Downloading USGS Streamflow Data from Internet into ArcGIS

Prepared by
Venkatesh Merwade
School of Civil Engineering, Purdue University
vmerwade@purdue.edu

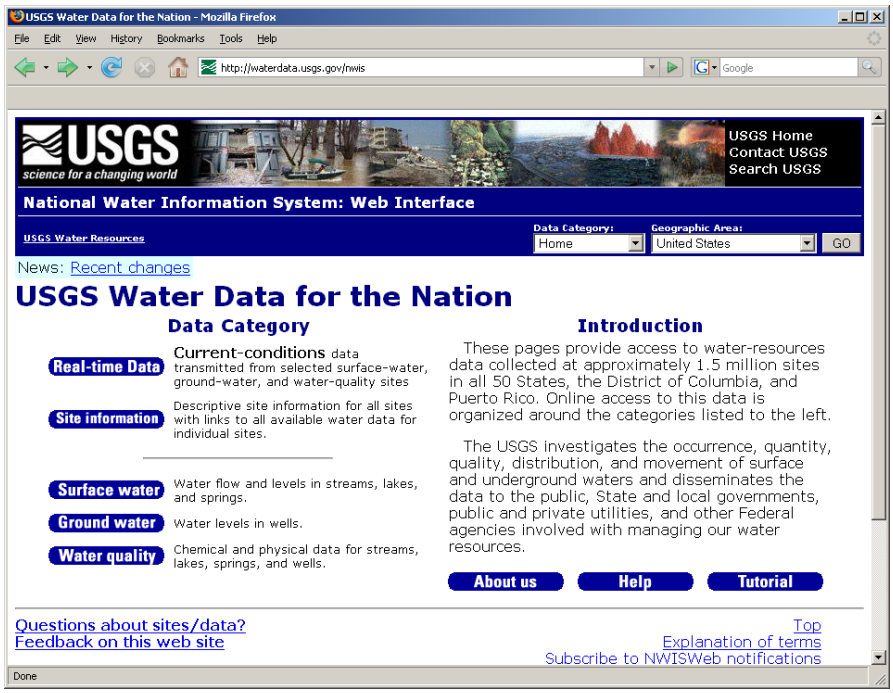
October 2012

Purpose

The purpose of this exercise to demonstrate: (a) the steps involved in downloading streamflow records for a gauging station in Indiana; (b) how to import the location of the station and its associated time-series data into ArcGIS; and (c) plotting and animation of time series data in ArcGIS.

USGS NWIS Web Interface

To download streamflow data, you need to **go to National Water Information System (NWIS) web interface** at the following link:
http://waterdata.usgs.gov/nwis

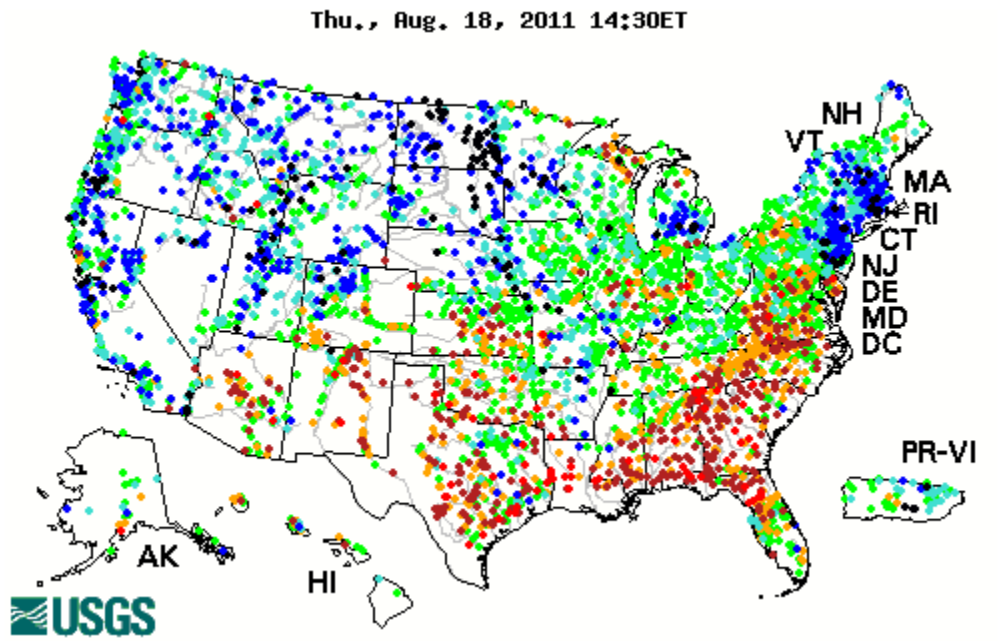


USGS cooperates with other federal, state and local agencies to maintain approximately 7700 streamflow gauging stations in the United States. The stage at most stations is recorded at a fixed interval (5 – 60 minute) which is then converted to streamflow

through stage-discharge rating curves. These fixed interval data are available under the Real-time Data link on the NWIS website. All fixed interval recordings are averaged to get daily average values that are available under the Surface water link on NWIS website.

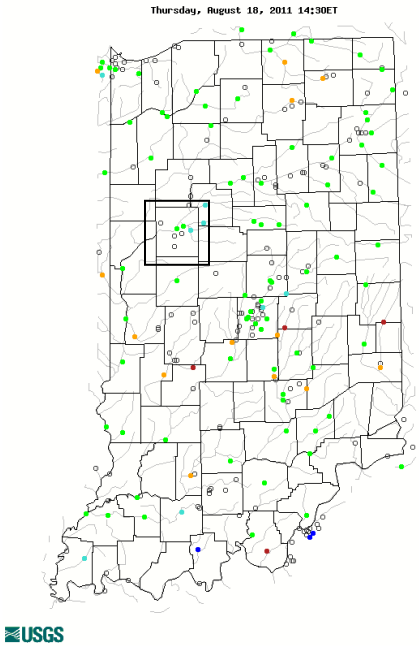
Exploring and Downloading Real-time/Daily Data

Click on Real-time Data button, and you will see the following map with colored dots on it.

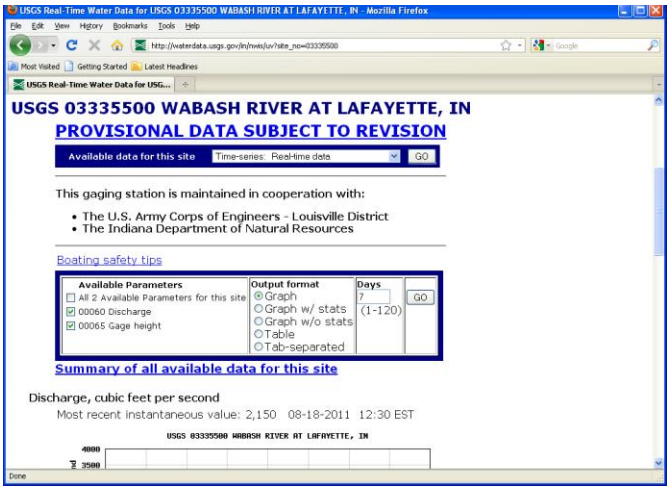


Each colored dot represents a USGS gauging station with at least 30 years of streamflow data. A color is assigned to a dot depending on the current flow at that location in terms of percentile, which is computed by using data for the current day of the year for all years. For example, 20th percentile flow means it is equaled or exceeded 80% of the flow values in the dataset. For the above map, a dataset for each dot includes all values on the current day for all available years. The website gives information on the percentile associated with each color. Typically, a flow equal to or greater than 75th percentile is considered above normal (high), a flow between 25th and 75th percentile is considered normal (average), and a flow equal to or smaller than 25th percentile is considered below normal (low).

Besides just looking at the current conditions at each point, you can also download real-time data in several ways. One (interactive) way is to click on the national map, and it will give you a map at the state level depending on where you click. Lets click on any point in Indiana, and you should get all stations within Indiana as shown below.



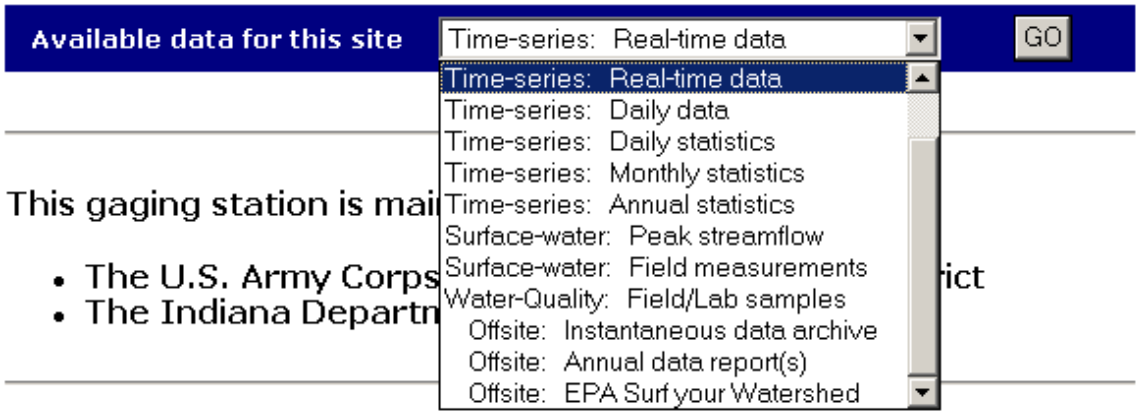
Move your computer cursor over Tippecanoe county (highlighted above), and you will see information associated with each dot in this county. If you find a station that is named "Wabash River at Lafayette, IN", click on it, and you can get access to the data associated with that particular gauging station.



You will also notice that each dot has a number besides its name. For Wabash River at Lafayette, the station number is 03335500. So each USGS gauging station has unique number associated with it. If you know the station number of a location for which you need data, you can use a quicker way to get the data instead of taking the interactive approach as is done in this exercise.

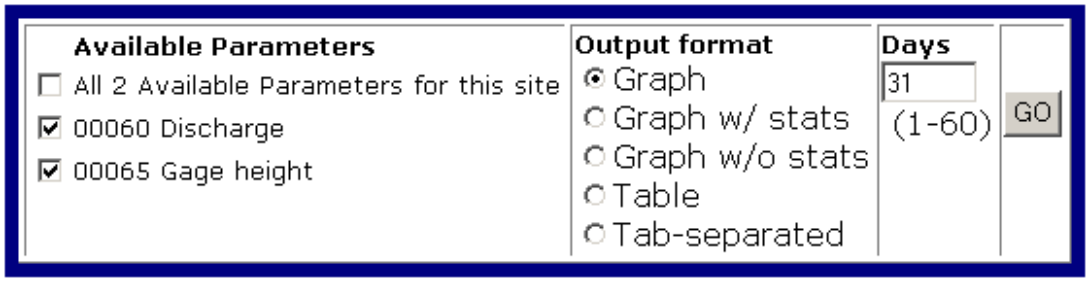
Because we started our search using the Real-time Data link, by default, you will see "Time-series: Real-time data" in the available data for this site bar. You can click on the

combo-box and choose any type of time series including daily data and other statistics as shown below.



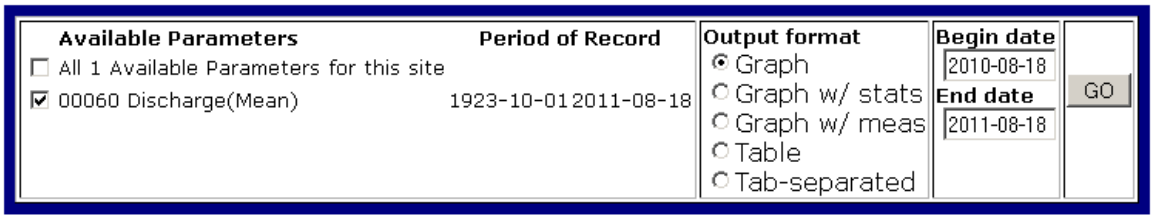
In this exercise we will only explore or look at real-time data, but will download daily data.

If you scroll down, you will see that by default seven days of discharge and stage data are provided as plots, but you can change the time span to up to 60 days and choose the output format using the options shown below (time is changed to 31 days with graph output).



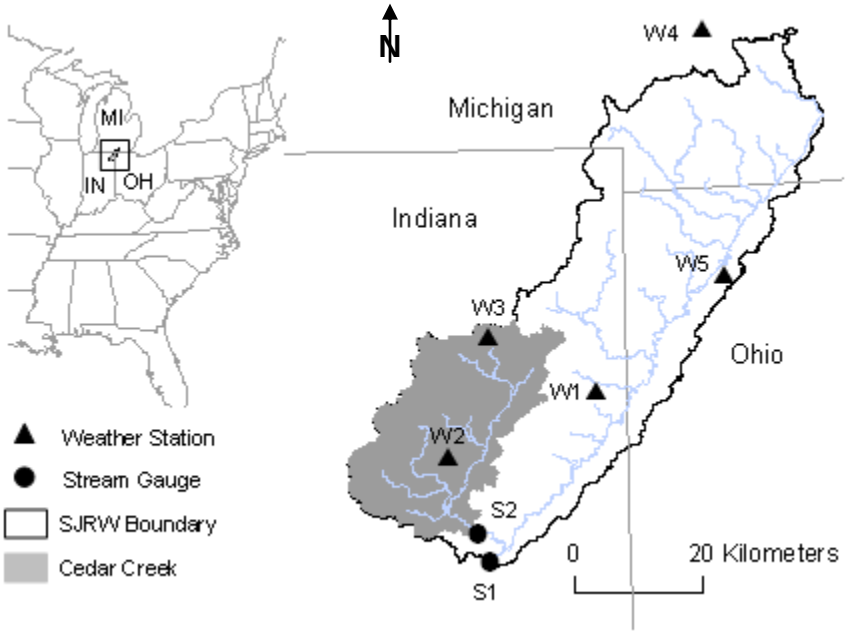
You can look at the graphs and see how flow and stage has changed at this station for last 31 days.

Next change the available data for this site to “Time-series: Daily data” and you will see that the download options also change as shown below.



Now you have the option of getting more data than just 60 days as with real-time data, and also notice that the discharge that you are getting now is mean (daily average of all short-interval real-time values). Again, you can select different output formats as shown in above options.

We are done exploring real time data for Wabash River near Lafayette. Lets download data for our study area which is St. Joseph River Watershed (SJRW) in northeast Indiana as shown below. We are interested in downloading data for streamflow station (S2) which serves as the outlet for Cedar Creek (one of SJRW sub-watersheds). S2 represents Cedar Creek near Cedarville, Indiana station with 04180000 station number as its unique identifier.



<u>Things to do on your own:</u>

1. Using the NWIS web interface download daily data for S2 from 01/01/2007 – 08/20/2009. (Hint: select the station by using "Site Name Identifier", and search the site name by choosing the "match any part" option)
2. Set the desired data range (check date format), and choose the output option as "Tab-separated data" save to file. (you can also display the tab separated data in a browser and then save it).
3. Also get the site description by using the "Site-description information" option, and report the following information for 04180000: HUC Number, latitude (decimal degrees), longitude (decimal degrees), geographic projection, drainage area and datum above sea level including reference.

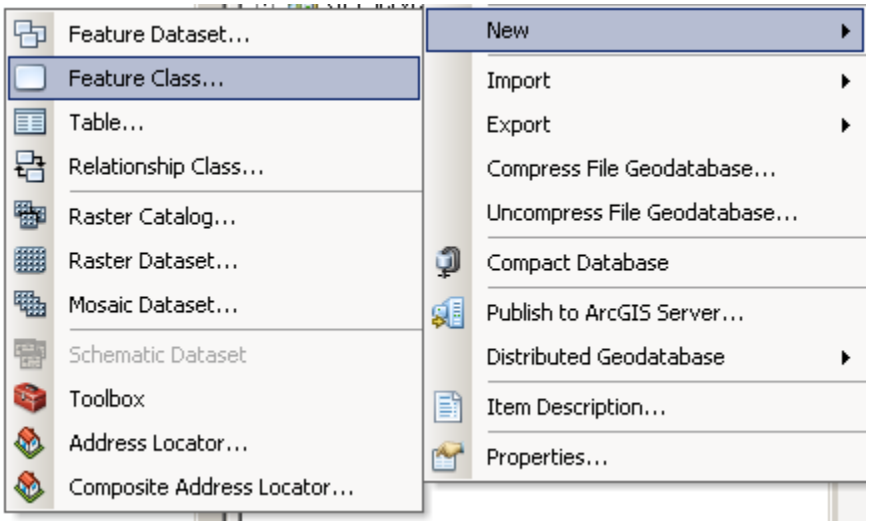
Save the excel file, and also save the sheet where you created the chart as a comma separated or a dbf file because we are going to use it later.

Bringing Data to ArcGIS – Creating geodatabase, feature class and table

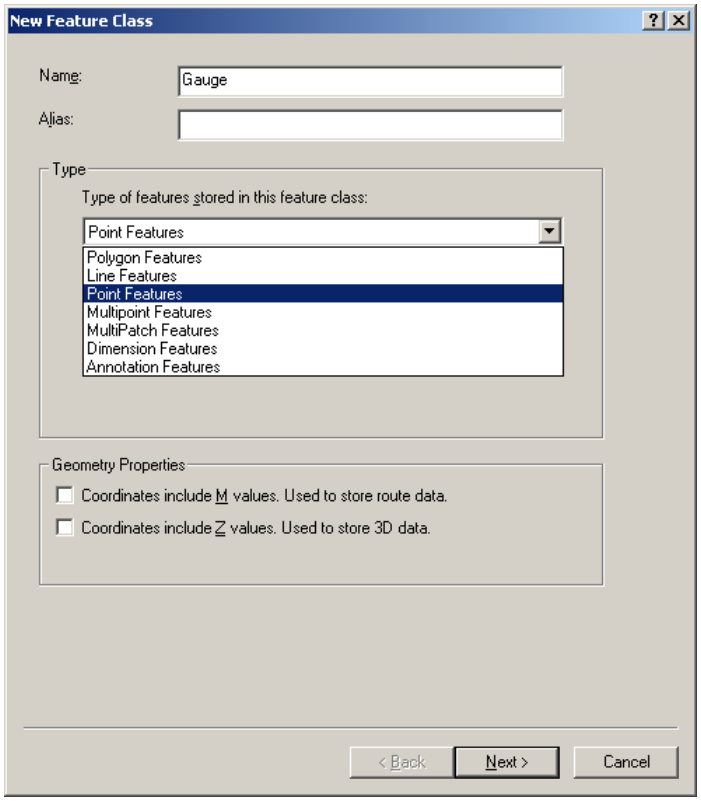
In this part, you will plot the location of 04180000 in ArcGIS and link the associated time series data with this point. There are several datasets available in public domain that will give location of all USGS gauging points in a GIS file (shapefile/feature class), but the objective here is to learn how to do this from scratch. The task of linking time series with a geographic location is a very useful task in day-to-day life of water resources researchers and professionals.

To accomplish this task of creating a point for a stream gauge and link time series data to it, lets create an new geodatabase using ArcCatalog with an empty point feature class and a table. **Open** ArcMap, and **save** it as *exe1.mxd* file in your working directory. In the Catalog window (if the catalog window is not open in ArcMap, click on the catalog window button). In your working directory, create an empty personal geodatabase named "exe1.mdb". In exe1.mdb, create a feature dataset named Data. Right click on exe1.mdb, and **Select** *New→Feature Dataset*. **Name** the feature dataset as "Data". **Assign** North American Datum 1927 geographic coordinate system (NAD 1927, which is same as 04180000) to Data feature dataset. For vertical coordinates use NGVD 1929 similar to 04180000.

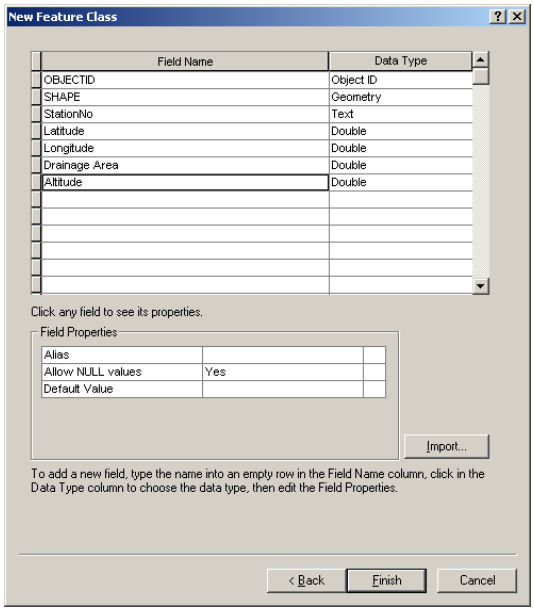
Next, **right click** on Data feature dataset, and create a new feature class as shown below.



Name the new feature class as "Gauge" and make it a point type feature class as shown below:

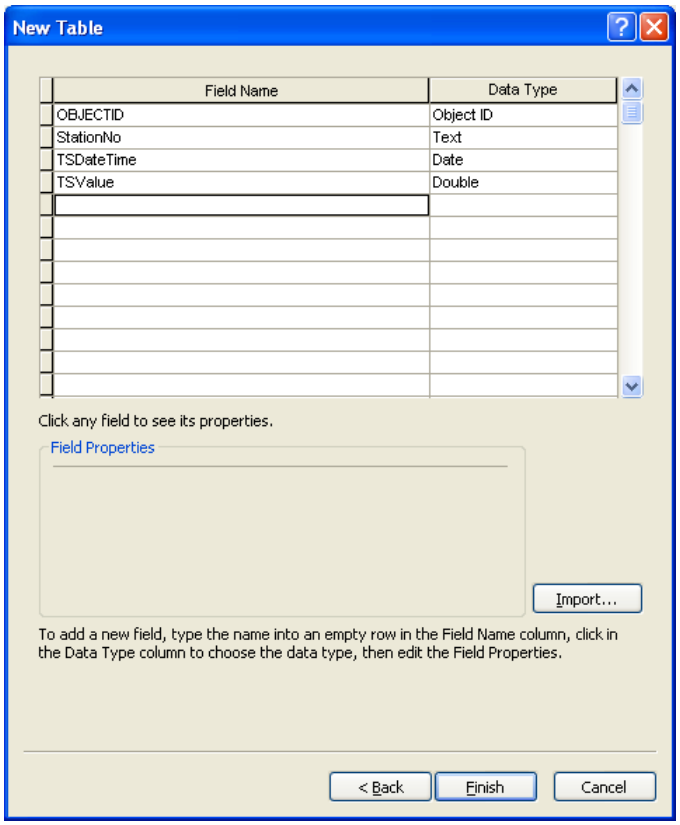


Leave include M values and Z values unchecked for this feature class. **Click** *Next*.
In the next window create some new fields to store information associated with a gauge such as its station number, location, name, etc. So create the following fields with the corresponding data types:



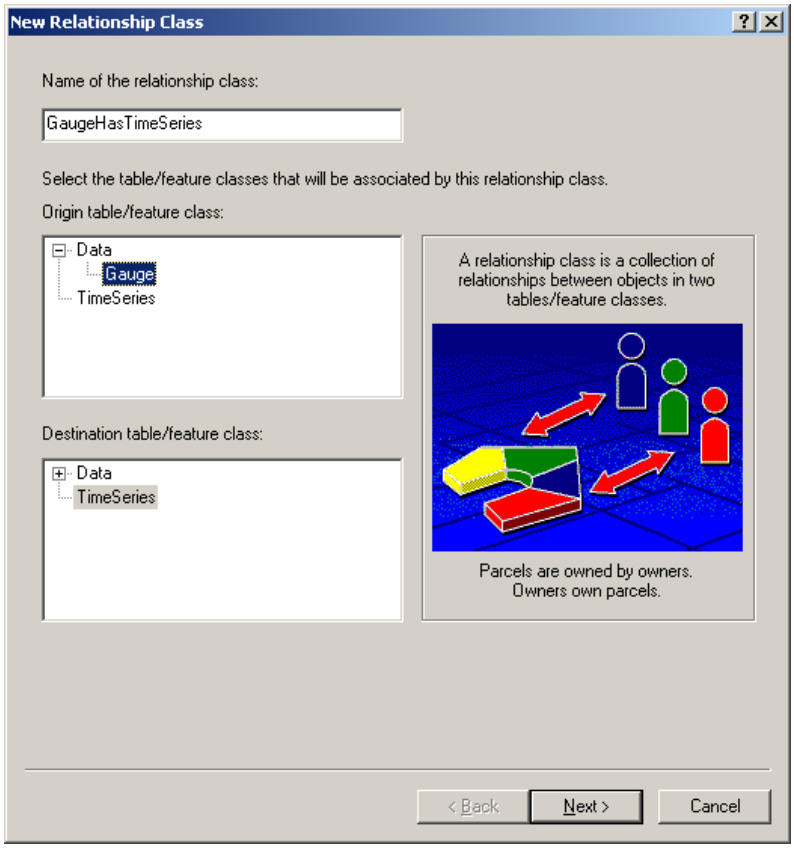
Click *Finish*.

Next, **create** an empty table in exe1.mdb, and name it as "TimeSeries". Follow same steps as above, but instead of creating a feature class, create a table by right-clicking on exe1.mdb (not on Data feature dataset). You will also realize that when we create a table, we do not specify its type (point, line or polygon), or coordinate system because we do not store any geographic data in a table. Define some fields for TimeSeries table as shown below:



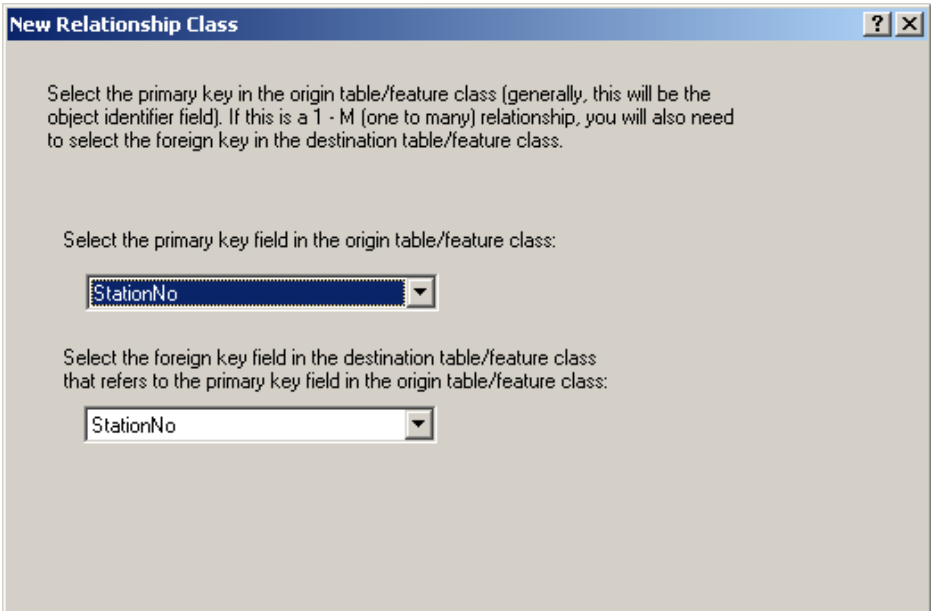
Click Finish. One last thing before we are done with creating data is to define a relationship class that will link Gauge with TimeSeries.

Right click on *exe1.mdb*, and **create** a new relationship class named "GaugeHasTimeSeries". **Select** Gauge as origin table and TimeSeries as destination table as shown below:



In the next few windows, **select** the following options: simple peer-to-peer relationship, no messages are propagated, one-to-many relationship (one point will have many discharge values), no attributes for the new relationship class.

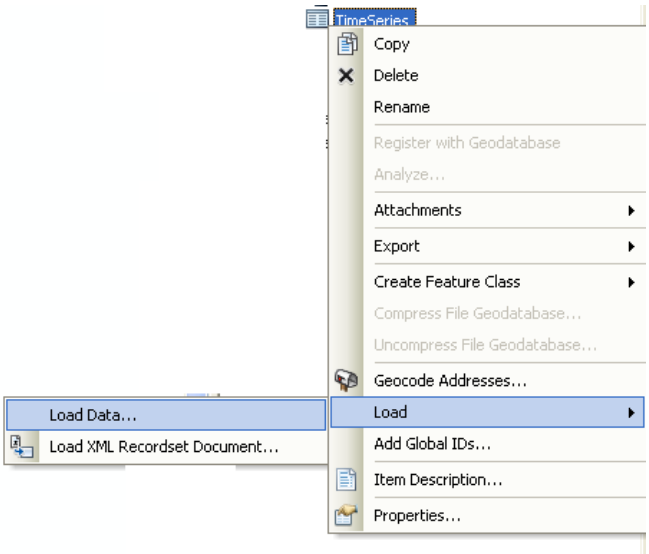
Finally, use StaionNo as the primary key to link these two tables.



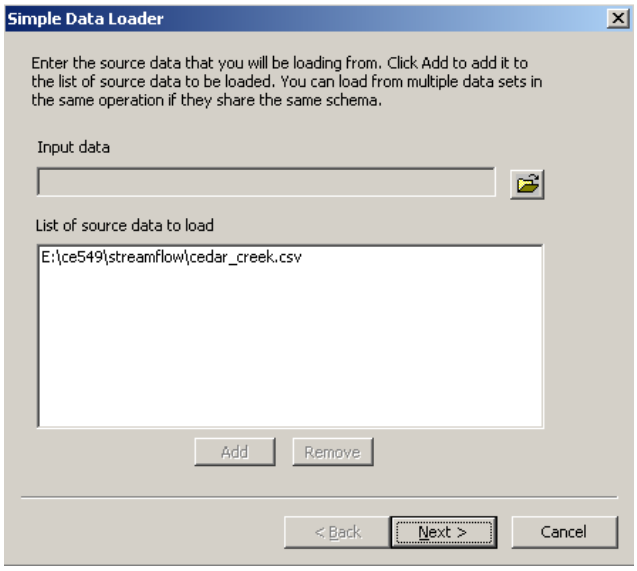
Click *Next* and then *Finish*. Your final geodatabase should look something like below if you use the same names as used in this exercise.

Because the discharge data we need for TimeSeries already exist in the comma separated file, we can load that data into TimeSeries in ArcCatalog.

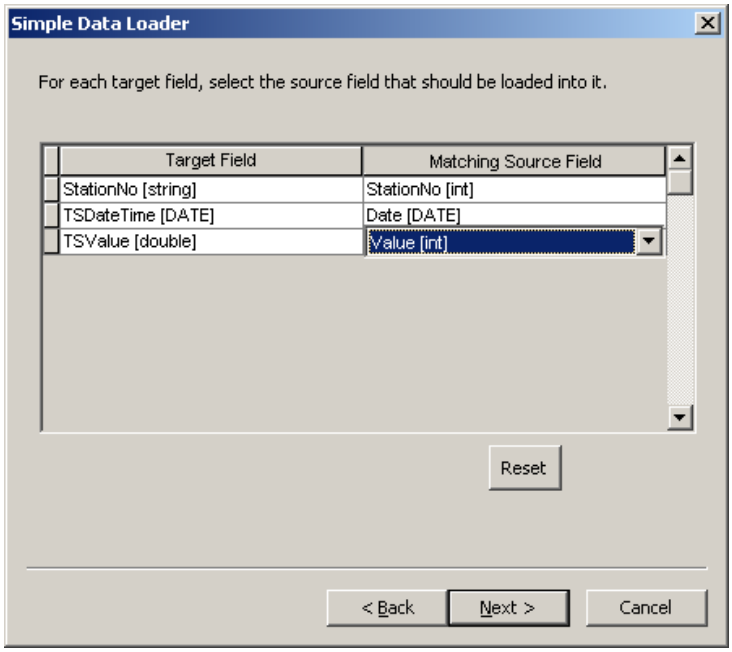
Right click on *TimeSeries*, and **select** the *Load* data option as shown below.



Browse to the CSV/DBF file, **click** *Add* and then **Click** *Next*.



Click next on the subsequent window, and then match the following fields as shown below.



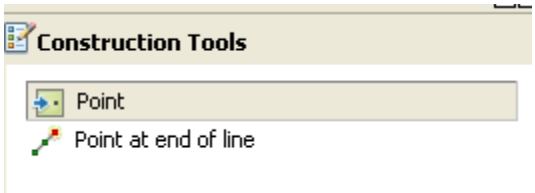
Click *Next* and then *Finish*. Now that we have data in TimeSeries, the next step is to create a point for the gauging station and explore this link through the relationship class that we just created. Very cool! **Close** the ArcCatalog window.

Populating Gauge data in ArcMap

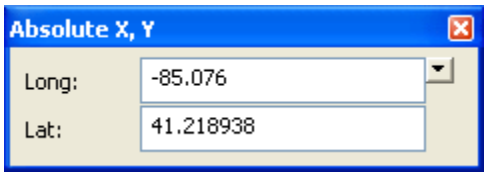
Your map document should already have the Gauge featureclass and TimeSeries table added in the table of contents view. If not, **add** Gauge and TimeSeries to the map document using the Add data button, or by going to the file menu. **Save** the map document.

Open the editor by pushing the editor toolbar button . To create a point for the gauge, **click** on the *Editor* toolbar and **Start** Editing.

Once the edit session is started, an editor window will be added to the map document that will show the features and the construction task. Select the Gauge feature class in the editor window, and the construction tasks associated with this feature will be listed in the Construction tools as shown below.

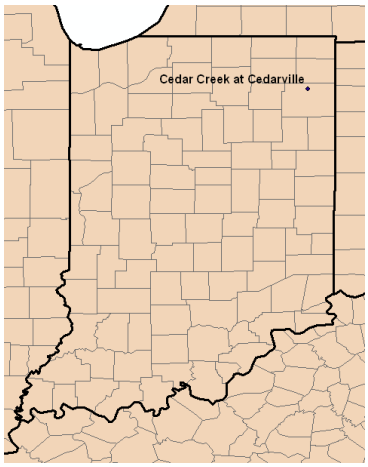


Select the Point construction tool, and bring the cursor to the map document. Now, to add a point by using the lat-long of the gauging station, **press** *F6*. You will be asked to enter the absolute X and Y for the new point. For X, enter the longitude for 04180000 with a negative sign (western hemisphere) and for Y, enter the latitude for 04180000 (both in decimal degrees), and **press** enter on your keyboard.



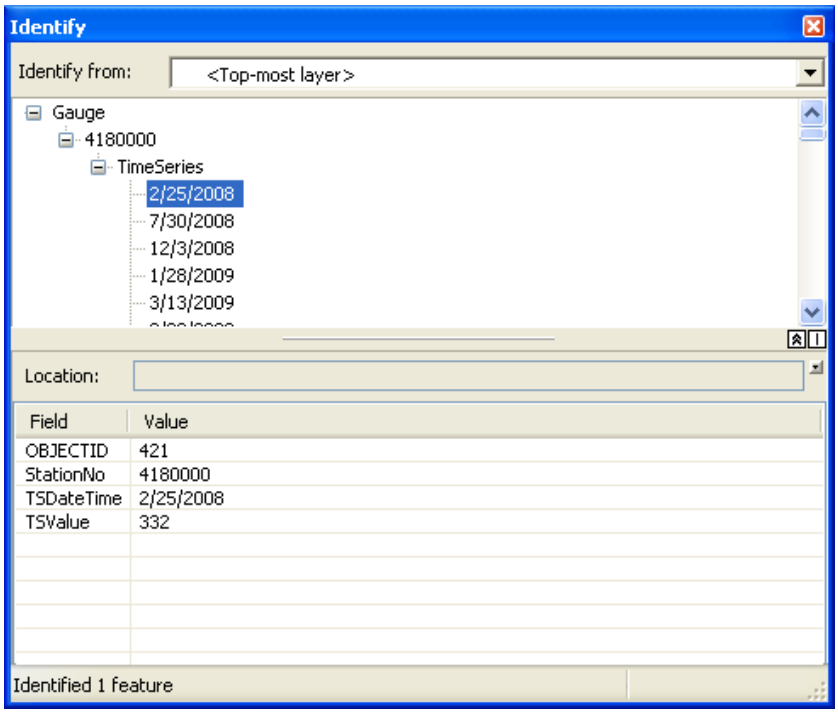
You will see that a new point is created at the specified location on the map document which is stored in Gauge feature class. Now open the attribute table for Gauge (by right-clicking), and populate all the fields using the information for 04180000. Close the attribute table, save edits by clicking on editor toolbar, and stop editing.

What you just did is very cool and also learned several GIS functionalities. We can now identify in GIS where 04180000 is located. To make this location relevant, you can add other GIS layers such as counties, state boundaries and stream lines that are provided in Program Files\ArcGIS\Bin folder (in TemplateData.gdb). If this folder is inaccessible on your computer, get these data from the National Atlas available at http://www.nationalatlas.gov/. You can see that the point we have created is in Allen County, Indiana. You can also add other layers such as roads, cities and rivers to get a good reference of this point.



Ok, now that we are confident about the location of our gauging station, the next step is to see how Gauge and TimeSeries are linked together.

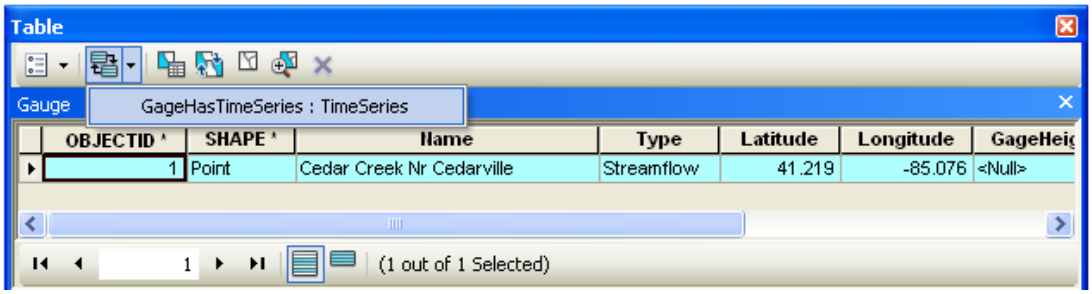
Use the identifier button, and click on the gauge. You will see the identify window, that will display all information about gauge attributes as shown below.



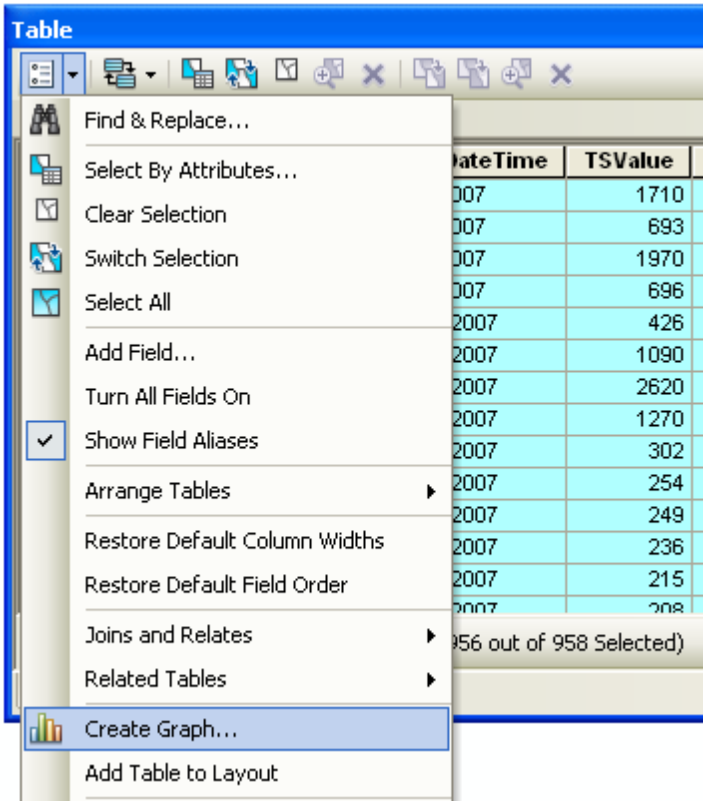
If you **expand** the box (click on +) next to Cedar Creek at Cedarville, you will see that TimeSeries is associated with this point, and if you expand TimeSeries, you will see all the data associated with this point. When you click at any number (ObjectID) associated with the TimeSeries, you will see the information associated with it such as date and streamflow!

Plotting of TimeSeries Data in ArcMap

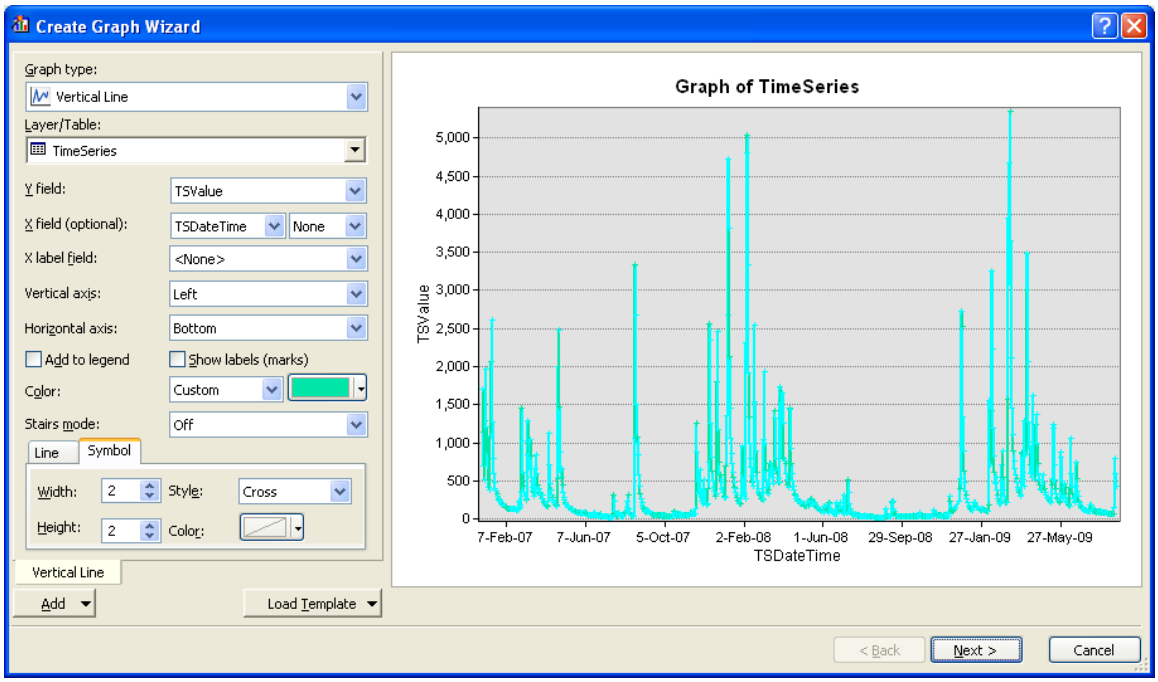
You can even plot the data associated with a gauging point by selecting the point, and getting access to its related time series. Open the attribute table of the Gauge feature class, and select a feature (we have only one). Next, click on the Related Table button in the attribute table menu, and click on the GageHasTimeSeries: TimeSeries button as shown below



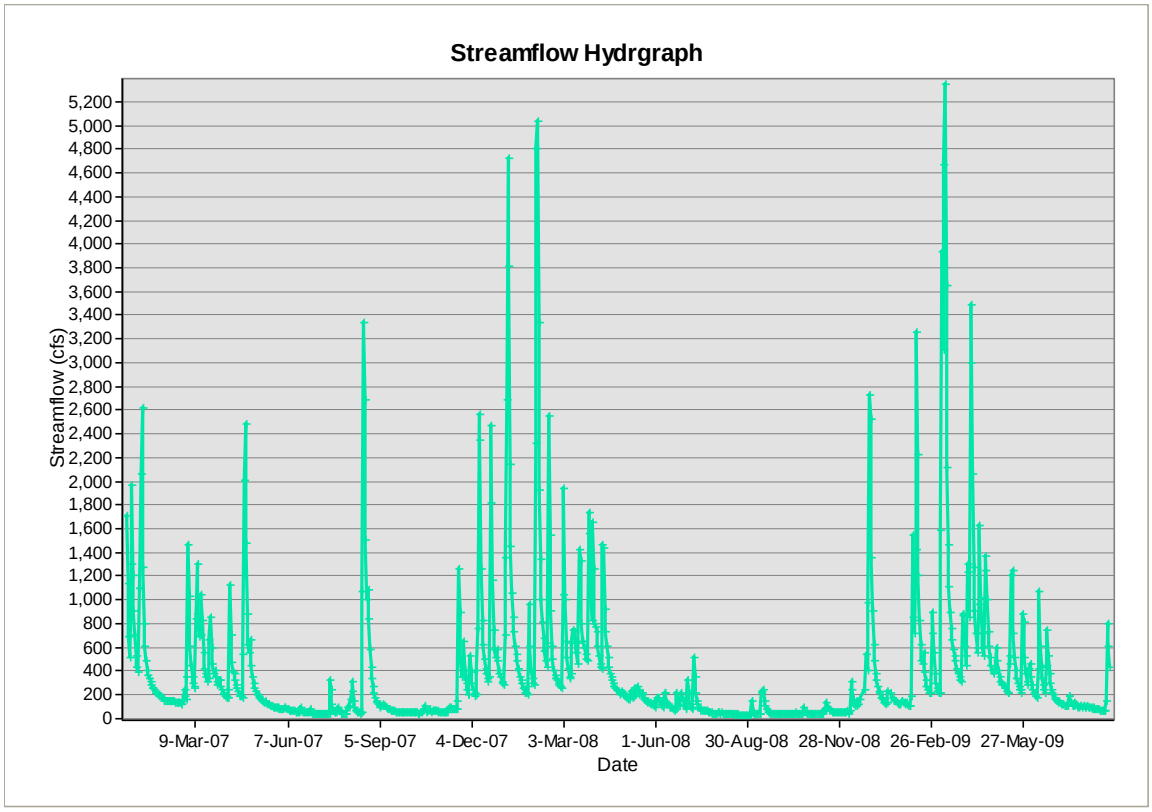
This will open the TimeSeries table with the time series values related to the selected point as selected rows. Now you can plot the selected rows in a graph as hydrograph by clicking on Table Options→Create Graph as shown below.



In the create graph wizard, choose the Y field to e TSValue and X field to be TSDateTime.

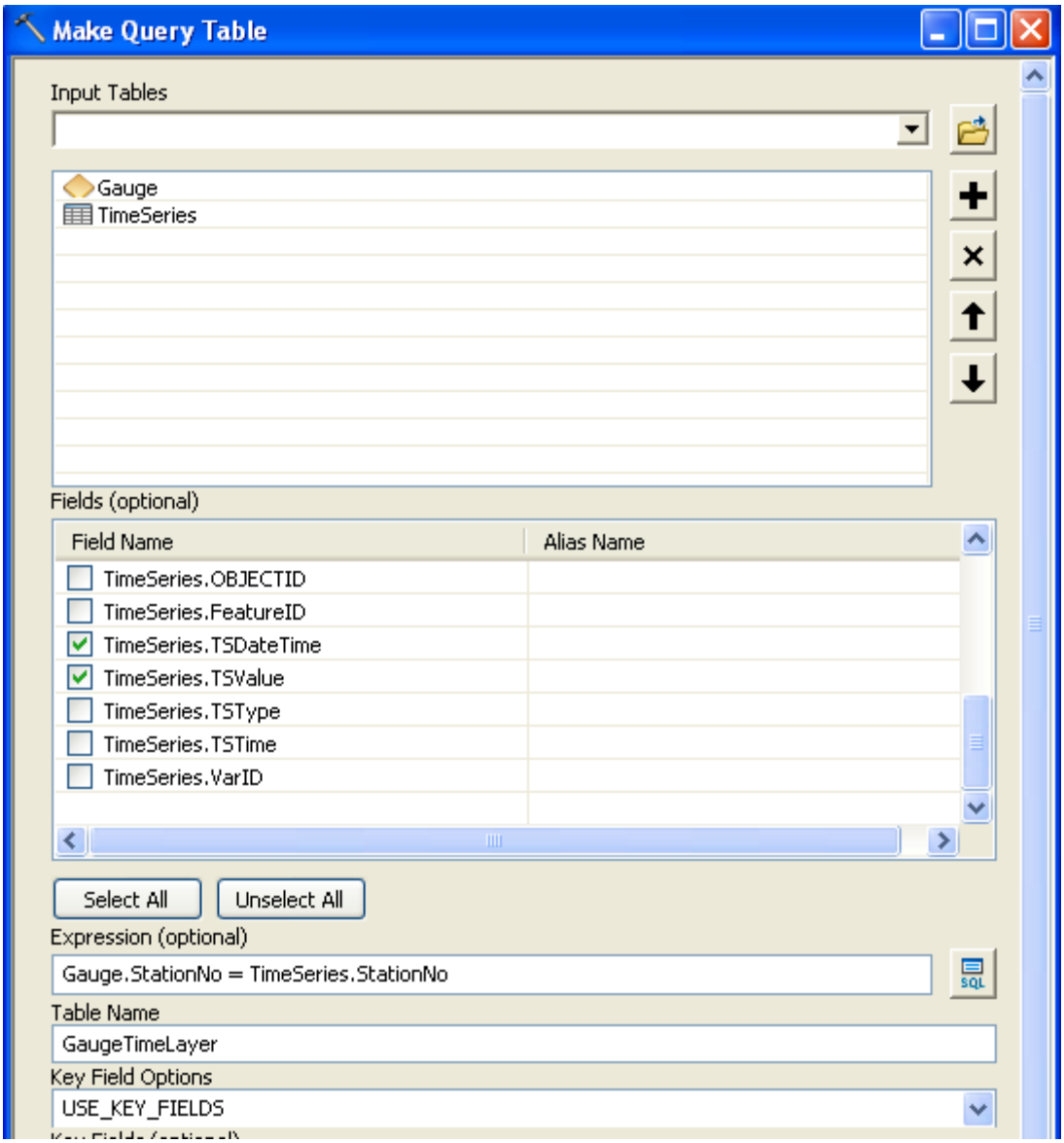


Change other variables as needed, add labels, title, etc, and create the graph.



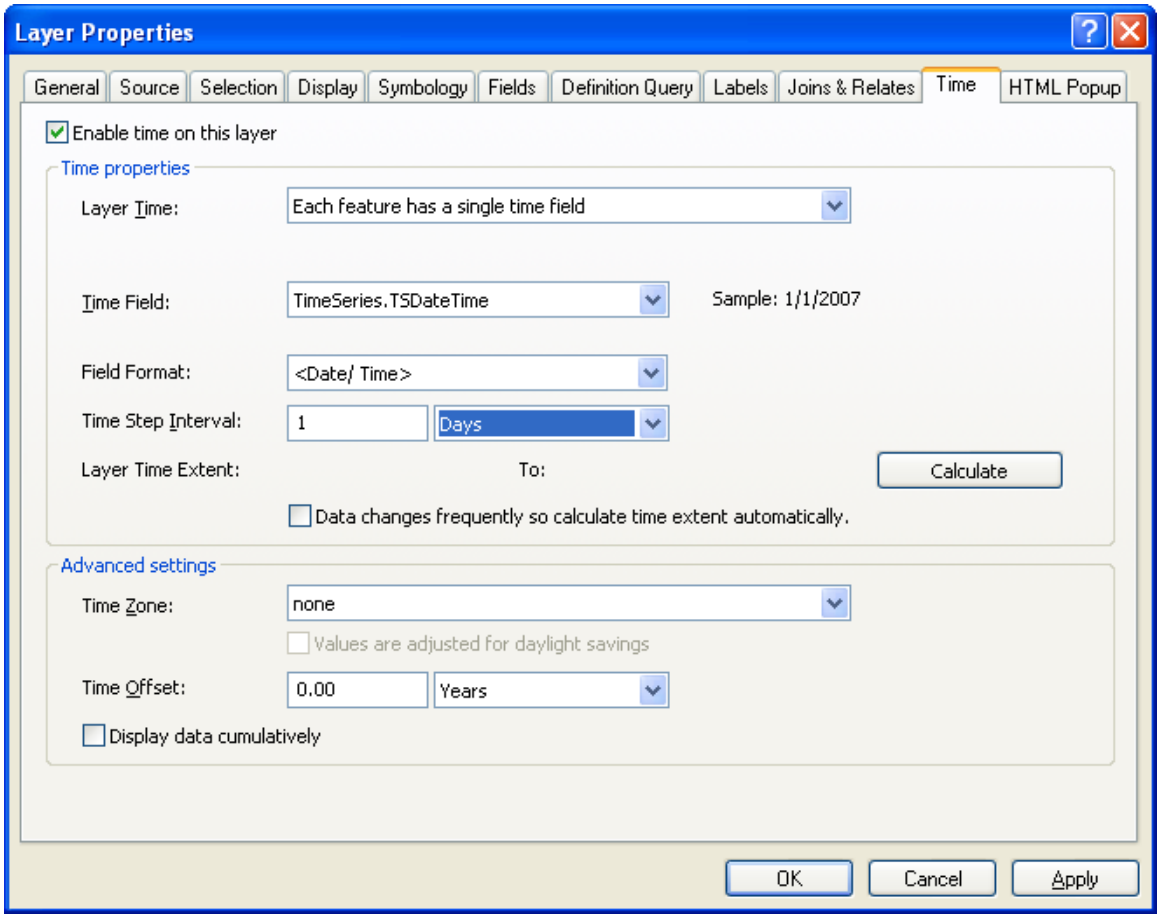
Creating Time Enabled Layer in ArcMap

By using the link between GIS features and related time series, we can also create a time enabled layer in ArcMap and produce animations. To create a time enabled gauge layer, lets first join the gage and time series through a query table. Open ArcToolbox, and select Data Management Tools→Layers and Table Views→Make Query Table. Add Gauge and TimeSeries as the input tables. For fields, make sure to select the StationNo and SHAPE in the gauge feature class, and TSDateTime and TSValue in the TimeSeries table. For the expression, use the SQL query button to equate the related fields in Gauge and TimeSeries (StationNo in our case as shown below). Name the output table as GaugeTimeLayer, leave the other defaults unchanged, and click OK.

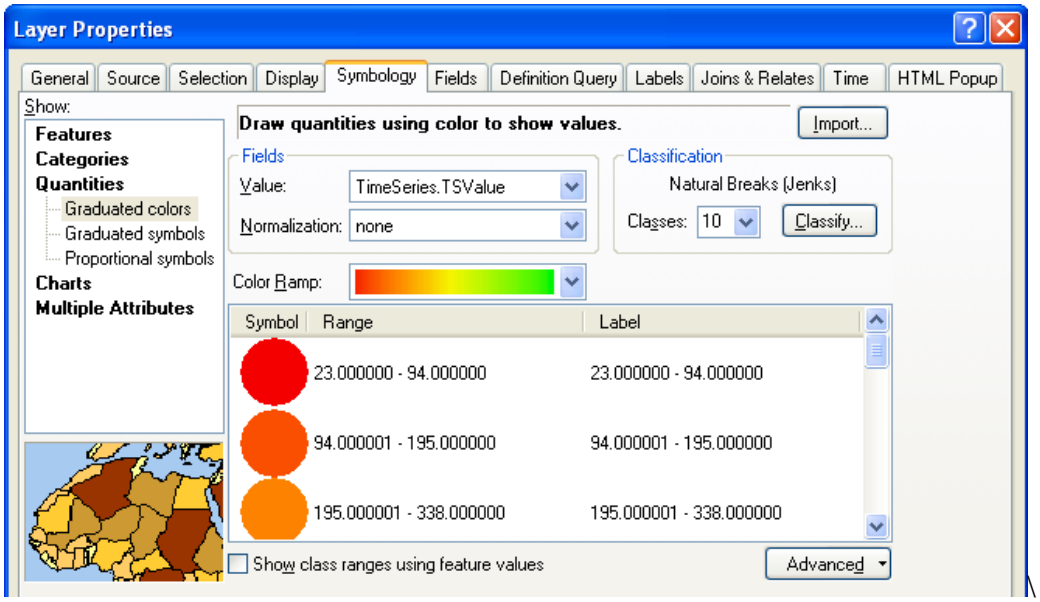


This will create a new temporary feature class named GaugeTimeLayer, and add to the map document. If you open the attribute table of this new layer, you will see that there is a point created corresponding to each related time series record for a gauge point.

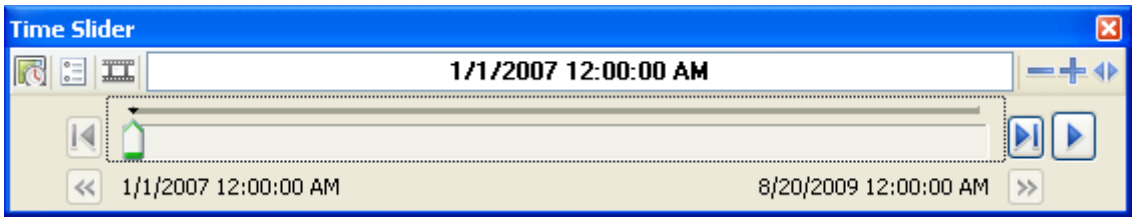
Next, go to GaugeTimeLayer properties, and select the Time tab. In the Time tab, check the Enable time on this layer box, make sure the right time field is selected, and change the time step interval accordingly (we have daily streamflow data so selecting on day is OK) as shown below.



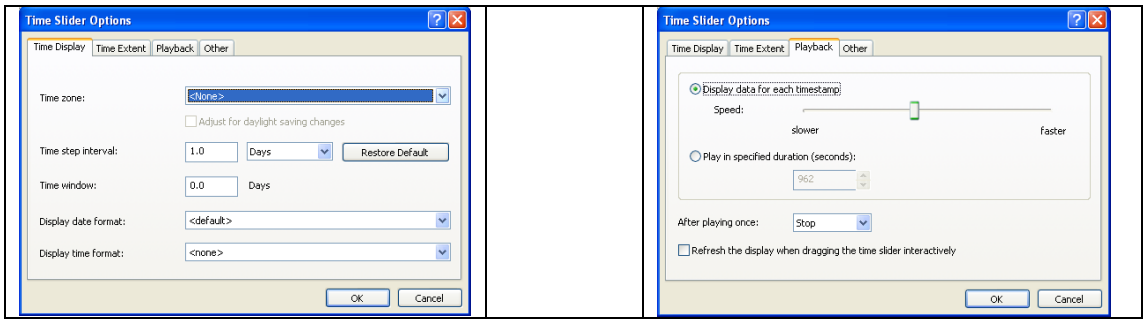
Next, select the Symbology Tab, and modify the symbology to have different color or shape for the point based on TSValue by using graduate colors or symbols.



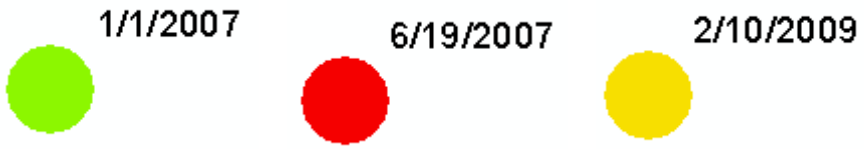
You can also label the feature to show the TSValue. That way you will see both the values and symbol as we navigate through time using the new layer. Once this is all done, the time slider tool  on the Tools toolbar will be activated. Click on that to see the time slider bar as shown below.



Click on the Options button on the time slider bar, and make sure the time step interval matches with the data, and change the playback speed appropriately so that the playback is not too slow or fast as shown below.



Now hit the play button on the slider bar, and see how the symbology of the point changes based on streamflow value for different days as shown below.



What you just learned is a powerful functionality in GIS to link geospatial and temporal information together that can be used for data analyses, visualization and animation!

Things to do on your own and turn-in:

1. Download data for St. Joseph River Near Fort Wayne, IN (#04180500) from 01/01/2010 – 12/30/2011. Bring the data into Excel. For these data, compute and **report** average, standard deviation, maximum and minimum flow values.
2. For the data you have downloaded, **report 10th, 25th, 50th, 75th and 90th percentile flow**. You can either used Excel functions to compute these percentiles, or you can do this through flow duration curve (graphical method).
3. In the gauge feature class create one more point for St. Joseph River Near Fort Wayne, IN (#04180500).
4. Store daily records at this point in the TimeSeries table you just created.
5. **Turn-in a screen shot** of ArcMap with locations of the two gauging stations (label them with their names) along with some reference layers.
6. **Create and turn-in screen-shot of time-series plot created in ArcMap for the new station.**