

---

# DV User Manual

## Table of Contents

<b>1. Introduction</b>	<b>2</b>
<b>1.1 Getting Started</b>	<b>2</b>
<b>1.2 Manipulate Field Angles</b>	<b>5</b>
<b>1.3 Applying Domain Constraints</b>	<b>6</b>
<b>1.4 Generate Confusion Matrix</b>	<b>7</b>
<b>1.5 Display Additional Options</b>	<b>8</b>
<b>1.6 Randomize Angles Optimization</b>	<b>9</b>
<b>1.7 Zoom Function</b>	<b>10</b>
<b>1.8 Image Options</b>	<b>11</b>
<b>1.9 Chart Properties</b>	<b>12</b>
<b>1.10 Saving the Project</b>	<b>13</b>
<b>1.11 Opening a Saved Project</b>	<b>14</b>
<b>2. 3+ Class Visualizations</b>	<b>14</b>

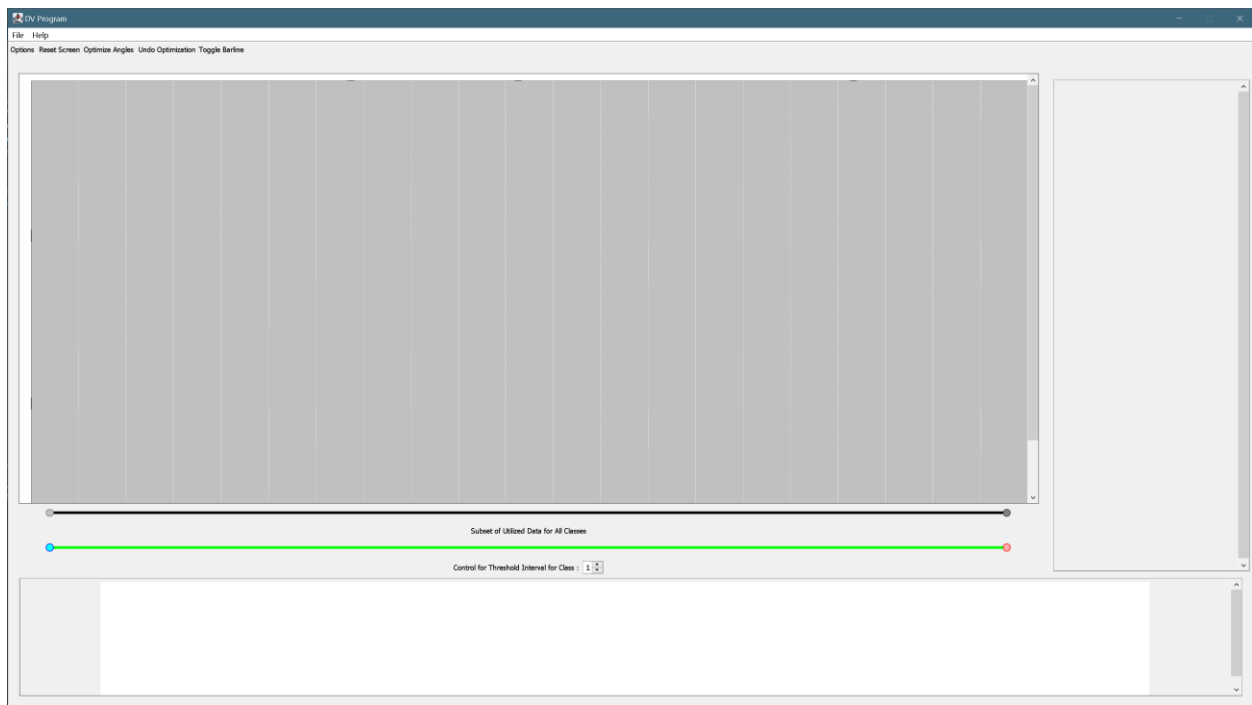
By: Daniel Van Houten, Morgan Leblanc, Tyler Swan, Fawziah Alkharnda,  
Stephan Adams, and Lincoln Huber

# 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.

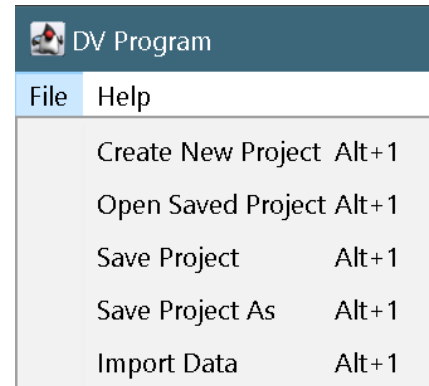
## 1.1 Getting Started

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

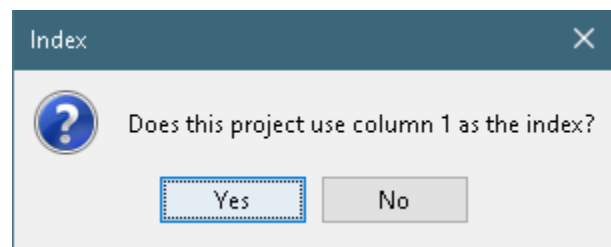


The initial screen is where a user will begin the process to create a new project or pick up from another project. To start a new project, click on the “File” drop-down menu.

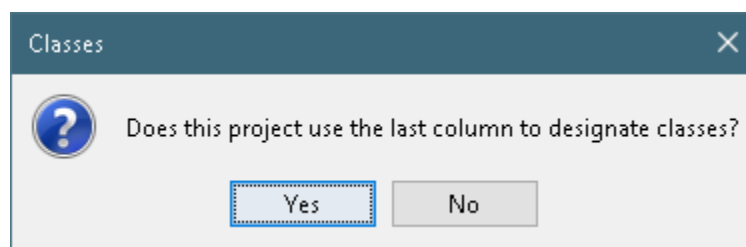
Underneath the “File” drop-down menu, choosing “Create New Project” opens a new window, beginning the process to create a new project from comma delimited data (e.g. .csv file). Select the desired dataset using the file browser displayed.



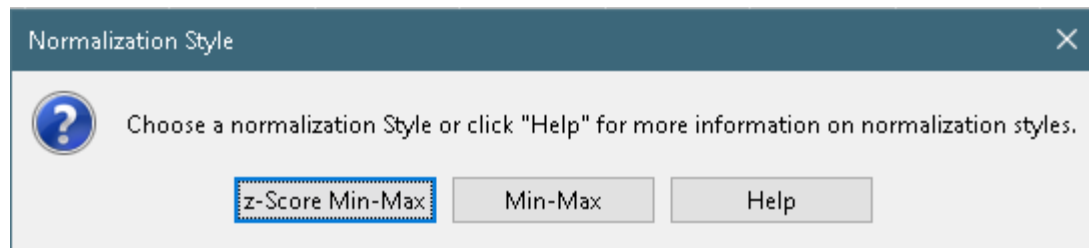
Commonly, the first column of a dataset may represent the index of the row, thus the program will ask the user if column 0 is used as the index upon initial click of “Create New Project.”



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

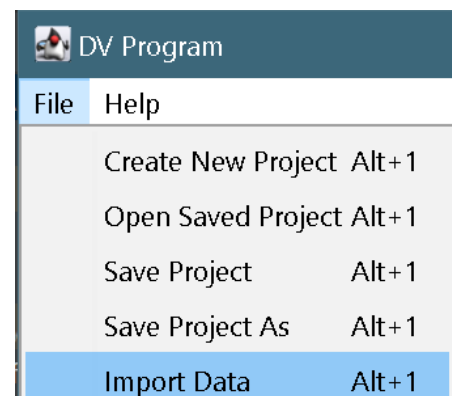


Then, the program will ask the user how to normalize the data. A z-Score Min-Max normalization will perform a z-Score standardization before normalizing the data from [0, 1] and a Min-Max normalization will directly normalize the data from [0, 1]. A user can also click “Help” for information of the exact mathematical formulas used for each option.

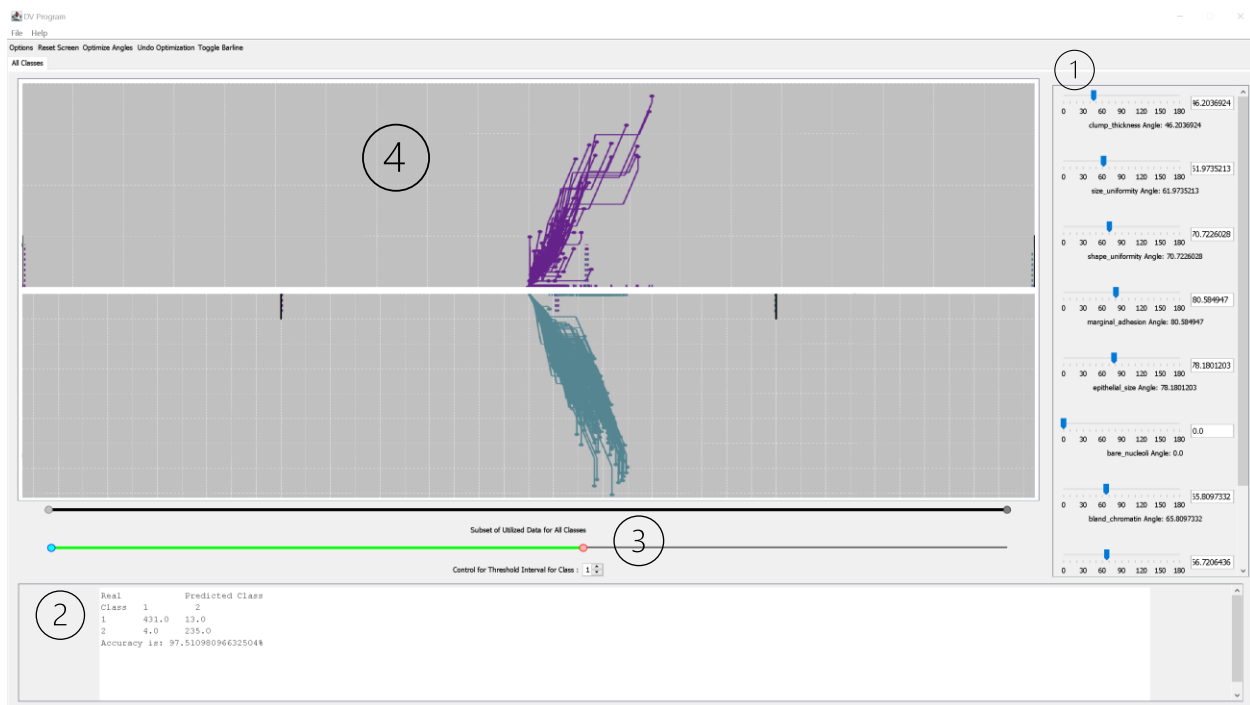


If a dataset's classes are separated by files, DV can still visualize all classes simultaneously. Again, the user will create a project in the same process as above, however they must say “No” when prompted “Does this project use the last column to designate classes?”

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.

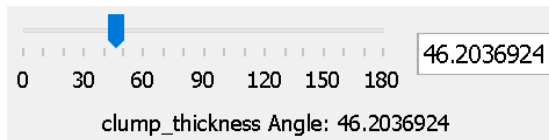


The below image is an example of what the program shall render once the user has finished the creation process. In this diagram are several important labels. The label (1) shows the location of the **field angles** for user interaction. Label (2) is where the **confusion matrix** of the dataset will 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.



## 1.2 Manipulate Field Angles

When rendering a visualization for a new project, the DV program will default use Linear Discriminant Analysis to calculate the angles and threshold, as shown on Figure 1. The user can change these using components found in the field angles panel (denoted by (1)).

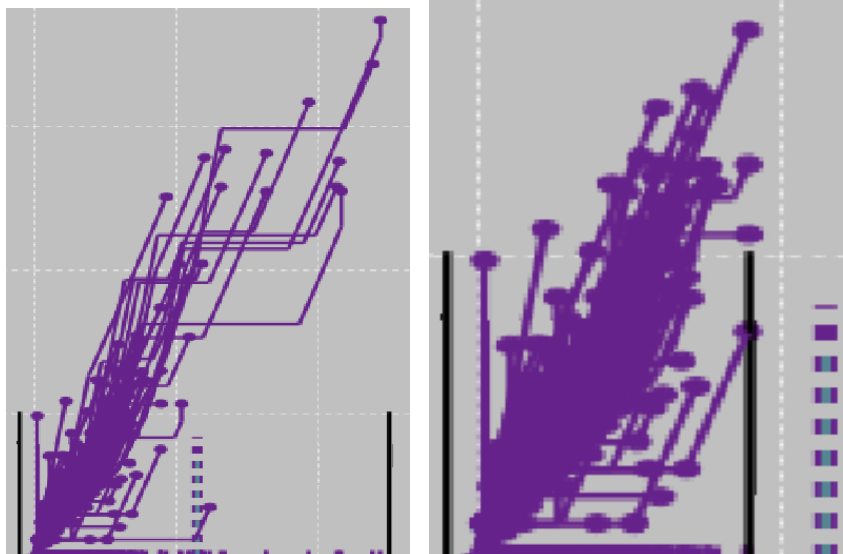


With these sliders, as user can manipulate the

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 "clump\_thickness."

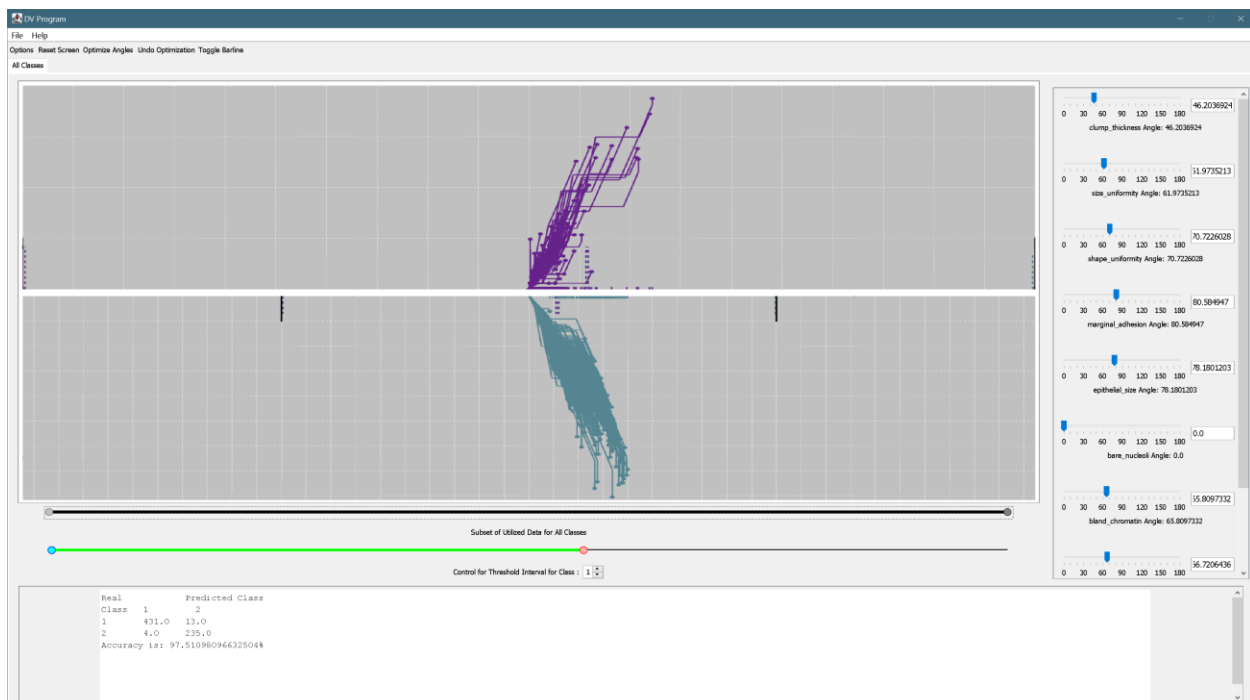
### 1.3 Applying Domain Constraints

The domain constraints can be utilized through the **Subset of Utilized Data for all Classes** slider show 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 close together the sliders are, the more data that will be constrained.

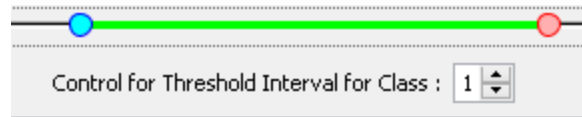


## 1.4 Generate Confusion Matrix

Below the domain slider, pointed to by label (3), the **Control for Threshold Interval** sliders can be found. By default, the sliders are set to the median domain location for their class, and then span one standard deviation to the left and the right. If the sliders are overlapping, the total interval for one of the sliders will be shrunk to give space for the other slider. Also, if the sliders have gaps between them and another slider or the edge of the total domain, the total interval for one of the sliders will be grown until the gap is filled. At this point a confusion matrix will be generated and display an initial accuracy of classification. The user can then adjust the sliders at will to designate different predicted classifications for different intervals. The confusion matrix will reflect these changes by displaying an updated accuracy of classification.

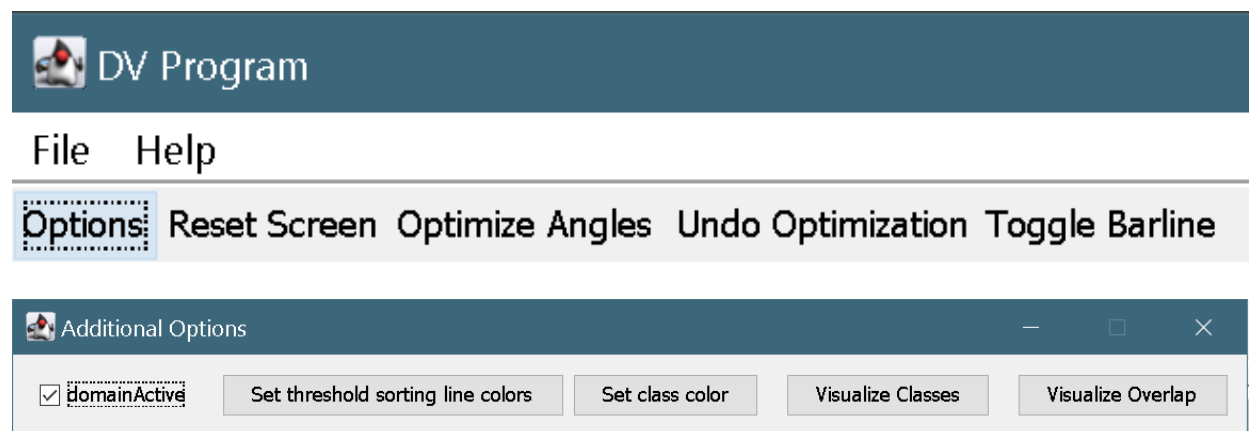


For sets with more than one class, the user must set the interval for each class. To select a specific threshold, the user must use the threshold control textbox.



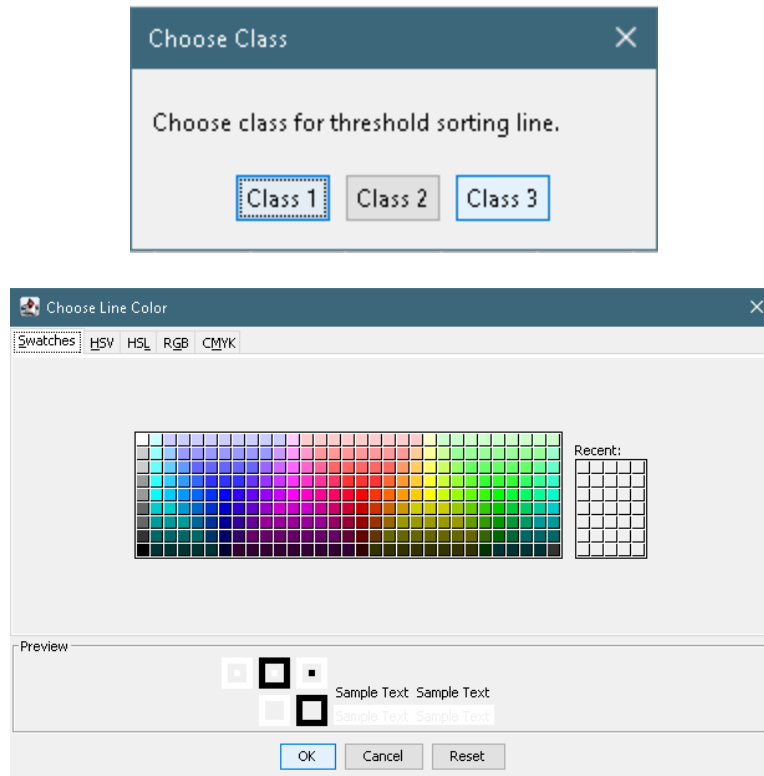
## 1.5 Display Additional Options

The user can find additional option by navigating to “Options” Underneath the main navigation bar. These options allow the user to toggle the domain, change the class colors and threshold interval colors, visualize different classes in the case of 3+ class datasets, and visualize the overlapping or inaccurate areas of the visualizes. For more information on 3+ class dataset visualization please see section 2.

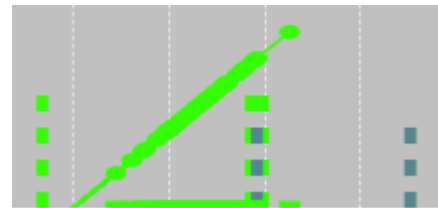


After choosing either “Set threshold sorting line colors” or “Set class color” the “Choose Class” menu will popup. After clicking on a class, a color choosing menu will appear. Choosing a color will change the sorting line color or class color for the chosen class.





On the right, the color for class one's threshold sorting lines and overall color were set to green.



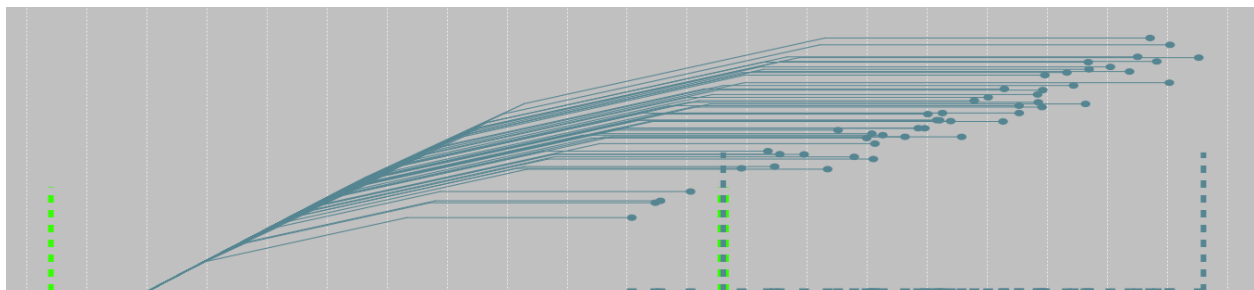
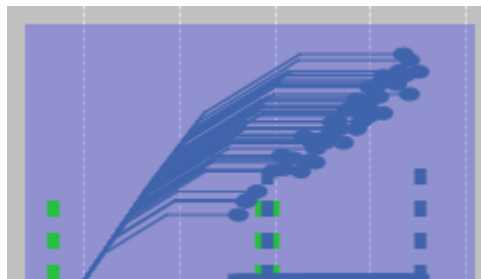
## 1.6 Randomize Angles Optimization

The DV program can generate a possible optimization of angles to find the highest accuracy. DV continuously look for a new set of angles until a new combination resulting in a higher accuracy is found. This can be initiated by clicking the “Optimize Angles” option underneath the main navigation bar. If the user wishes to revert to the configuration prior to random optimization, an “Undo Optimization” option 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. An Example of this is shown below; the first image is the selection and the second image is the zoomed in area selected.



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

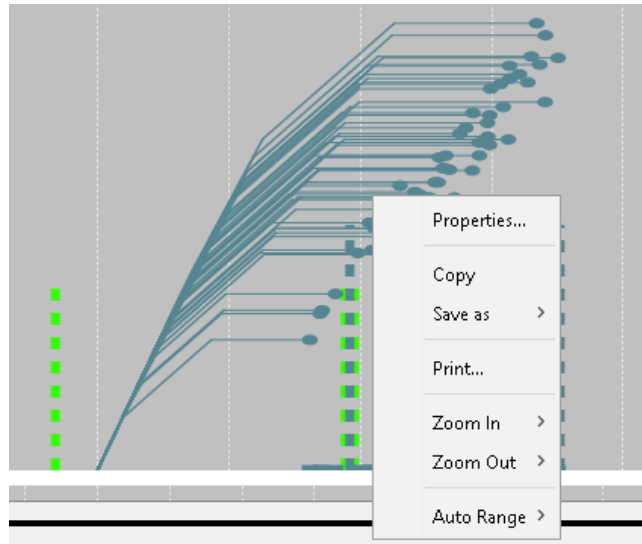
original view, just click the “Reset Screen” button on the tool bar at the top of the screen.



The last zoom option is using the scroll wheel, the user can zoom in and out for a more precise zoom.

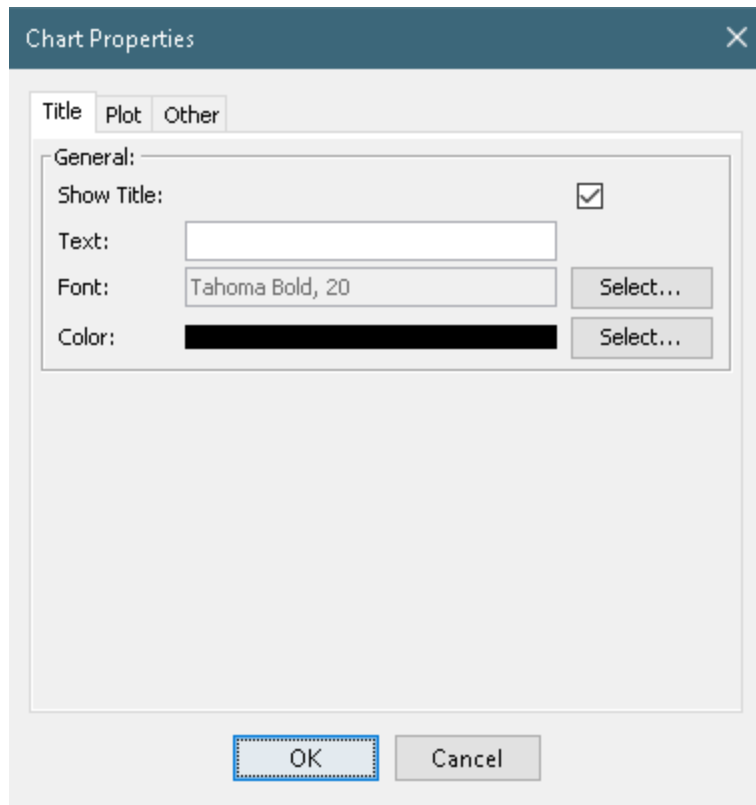
## 1.8 Image Options

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

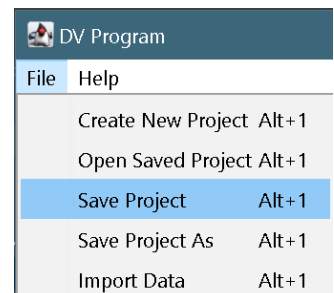
Additional customization options can be set using the Chart Properties window, which can be found in the “Properties” option when the user right clicks on a graph (shown in section 1.8). This will display the Chart Properties window where a user can add titles, label axes, and change colors of the graph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

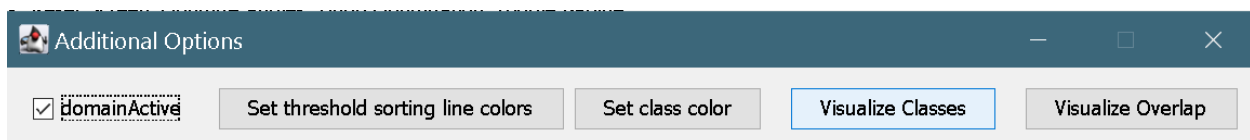


## 1.11 Opening a Saved Project

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 be of the type “.datv” to open past projects.

## 2. 3+ Class Visualizations

Since linearly separating more than two classes with parallel lines is generally not feasible, the DV program only allows the visualization of 2 classes at once. In the case of 3+ class datasets, the DV program will visualize one class against all other classes in the dataset. For instance, if a dataset had three classes the DV program could visualize class one against classes two and three. To change which class is being visualized individually the “Visualize Classes” option can be found after clicking the “Options” button under the main navigation bar.



After selecting the “Visualize Classes” button a new menu will appear with all available classes. Simply choose a class to visualize that class against all others. At this point the chosen class will be visualized.

