

Data Exploration Guide

TwoRavens: Tabular Data Exploration

Exploring and Analyzing Tabular files in Dataverse

On the files tab, click on the “Explore” button to initiate TwoRavens Data Exploration and Analysis Tool.

Selection/Left Panel:

The left panel contains two sets of buttons: (1) Original Data and Subset Data; and (2) Variables, Models, and Summary.

Original Data and Subset Data:

When TwoRavens is initiated, you begin with the original dataset, and thus the Original Data button is selected. You will not be able to select the Subset Data until you have subsetted the data using the subset and select features in the right panel. After you have selected a subset of your data, you may toggle between that subset and the original data. If you wish to select a different subset, you may do so, but note that only one subset is supported at a time.

Variables, Models, and Summary:

Each of these tabs displays something different in the left panel when selected. The Variables tab shows a list of all the variables. When a variable name is hovered over, you can see that variable’s summary statistics. The first three variables are selected by default and displayed in the center panel, but you may add or remove variables by clicking on their name in the Variables tab.

The Models tab displays a list of Zelig models that are supported by TwoRavens. A brief model description is visible when hovering on the model name. Depending on the level of measurement of the dependent variable (continuous, ordinal, dichotomous, etc.), particular models may or may not be appropriate.

Note that to estimate a model, you must select one from the list. Currently, please use only Ordinary Least Squares (ls) as we are working on making other models available. (Suggestion: maybe we need to gray out the ones the other ones for the time being)

The Summary tab shows summary statistics when a pebble is hovered over. If one removes the pointer from hovering over a pebble, the previous tab will be displayed. So, if you wish to have the summary tab displayed regardless of where the pointer is, click on Summary before hovering over a pebble. Otherwise, if Variables or Models has been the last tab selected, when the pointer is no longer hovering over a pebble, the table will return to Variables or Models.

Modeling/Center Panel:

The center panel displays a graphical representation of variable relations and denotes variables that have been tagged with certain properties. Variables may be tagged as either a dependent variable, a time series variable, or a cross sectional variable. Each of these are accomplished by clicking on the appropriate button at the top of the screen, and then clicking on a pebble in the center panel. You'll notice that when a variable is tagged with a property, the fill color becomes white, and the outline (or stroke) of the pebble turns the color of property's button. Note that to estimate a model, the dependent variable must be selected.

Variable relations are specified by point-click-drag from one pebble to the other. When a path between pebbles has been specified, it is visually presented with a dotted arrow line and may be removed by pushing the delete key on your keyboard.

Results/Right Panel:

This section allows you to specify a subset of the data that you wish to estimate the model on, or that you wish to select and see updated summary statistics and distributions, and to set covariate values for the Zelig simulations (this is Zelig's `setx` function).

Subset

To subset the data, click on the subset button and highlight (or brush) the portion of the distribution that you wish to use. You may remove the selection by clicking anywhere on the plot that is outside of the selected area. Or, if you wish to move the selected region, click inside the selection and move it to the left or right. If no region is selected, then by default the full range of values are used. If more than one variable is selected for subsetting, only the overlapping region is used in the model. If there are no overlapping regions, (i.e., if subsetted there would be no data), then only the first variable is used. Notice that range (or extent) of the selection for each variable is displayed below.

With a region selected, you have two options. First, you may click the Estimate button to estimate a model on using only the specified subset. Second, you may click the Select button. This will not estimate a model, but it will subset the data and return new summary statistics and plots for the subset. You may wish to use this feature to see how a subset will change the Set Covariate (Zelig's `setx`) defaults, for example. After selecting a subset, you may toggle back and forth between the subsetted data and the original data by activating the appropriate button in the left panel.

Set Covariates

The Set Covariates button plots the distributions of each of the pebbles with an additional axis that contains two sliders, each of which default to the variable's mean. This is TwoRavens' equivalent of Zelig's `setx` function. Move these sliders to the left or right to set your covariates at the desired value prior to estimation. Notice that the selected values appear below each plot.

After clicking the Estimate button, the model will be estimated and, upon completion, results appear in the Results tab. The results include figures produced by Zelig (and eventually the equation that has been estimated, the R code used to estimate the model, and a results table).

Additional Buttons:

Estimate

This executes the specified statistical model. Notice the presence of blue highlight on the “Estimate” button while process is running, turning into green upon completion. Note: you cannot use estimate without selecting a dependent variable and a model.

Force

The Force button allows you to control the way layout of the pebbles. To use this feature, first make sure none of the pebbles are highlighted. If one is, simply click on it to remove the highlighting. Second, press and hold the control key. Third, while holding down the control key, click the Force button. Fourth, continue to hold the control key and click on a pebble. You may now release the control key. Click on a pebble and drag it around on your screen.

Reset

This is your start over button. Clicking this is equivalent to reloading the Web page or re-initiating TwoRavens.

Scenario Example:

WorldMap: Geospatial Data Exploration

WorldMap

[WorldMap](#) is developed by the Center for Geographic Analysis (CGA) at Harvard and is an open source software that helps researchers visualize and explore their data in maps. The WorldMap and Dataverse collaboration allows researchers to be able to upload shapefiles to Dataverse for long term storage and receive a persistent identifier (through DOI) as well as be able to easily move into WorldMap to interact with the data and save to WorldMap as well. GeoConnect is the platform integrating Dataverse and WorldMap together and what you will use to visualize your data.

Uploading Shapefiles to Dataverse

To get started, you will need to create a dataset in Dataverse. For more detailed instructions on creating a dataset, read the [Dataset + File Management](#) portion of this user guide.

Dataverse recognizes ZIP files that contain the components of a shapefile and will ingest them as a ZIP.

Once you have uploaded your ZIP files comprising a shapefile, a Map Data button will appear next to the file in the dataset.

Mapping your data with Geoconnect

In order to use the WorldMap and Dataverse integration, your dataset will need to be published. Once it has been published, you will be able to use the MapData button. Click on the Map Data button to be brought to GeoConnect, the portal between Dataverse and WorldMap that will process your shapefile and send it to WorldMap.

To get started with visualizing your shapefile, click on the blue Visualize on WorldMap button in GeoConnect. It may take 30 seconds or longer for the data to be sent to WorldMap and then back to GeoConnect

Once the visualizing has finished, you will be able to style your map through Attribute, Classification Method, Number of Intervals, and Colors. At any time, you can view the map on WorldMap if you would like to see how it will be displayed there.

After styling your map, you can delete it or return to Dataverse. If you decide to delete the map, it will no longer appear on WorldMap. By returning to Dataverse, you will send the styled map layer to WorldMap as well as to Dataverse where a preview will be available of the map layer you styled using GeoConnect.

To map the shapefile again, all you will need to do is click the Map Data button again.