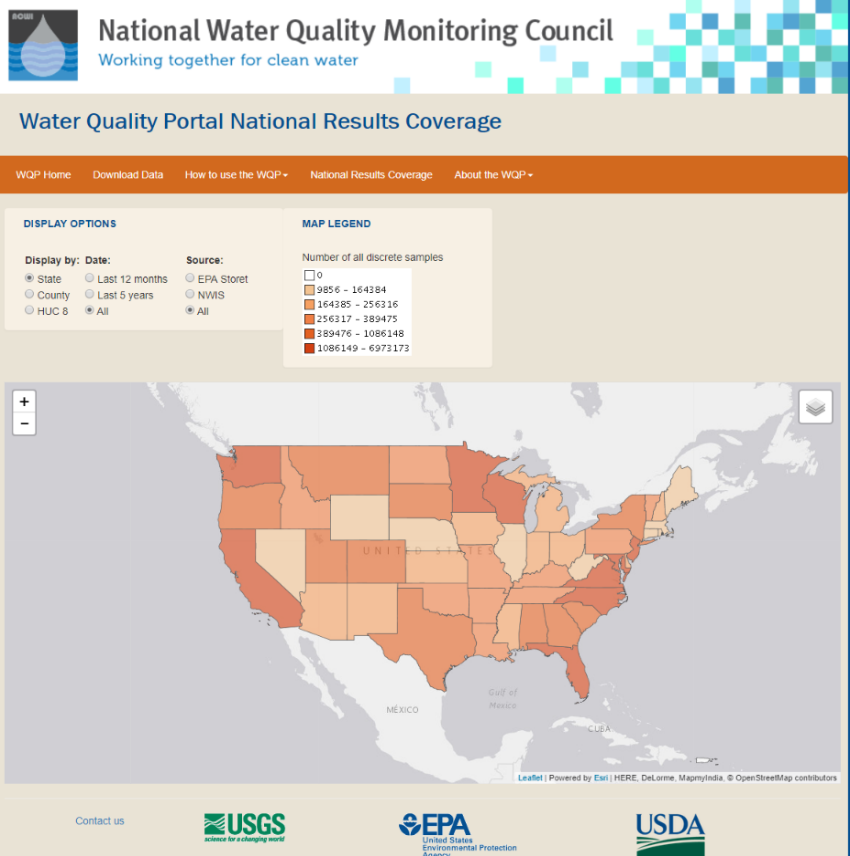
Unit 3: Water Quality

# FOCUS OF THIS LESSON: Data Visualization

# Tools: Tableau, with intro to Python Notebooks, and Shiny R

# What data do I need? & Do the data exist

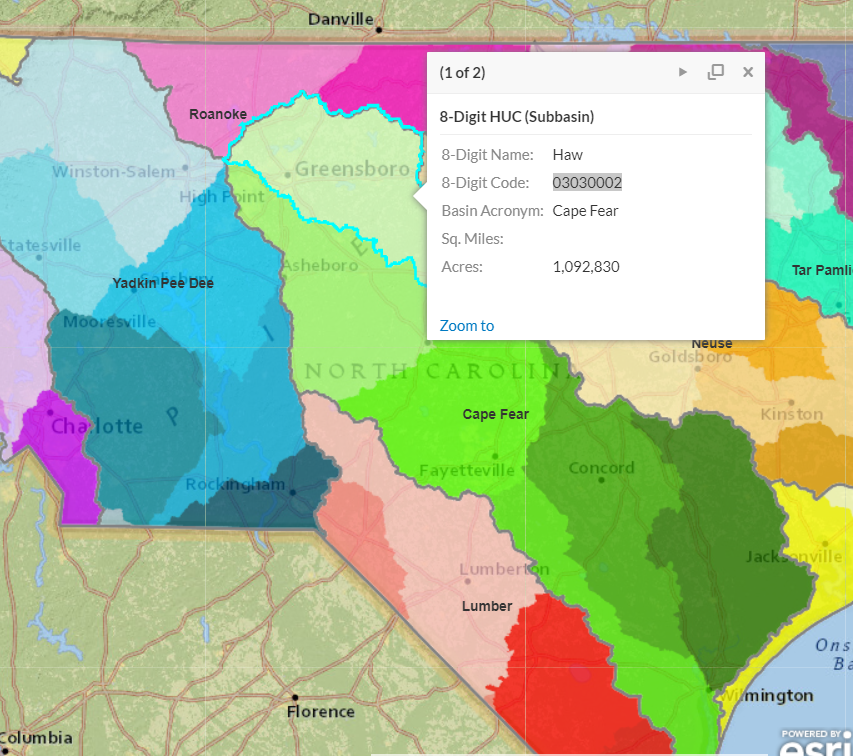
1. Where to find water quality data: the water quality portal
   1. The Water Quality Portal uses the water quality exchange data standards to integrate publicly available water quality data from the USGS NWIS, EPA STORET, and the USDA STEWARDS databases. Other data are included from state, federal, tribal and local agencies. It’s a great first place to go to find data.



**Figure:** Snapshot of the number of water quality samples in the portal by state

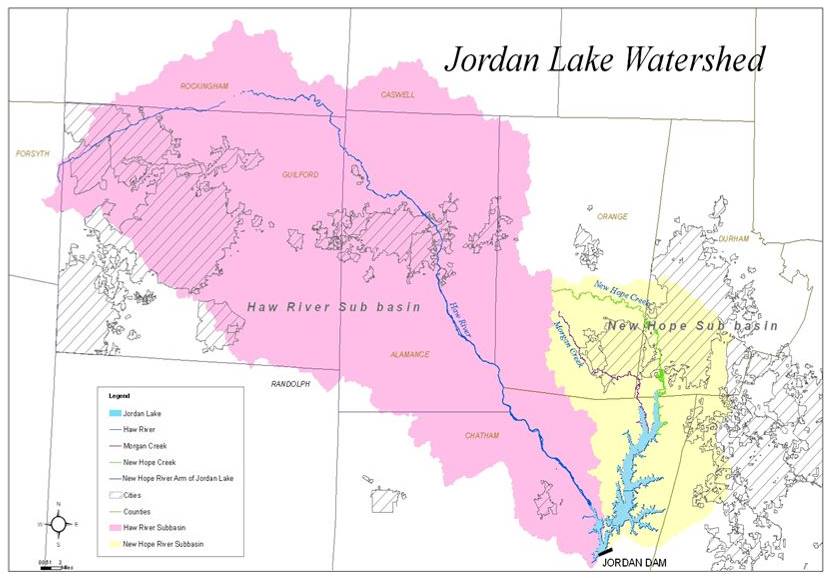
* 1. On the download data page you can search for data based on location, parameters, characteristics, etc. To search for nitrogen contributions upstream of Jordan Lake, it would be helpful to know the HUC.

1. Find your HUC8 upstream of Jordan Lake. The North Carolina Department of Natural Resources has put together an arcgis platform for sharing data.
   1. <http://data-ncdenr.opendata.arcgis.com/>
      1. Haw: 03030002



**Figure:** Snapshot of NC DENR arcgis maps highlighting the Jordan Lake watershed.

1. Download the necessary water quality data. We are going to be interested in nutrients, particularly nitrogen.
   1. Narrow the search by listing the following:
      1. State 🡪 North Carolina
      2. Site Type 🡪 Lake, Reservoir, Impoundment and Stream
      3. HUC: 03030002
      4. Sample Media: Water
      5. Characteristic Group: Nutrient
      6. Date Range: 01-01-1970, 12-31-2017
      7. Minimum results per site: 500
   2. **MAKER SURE TO KEEP NOTES OF YOUR DOWNLOAD SELECTIONS FOR REPLICATION PURPOSES**
   3. Click show sites on map (n = 91)
      1. Download sites data as a comma separated file
      2. Download Sample results (physical/chemical metadata) as a file
2. What are the thresholds we will be assessing for North Carolina and Falls Lake? It’s not so easy to find this information.
   1. <https://www.epa.gov/nutrient-policy-data/progress-towards-adopting-total-nitrogen-and-total-phosphorus-numeric-water>
   2. NC: <https://www.epa.gov/sites/production/files/2014-12/documents/nc-classifications-wqs.pdf>
   3. <https://deq.nc.gov/nc-stdstable-09222017>
   4. <http://www.orangecountync.gov/document_center/PlanningInspections/Fallsrulesfactsheet.pdf>
   5. It turns out some states have nitrogen limits. In North Carolina there is a nitrate nitrogen limit of 10 mg/l. The big news for both Jordan and Falls Lake centers around rules that have been passed regarding how much nitrogen (and phosphorus) are acceptable to enter the lake annually. For Jordan Lake, different arms of the lake have different acceptable annual loads:
      1. Upper New Hope: nitrogen load of 641,021 pounds/yr
      2. Haw River: nitrogen load of 2,567,000 lbs/yr



**Figure:** Sub-basins within the Jordan Lake watershed

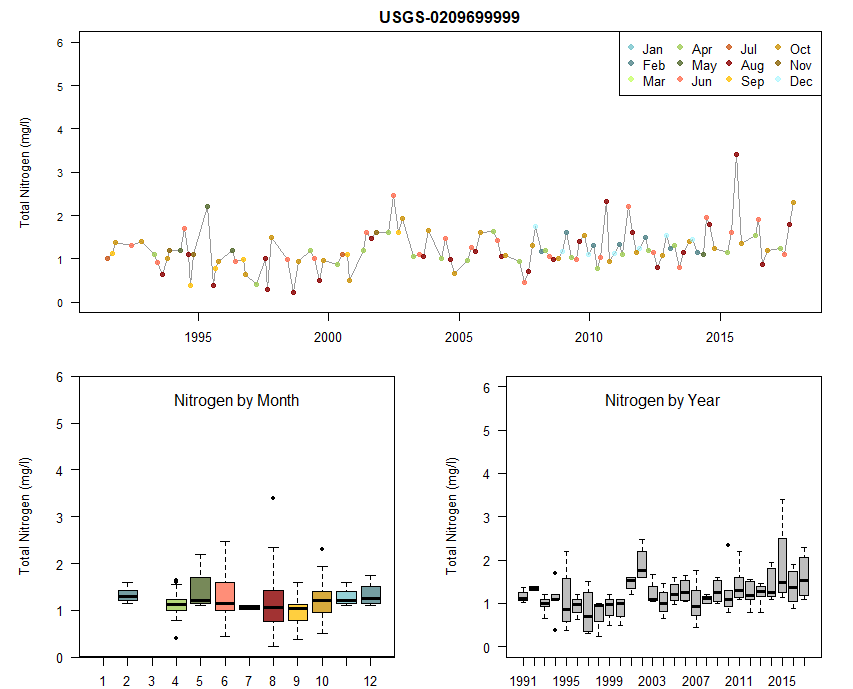
# Exploratory Data Analysis

## Find the data we need to use

1. Water quality data tends to have a **lot** of rows and columns over large areas. It is good to look at the data in excel, but here were will do the analysis in R and python.
2. Load in the site data and plot a map.
   1. Let’s only look at sites with a drainage area greater than 25 mi2
   2. What’s the problem you instantly notice? Many of the sites in Jordan Lake have disappeared because there is no data in that field. Let’s not filter sites that way.
3. Load in the results data
   1. Let’s clean up and filter the data so we are getting the data from routine samples and not from extreme events.
   2. We also found that nitrogen used for regulations include: nitrate, nitrite, ammonia and organic forms. There is a “Nitrogen, mixed forms attribute in the characteristics that incorporates all those forms of nitrogen. Filter the data to only include those results.
   3. We also want to make sure we have the total nitrogen, so filter the Results Sample Fraction Text to include “Total”.
      1. Remember – it’s always good to check to make sure your code is doing what you think it is doing. Use unique() to make sure the filters worked.
4. Detection Limits and Measurement Units
   1. You may have noticed that many sample sites state “not detected”. This is important data that are not currently being represented. Create a new column and set the value equal to the results, unless it is below the detection limit – in which case set it equal to ½ of the detection limit.
   2. You may also have noted that the total nitrogen was sometimes reported as mg/l or mg/l NO3. We want mg/l. Convert those values to mg/l
      1. The atomic weight of nitrogen is 14.0067 and the molar mass of the nitrate anion (NO3) is 62.0049 g/mole. Therefore to convert Nitrate-NO3 to Nitrate-N:
      2. Nitrate-N (mg/L) = (14.0067/62.0049)\*Nitrate-NO3 (mg/l)
5. Merge the sites data with the results data.
   1. Group info by start year, end year, and range
   2. Keep only those sites with more than 10 years of data that are still in operation
   3. Plot sites

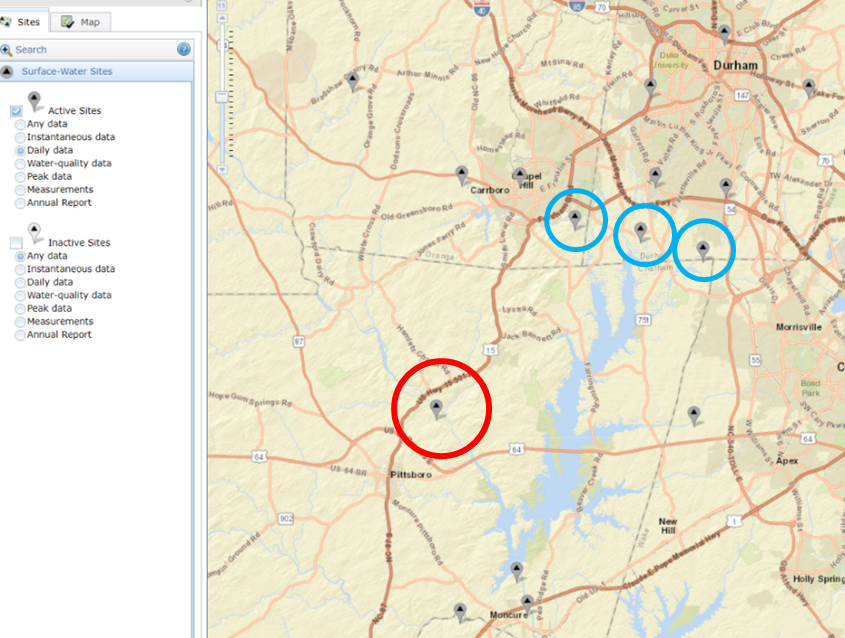
## What are the nitrogen trends over time at each site?

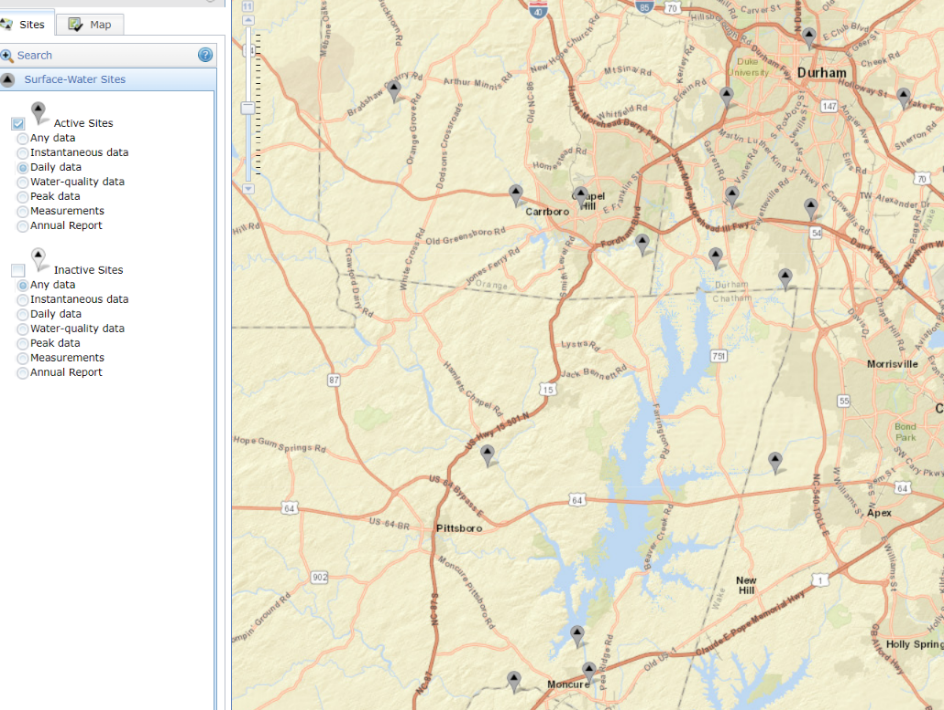
1. Some dates have multiple samples taken. Calculate the average nitrogen for each date within a site. Plot nitrogen over time.
2. Add two boxplots below. One showing samples by month, the other by year.



**Figure:** Nitrogen plots for the USGS site on the Haw River.

1. Plot the average nitrogen for each site in 2017.
   1. You can make a series of maps by looping through and plotting Nitrogen each year over time.
2. Now let’s look at how the Haw River and New Hope River thresholds are doing. The water quality data reports Nitrogen as mg/l. In order to convert to an annual load (lbs/yr), we need to know the volume of water flowing through each site. Go to the NWIS Mapper to find which USGS gauges are closest to:
   1. USGS-0209699999 (Haw River Arm)
   2. USGS-0209768310 (New Hope Arm)





**Figure:** NWIS mapper USGS gauges near water quality monitoring sites. The red circle is the gauge of interest for the Haw River Branch. The blue circles are the gauges of interest for New Hope Creek.

1. Use readNWISdv to download the daily discharge for the Haw River.
2. Plot the USGS site and the water quality site together to see how close they are and if any major tributaries intersect between them.
3. Convert flow to MGD and calculate the total annual flow.
4. Calculate the average annual nitrate (mg/l)
5. Merge the annual discharge and nitrate load together. Calculate the pounds of nitrate per year based on the flow (rough estimate).
   1. Pounds = MGD \* Average N \* lbspergal (water is 8.34 pounds per gallon)
6. Repeat this process for the Upper New Hope Creek.
   1. Take the sum of the flow for all three sites over a year.
   2. Notice that the lbs are far below the threshold. Why might that be?
      1. The drainage area for Jordan lake is 1689 mi2
      2. The Haw River site accounted for 1,275 mi2
      3. The three sites on for New Hope Creek accounted for 76.9, 21.2, and 41 mi2
         1. The downstream site on the right accounts for 12 mi2
      4. This leaves 402 mi2 unaccounted for
   3. Let’s assume that 75% of this is being missed and should be included in the analysis. Adjust the pounds accordingly and plot both on the graph.
7. Write files out in csv to be loaded into RShiny, Python, Javascript, or Tableau….

# Data Visualization: Tableau

1. Students have access to Tableau for free. Everyone has access to Tableau Public.
2. Download and open Tableau 🡪 they offer training videos that are helpful for getting started. Googling specific questions is also helpful.
3. Start by uploading the three csv files you saved out. Notice that Tableau automatically makes the connection between the sites.
4. Create a new worksheet and name it: Map
   1. Create a map with latitude and longitude for the nitrogen data. You can highlight whatever features you desire. I chose to:
      1. Color the sites using the average Total N value.
      2. Make the site size relative to the number of samples.
      3. Included the attribute name in the tool tip description.
      4. Add a (filter) dropdown list that lets the user look at Nitrogen symbols over time.
5. Create a new worksheet and name it: Temporal
   1. Here we can explore temporal trends in nitrogen
   2. Create a plot showing the average nitrogen (mg/l) by sample date
6. Create a new worksheet and name it boxplot month
   1. Create a plot showing the distribution of nitrogen samples by month
7. Repeat the boxplot worksheet with year
8. Create a dashboard showing the map, time series, and boxplots.
   1. Make the map the filter for the other plots
   2. Click on the “use as filter” and then go to Analysis and click “Aggregate Measures”
9. Create a new worksheet for Haw River and plot the pounds by year. Include a reference line showing the threshold limit.
10. Repeat for Hope Creek using the adjusted limits. Add the actual calculated pounds to the tooltip.
11. Create a new map with just those two points
12. Create a dashboard showing the map and both line charts.
13. Create a storyboard
    1. Add an introduction to the story.
    2. Add the worksheets or dashboards in as desired to create a flow to what you want to communicate.

Sample: <https://public.tableau.com/profile/lauren3839#!/vizhome/Unit3-Nitrogen/Story1?publish=yes>

Other examples of water quality using Shiny R:

Mark Ziman’s tool: <https://mark-ziman.shinyapps.io/WQRAT_MZ/>

CA Stormwater Tool: <https://daltare.shinyapps.io/Stormwater_Enforcement_Tool/>