Graph Example Applications

These example applications for S-PLUS are designed to demonstrate the use of graphic functions in the USGS library. A general knowledge of S-PLUS is assumed. For users needing introductions to S-PLUS, several are available in bookstores and online. Examples include Burns, Patrick, 2008, S Poetry, last accessed March 19, 2009 at *http://www.burns-stat.com/pages/spoetry.html*, and Lam, Longhow 2001, An introduction to S-PLUS for Windows: Amsterdam, The Netherlands, CANdiensten, 230 p., last accessed March 26, 2010, at *http://www.splusbook.com*. A detailed introduction to the S language is Kraus, A. and Olson, M., 2002, The basics of S-PLUS:, New York, Springer, 420 p. Another example of an introduction that concentrates on statistical applications is Venables, W.N., and Ripley, B.D., 2002, Modern applied statistics with S: New York, Springer, 495 p.

A few type face conventions are used to assist the user in interpreting the guidelines. S-PLUS object names and function arguments are in *italics* and function names are followed by parentheses *()*, window names are highlighted in **bold**, and object class names are in a sans-serif font. User commands, called calls, are in `plain text` and on separate lines. Any printed output from a call, whether in a **Commands Window** or **Report Window**, is shown in `plain text` in a box. Plots are shown as is, copied from the default settings for an S-PLUS **Graph Sheet**; however, individual user preferences can change the way the plots are displayed. Any S-PLUS dialog menu directions are shown in **bold** separated by a vertical bar (|). Column names and other dialog box entries are underlined.

An important way to manage data is to use the **Object Explorer Window**. The **Object Explorer Window** can be opened by clicking on the small icon that is a little yellow folder with two blue dots above it .

Note that none of these graphs should be considered as being publication-ready. These graphs are intended as easy-to-use data visualization tools.

Introduction. *Plot the Data*
Application 1. *Create a ground-water hydrograph*
Application 2. *Create a surface-water hydrograph*
Application 3. *Create a Piper diagram*

Plot the Data

The following text was modified from Helsel and Hirsch (2002). Perhaps it seems odd that a chapter on graphics appears at the front of a text on statistical methods. We believe this is very appropriate because graphs provide to the data analyst crucial information that is difficult to obtain in any other way. For example, figure 2.1 shows eight scatterplots, all of which have exactly the same correlation coefficient. Computation of statistical measures without reference to a plot is an invitation to misunderstanding data, as figure 2.1 illustrates. Graphs provide visual summaries of data that more quickly and completely communicate essential information than do tables of numbers.

Graphs are essential for two purposes:

1. they provide insight for the analyst into the data under scrutiny, and
2. they illustrate important concepts when presenting the results to others.

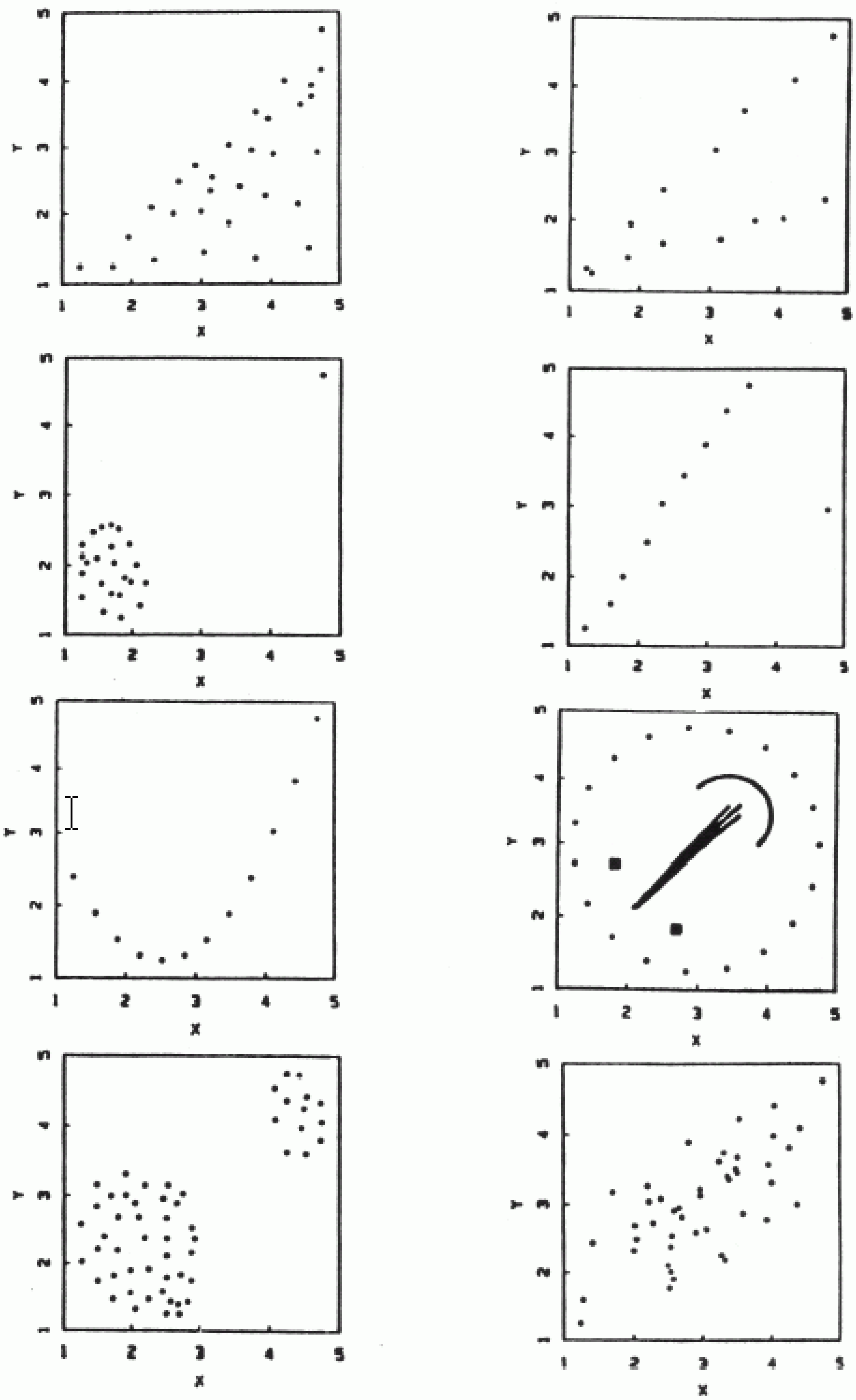


Figure 2.1 Eight scatterplots all with correlation coefficient r = 0.70 (Chambers and others, 1983).

The first of these tasks has been called exploratory data analysis (EDA), and is the subject of this chapter. EDA procedures commonly are (or should be) the "first look" at data. Patterns and theories of how the system behaves are developed by observing the data through graphs. These are inductive procedures; that is, the data are summarized rather than tested. Their results provide guidance for the selection of appropriate deductive hypothesis testing procedures.

After an analysis is complete, the findings must be reported to others. Regardless of whether the findings are documented in a written report or oral presentation, the analyst must convince the audience that the conclusions reached are supported by the data. No better way exists to do this than through graphics. Many of the same graphical methods that have concisely summarized the information for the analyst will also provide insight into the data for the reader or audience.

REFERENCES:

Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A., 1983, Graphical methods for data analysis, Boston, PWS-Kent Publishing Co., 395 p.

Helsel, D.R. and Hirsch, R.M., 2002, Statistical methods in water resources: U.S. Geological Survey, Techniques of Water-Resources Investigations book 4, chap. A3, 522 p.

Create a ground-water hydrograph

This application illustrates how to create a groundwater hydrograph using the menu dialogue. Hydrographs are a specialized kind of time-series plot that shows hydrographic data. Any time-series plot that shows hydrographic data would be considered a hydrograph, so a hydrograph could be made from any time-series plotting function or application.

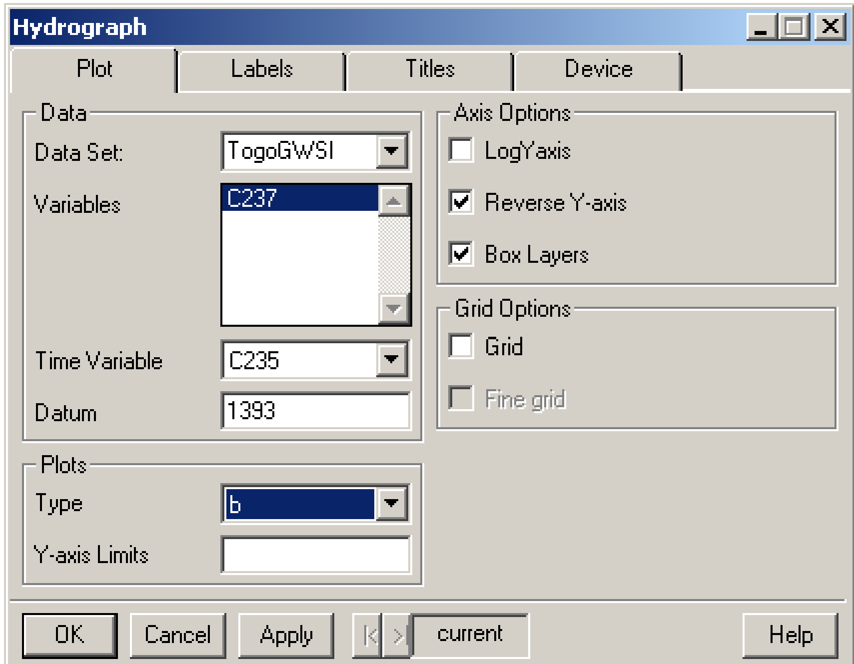
Step 1. Import the data

This example uses the *TogoGWSI* groundwater dataset. The source file is TogoGWSI.RDB and is in the USGS library folder. For details on importing this dataset, see the *Import an RDB file* section (application 1) in the USGS Data Example Applications help file, available by choosing **Help | USGS Example Apps | Importing Data Examples**.

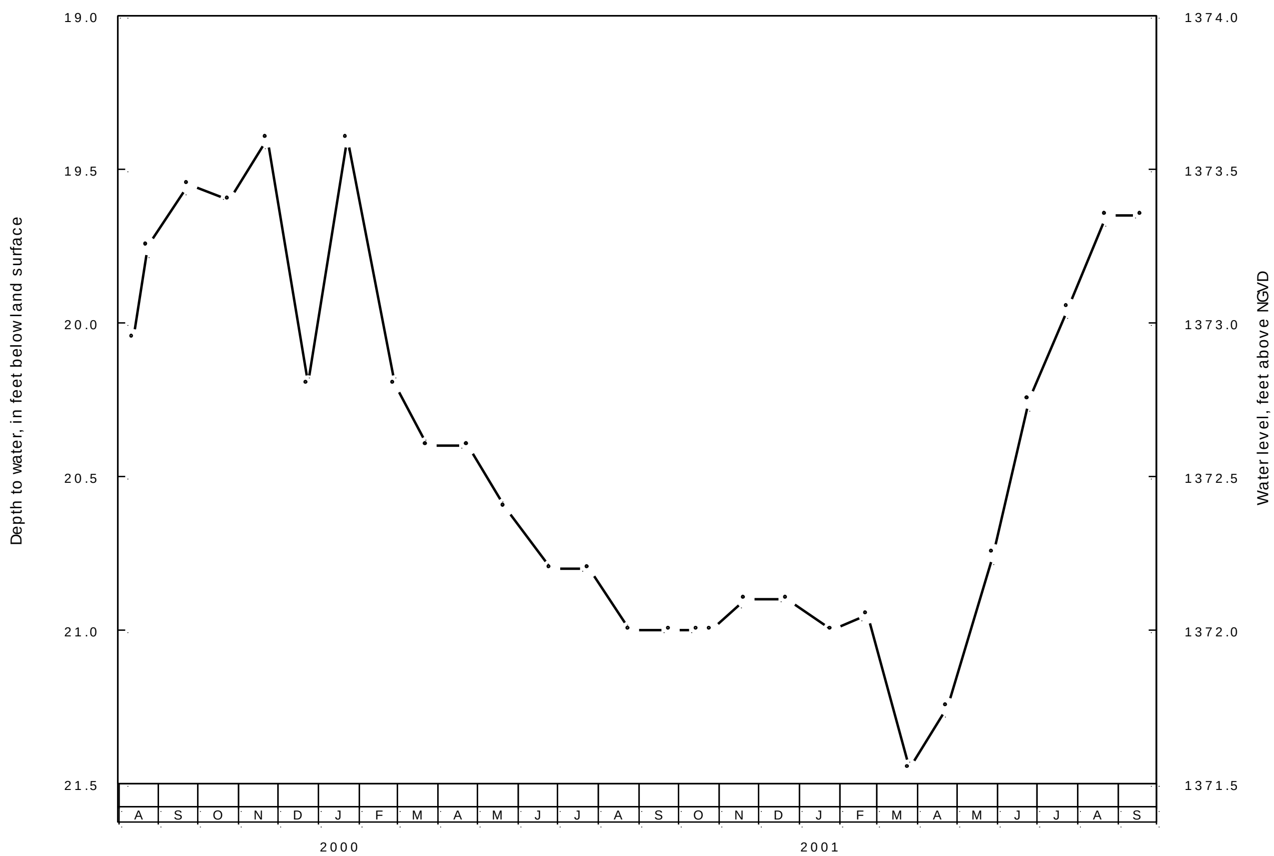
There are three columns in the *TogoGWSI* dataset. The column C235 is the date of observation, C709 is the time of observation, and C237 is depth below land surface. The land-surface elevation is 1,393 feet.

Step 2. Create the hydrograph

Click on **USGS | Hydrograph…** to open the **Hydrograph Window** and select the options shown on the Plot page.



Make sure to select C237 in the **Variables** box. You must select at least one of the columns listed in **Variables**, otherwise you will get an "undefined columns selected" error message. Any value entered for **Datum** automatically checks the **Reverse y-axis** and changes the **Left** and **Right Y-axis Labels** on the **Titles** page. The **LogYaxis** option must be unchecked. You may choose to click on the **Titles** page and change the labels if you wish. Click **OK** to produce the graph.



Depth to water, in feet below land surface
19.0
19.5
20.0
20.5
21.0
21.5
Water level, feet above NGVD
1374.0
1373.5
1373.0
1372.5
1372.0
1371.5
A S O N D J F M A M J J A S O N D J F M A M J J A S
2000
2001

Create a surface-water hydrograph

This application illustrates how to create a surface-water hydrograph using the menu dialogue. Hydrographs are a specialized kind of time-series plot that shows hydrographic data. Any time-series plot that shows hydrographic data would be considered a hydrograph, so a hydrograph could be made form any time-series plotting function or application.

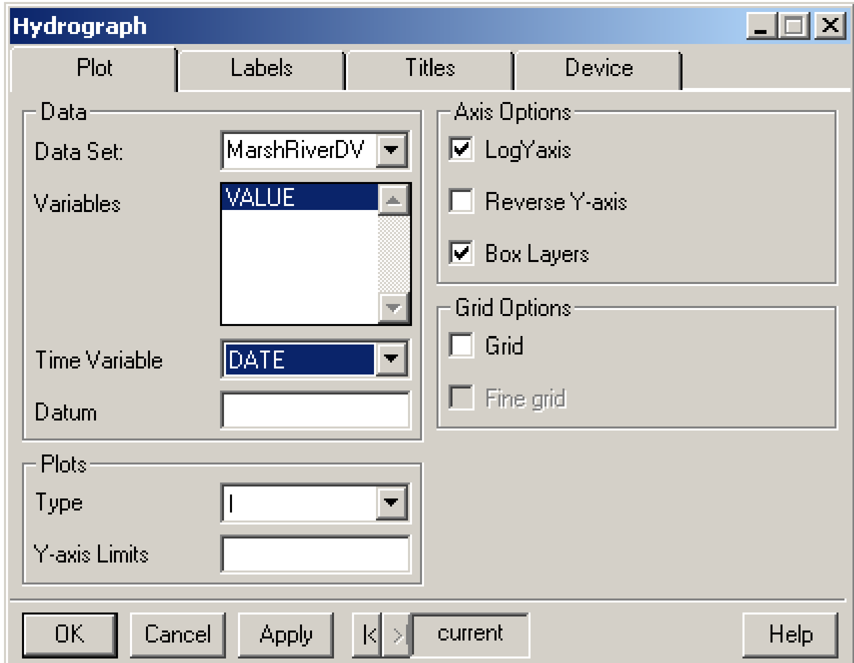
Step 1. Import the data

This example uses the *MarshRiverDVA* dataset. The source file is MarshRiverDVA.RDB and is in the USGS library folder. For details on importing this dataset, see Import an RDB file in the USGS Data Example Applications help file, available by choosing **Help | USGS Example Apps | Importing Data Examples**.

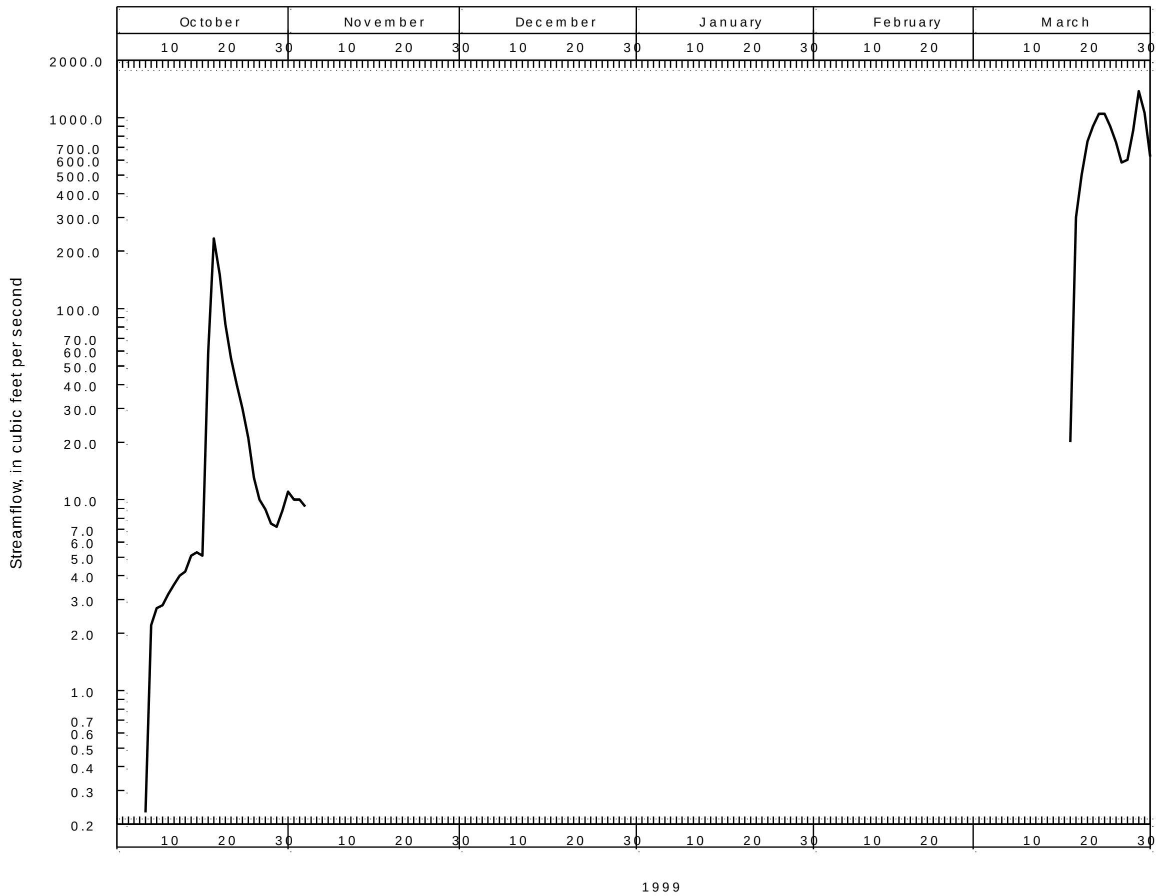
There are eight columns in the *MarshRiverDVA* dataset. Only two columns are needed for the hydrograph. The column named VALUE is the mean daily flow and DATE is the day of the observed flow.

Step 2. Create the hydrograph

Click on **USGS | Hydrograph…** to open the **Hydrograph Window** and select the options shown on the **Plot** page.



Make sure to select VALUE in the **Variables** box. You must select at least one of the columns listed in **Variables**, otherwise you will get an "undefined columns selected" error message. The **Left Y-axis Labels** on the **Titles** page is set by default to be correct for streamflow. You may choose to click on that page and change the text if you wish. Click **OK** to produce the graph.


October
November
December
January
February
March
Streamflow, in cubic feet per second
1999

Create a Piper diagram

This application illustrates how to create Piper (1944) diagram using the menu dialogue. Piper diagrams are a specialized kind of water-quality plot that shows the relative proportions of anions and cations in separate ternary diagrams and the overall composition of the water in a central rhombus. They are useful graphs to visually identify samples from different sources or to identify processes, such as solution or dilution, which affect water quality.

Step 1. Get the data

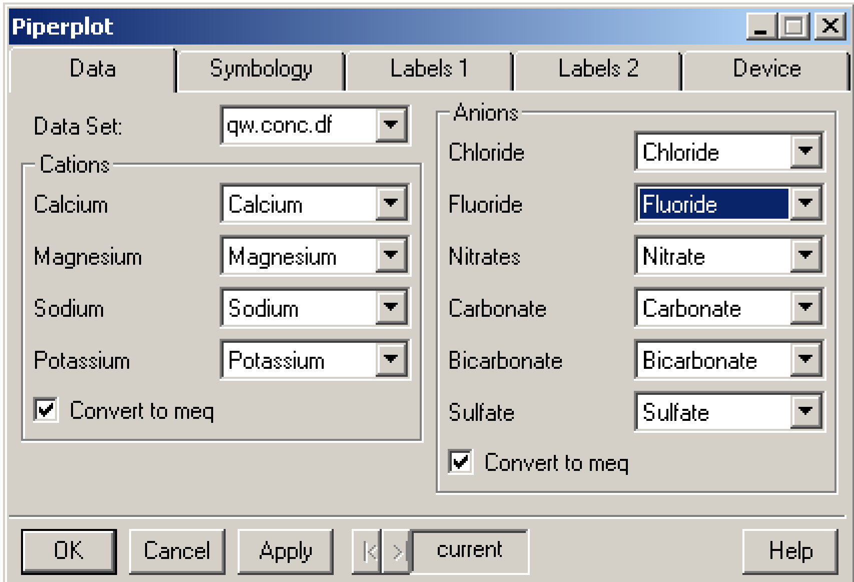
This example uses the *qw.conc.df* water-quality dataset. It is an S-PLUS dataset in the USGS library. It may be copied to the user's chapter by clicking on the dataset in the USGS library and dragging to the current working chapter using the left-hand pane in the **Object Explorer Window**, or by typing the following call in the **Commands Window**.

```
qw.conc.df <- qw.conc.df
```

There are 24 columns in the *qw.conc.df* dataset. The first three columns are the station identifier and the date and time of the same. The next columns are the concentration values and remark code pairs for sodium, potassium, magnesium, calcium, chloride, fluoride, sulfate, bicarbonate, carbonate and nitrate. The last column indicates whether the well is deeper or shallower than 20 feet.

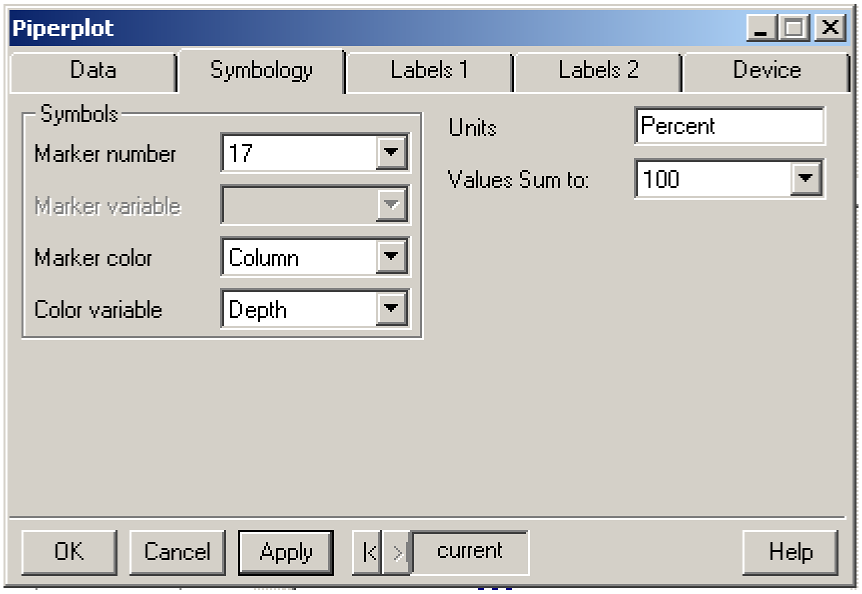
Step 2. Create the plot

Click on **USGS | Piperplot…** to open the **Piperplot Window** and select the options shown on the **Data** page. The dialog window will try to select columns names for each of the **Cations** or **Anions** from the data in **Data Set**. If the columns represent concentrations in milligrams per liter, then the **Convert to meq** should be checked in each of the **Cations** or **Anions** groups.



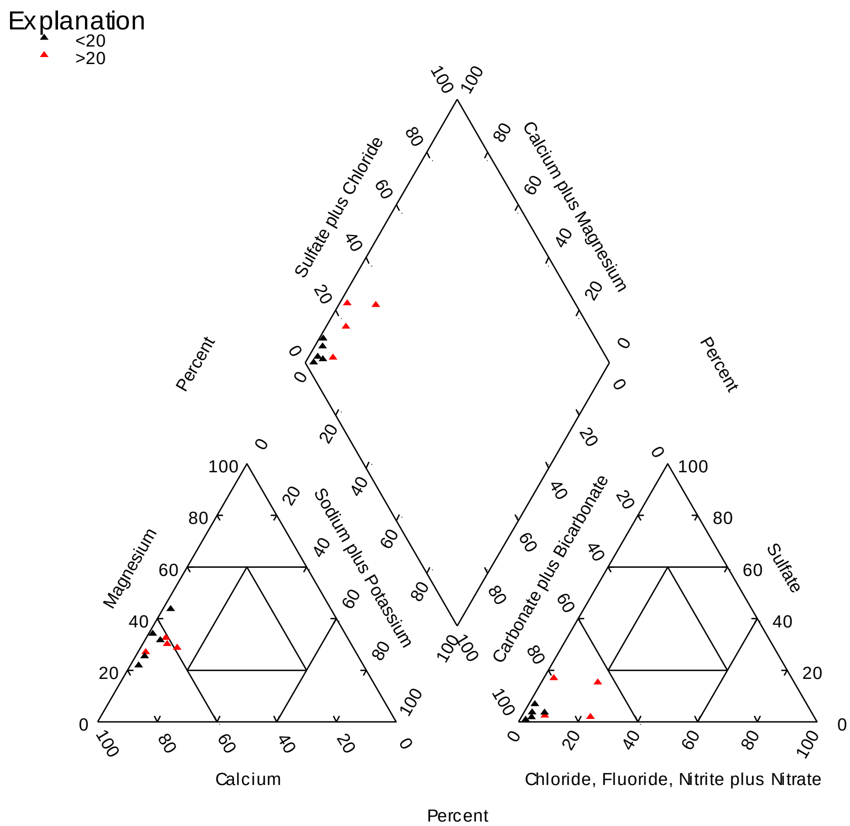
Note that if the value for any constituent missing (coded as NA), then the corresponding observation for the anions, or the cations and the overall cannot be plotted. For example, if only the value for carbonate is NA, then the anions can be plotted, but there will be no symbol for that observation in the cations ternary diagram or in the overall graph.

The **Symbology** page allows the user to change the symbol or its color. By default, the symbol is a black filled up-pointing triangle. The symbols are 0, open square; 1, open circle; 2, open up-pointing triangle; 6, open down-pointing triangle; 15, filled square; 16, filled circle; 17, filled up-pointing triangle; and 18 filled diamond. The user can also select a column in the data set to control the symbol. If a column is selected and if that column is numeric, then the symbol number is used, otherwise the factor level is used. A factor will be created automatically if the column is not already of type factor. The colors refer to the **User Colors** in the **Graph Sheet** window, which is available by right-clicking in the margins of the graphsheet. Colors are always converted to numbers and if the column is a factor, then the level is used. The **Piperplot** window below shows how to vary the symbol color by the column Depth.



If the symbol (**Marker number**) or the color is set by a column and not both, then an explanation is generated and automatically plotted. No explanation is created if only a single color and marker symbol is used or if both are set by a column.

The **Labels 1** page has options for the axis titles for the ternary diagrams. By default, they show all of the constituents listed on the **Data** page. If any constituent is not actually plotted, then it must be removed from the title. The **Labels 2** page has options for the secondary axis titles for the overall graph and an option for the **Main Title**, which would be placed at the top of the diagram. This page is not shown in this example. The resulting Piper diagram is shown below. Note in the example shown below that the color number 2, for depth ">20," is red but color number 2 can be different on any given graphsheet.



REFERENCE:

Piper, A.M., 1944, A graphic procedure in the geochemical interpretation of water analyses: Transactions of the American Geophysical Union, v. 25, p. 914-923.