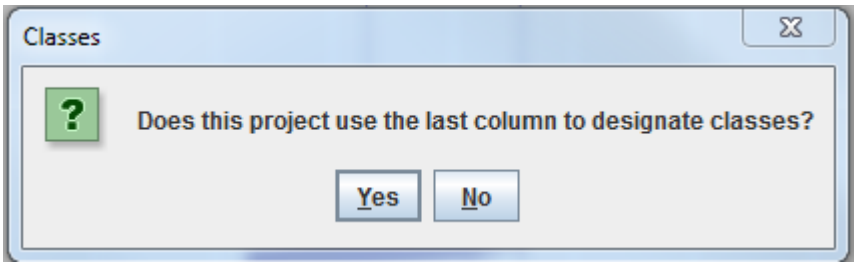
Underneath the “File” drop-down menu, choosing “Create New Project” opens a new window, beginning the process to create a new project from comma delimited data (e.g. .csv files). Select the desired dataset using the file browser displayed.



Commonly, the first column of a dataset may represent the index of the row, thus the program will ask the user if column 0 is used as the index upon initial click of “Create New Project.”

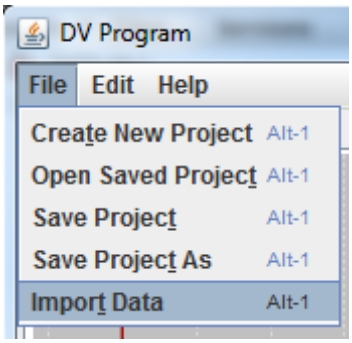


Similarly, a dataset may use the last column as the class indicator. Select the option yes if this is true, and no if not.



If a dataset’s classes are separated by files, DV can still visualize all classes simultaneously. Again, the user will create a project in the same process as above, however they must say “No” when prompted “Does this project use the last column to designate classes?”

Next, DV will render this single-class dataset. To load more classes from a different dataset, the user should click the “Import Data” option underneath the “File” drop-down menu. This option opens the file browser where the user can select another dataset.



The below picture, Figure 1, is an example of what the program shall render once the user has finished the creation process. In this diagram are several important labels. The label (1) shows the location of the **field angles** for user interaction. Label (2) is where the **confusion matrix** of the dataset will be displayed. Label (3) show slider locations for **range control** above, as well as **threshold control** below. Lastly, (4) indicates where the data will be visualized.

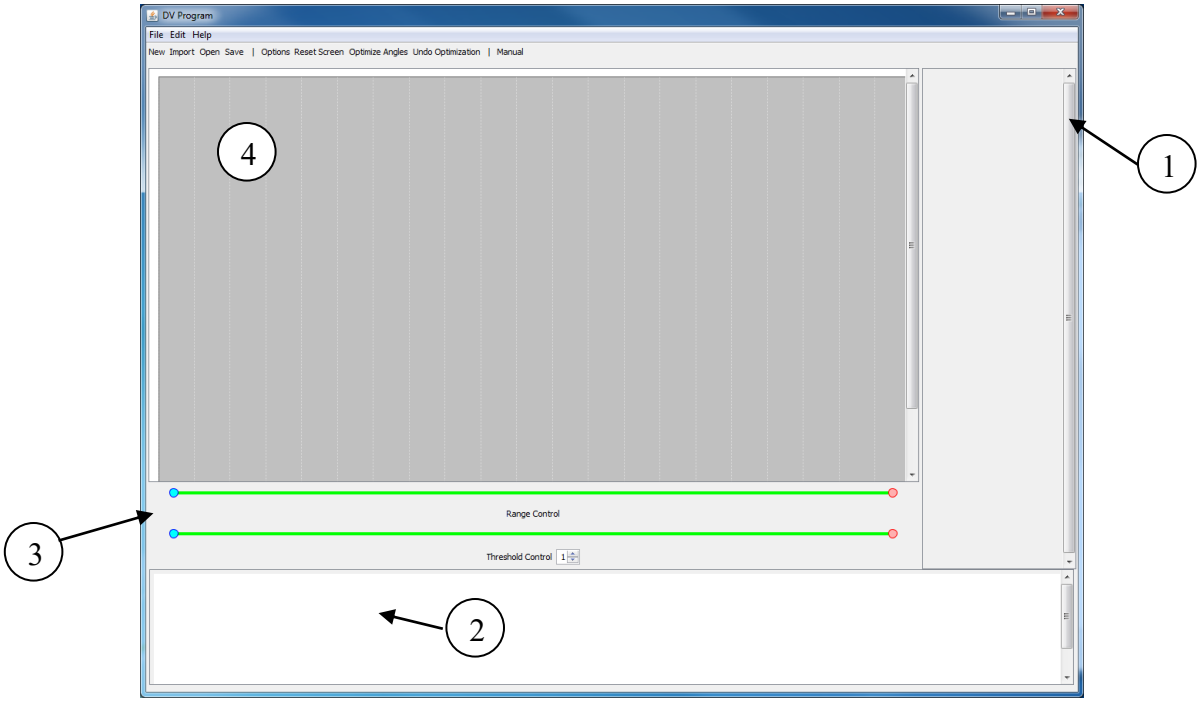
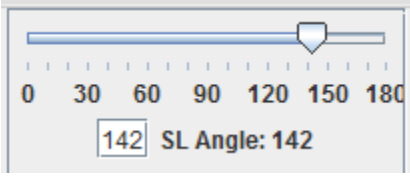


Figure 1

1.2 Manipulate Field Angles

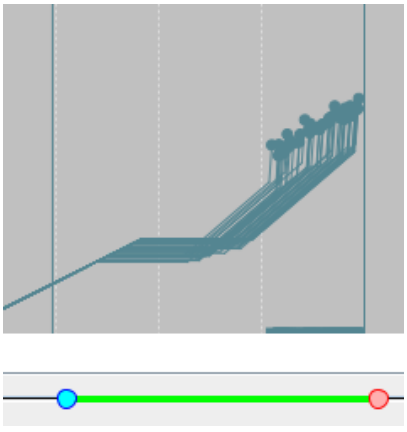
When rendering a visualization for a new project, the DV program will default every field's angle to 45 degrees, as shown on Figure 1. The user can change this using components from found in field angles panel (denoted by (1)).

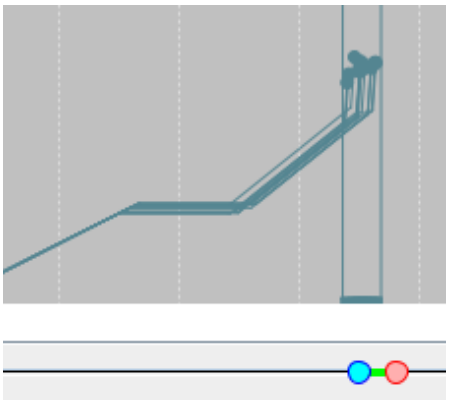


With these sliders, a user can manipulate the angle of the dimension which will be rendered in real time as the user interacts. This can also be done by inputting a number to the textbox. It should also be noted that the title of the dimension is displayed in the field angle's label. For example, the name of this dimension would be "SL."

1.3 Applying Domain Constraints

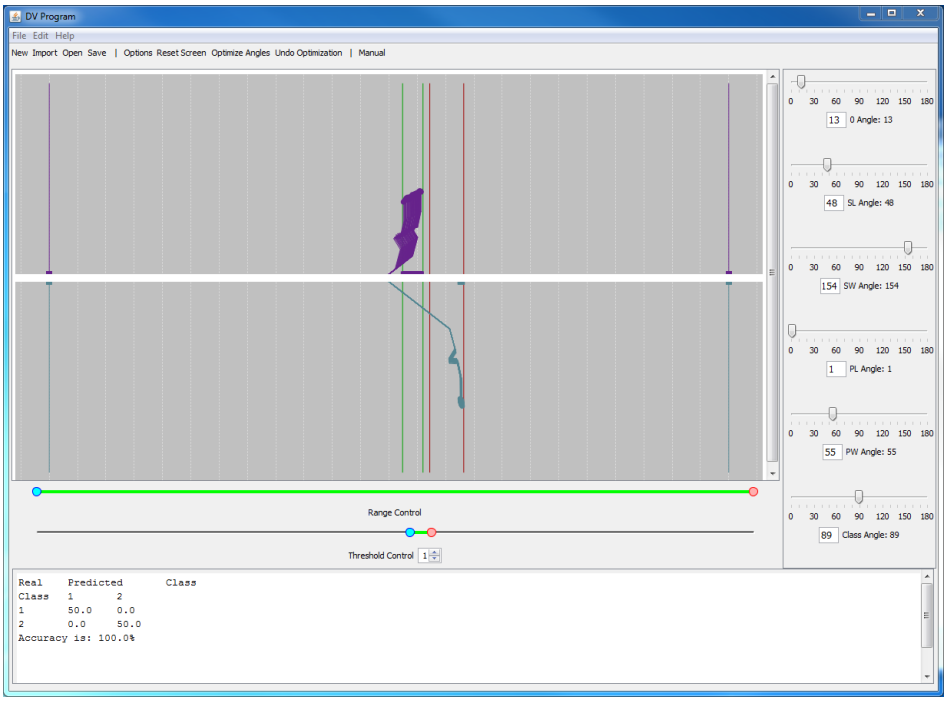
The domain constraints can be utilized through the **range control** sliders shown by label (3). By dragging this slider, the user can apply domain constraints to the visualization. This will exclude data from the visualization which resides outside of the specified domain. The closer together the sliders are, the more data that will be constrained.



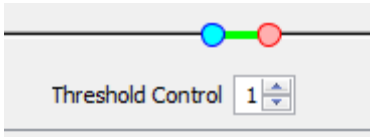


1.4 Generate Confusion Matrix

Below the domain slider, point to by label (3), the **threshold sliders** can be found. By default, the sliders are unset. Initially, the contents of the confusion matrix will be arbitrary until the user manipulates the sliders. To generate a useful confusion matrix, the user must set an interval with the sliders designating the predicted classification for that interval. The confusion matrix will reflect the change and display an accuracy of classification.



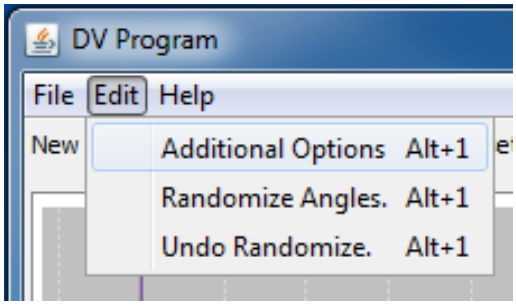
For sets with more than 2 classes, the user must set the interval for each class. To select a specific threshold, the user must use the threshold control textbox.

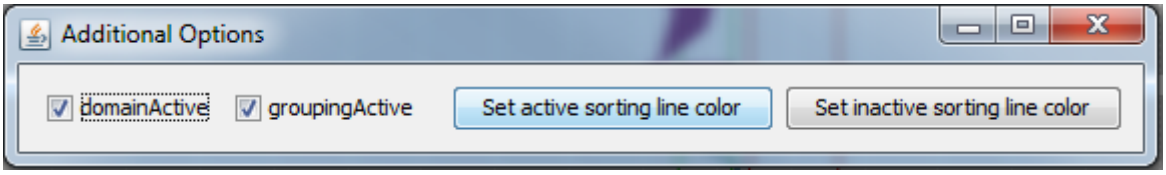


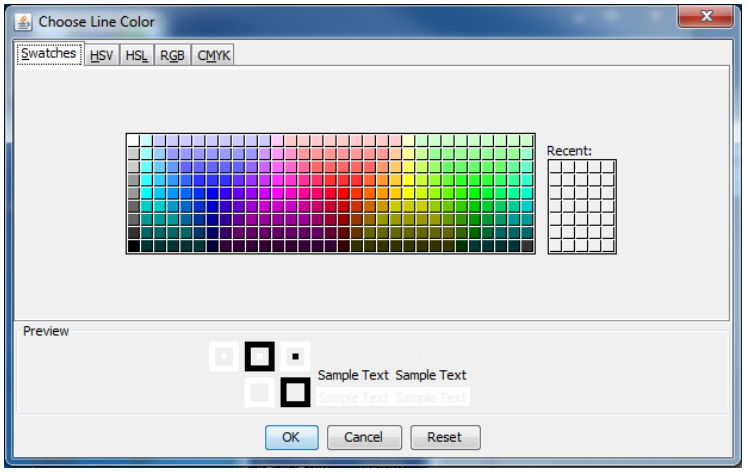
This allows the user to set any threshold at any time.

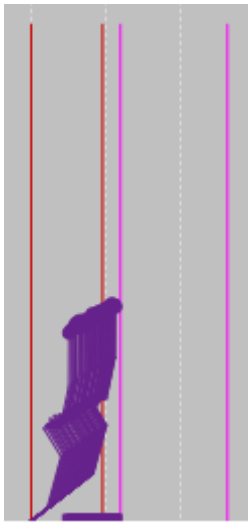
1.5 Display Additional Options

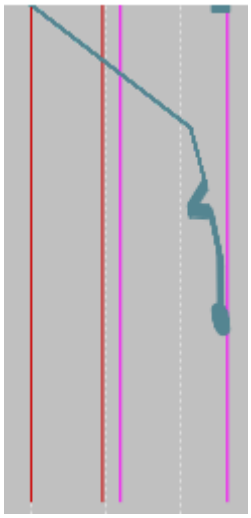
The user can find additional options by navigating to "Additional Options" underneath the "Edit" drop-down menu. These options allow the user to toggle the domain constraints, or grouping lines. As well, the user may change the color of lines representing thresholds.





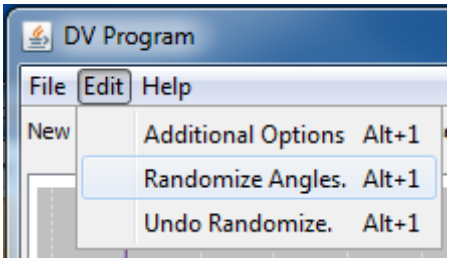






This menu is the result of the "Set active sorting line color" buttons, as well as the inactive line colors. Choosing a color and clicking OK will set the color of the respective line. On the right, the active threshold sliders were set to pink.

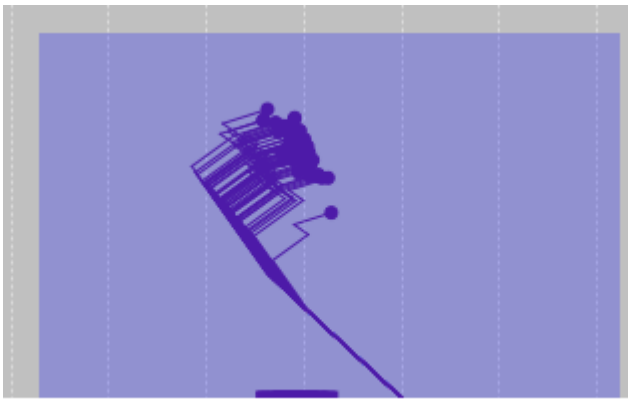
1.6 Randomize Angles Optimization

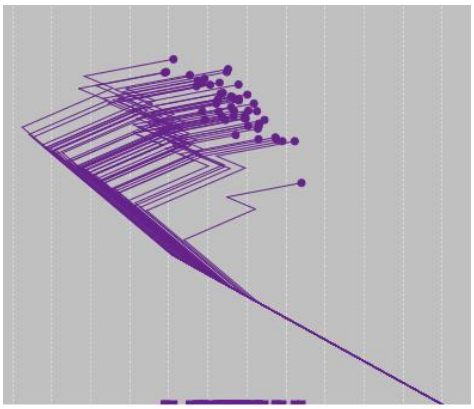


The DV program is capable of generating a possible optimization of angles to find the highest accuracy. DV continuously looks for a new set of angles until a new combination resulting in a higher accuracy is found. This can be initiated by clicking the “Randomize Angles.” option underneath the “Edit” drop-down menu. If the user wishes to revert to the configuration prior to random optimization, an “Undo Randomize.” options exists in the same menu.

1.7 Zoom function

The DV program allows users to zoom into the visualization by using a drag and drop box method to select an area to zoom, starting from the top left corner, moving down and right.

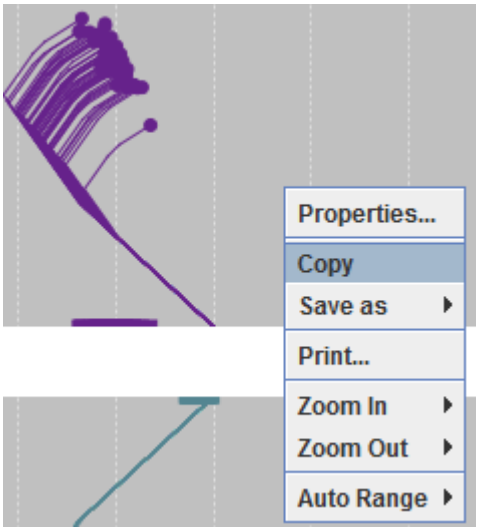




Similarly, the user can zoom back out by using the drag and drop box method in reverse, i.e. starting from bottom-right and moving to the top-left. To return to the original view, adjusting any of the sliders will reload the visualization. The last zoom option is using the scroll wheel, the user can zoom in and out for a more precise zoom.

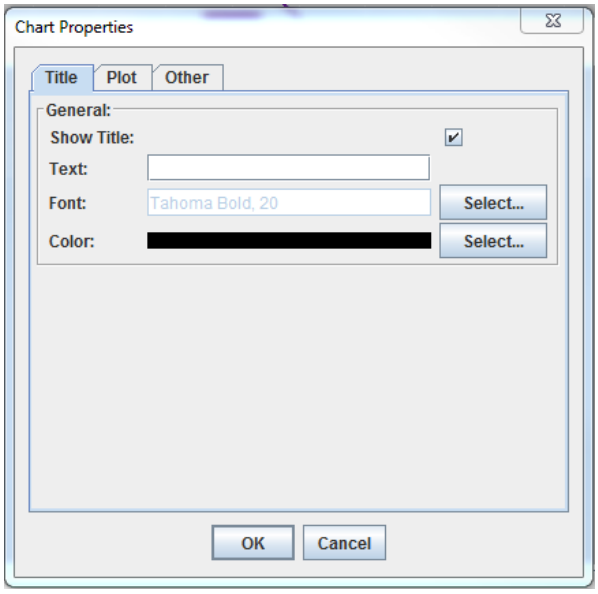
1.8 Image Options

The DV program also can export the visualization to a few different image options. By right clicking on a specific graph, the user may choose one of three options to export the image. Should the user choose “Print…” a new window will open prompting the user to confirm settings to print the image. If the user chooses “Save as,” DV will open a file browser prompting the user to select where to save the PNG image. The last possible image option, is to simply copy the PNG image to the user’s clipboard.

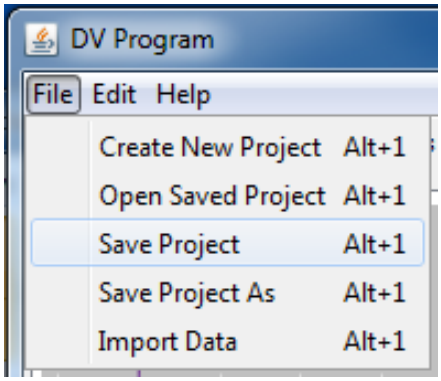


1.9 Chart Properties

Additional customization options can be set using the Chart Properties window, which can be found in the “Properties” option when the user right clicks on a graph (shown in section 1.8). This will display the Chart Properties window where a user can add titles, label axes, and change colors of the graph elements. Options to add custom labels to the axes are found under the “Plot” tab.



1.10 Saving the Project



If a user wants to save a project for later use, the DV program has this option located under the “File” drop-down menu. For the initial save of a project, the user should click “Save Project As” bringing up the file browser. The DV Program specifically needs .datv files for later use, thus the user should save a file with the “datv” extension. This can be done by entering “.datv” on the end of the file name within the file browser text field. Once the .datv project file has been created, the user may use the “Save Project” option from then on.

1.11 Opening a Saved Project

When a user wants to open a saved project, the option “Open Saved Project” under the “File” drop-down menu will open past projects. This button will bring up the file browser, where the user can search for their desired project. The DV program requires that all files are of the type “datv” to open past projects.