ZOO 315 Limnology: Conservation of Aquatic Resources
Analysis of long-term data assignment (worth 16% of final grade)

Part 1: to be submitted on Canvas on Oct 18th (6pts)

Part 2: to be submitted on Canvas Dec 8th (10pts)

Real world limnology:

The objectives of this exercise are to analyze and interpret limnological data from a Wisconsin lake and provide you with an opportunity to improve your data handling and data interpretation skills. The North Temperate Lakes Long Term Ecological Research (NTL-LTER) program began monitoring several physical, chemical, and biological attributes of 7 lakes in Wisconsin in the 1980s. Your task is to acquire the data set for the lake you've been assigned, do basic data manipulations and analyses, create two graphs of these data, and answer questions about your variable. Acquiring the data from the NTL-LTER is relatively straightforward, but there are lots of numbers, so please follow the directions carefully so that you get the appropriate data set. You are to work alone on this assignment.

Learning Outcomes

- Access and organize open-source data
- Access and read metadata
- Manipulate data to a format that enables analysis
- Plot visualizations of data that are scientifically correct and visually appealing
- Analyze data for long-term change in lakes
- Interpret and deduce factors that account for long-term change in lakes

Words of Advice:

- **Download your data soon to make sure you can get it without a problem.**
- Please read this handout carefully.
- If you are new to data analysis, this is your chance to get better. Start early so that you can get help.
- We have provided instruction for using Google Sheets here. If you want to use a more sophisticated coding language (e.g., R, Python), that is fine too as long as it is open source. Save the code. We may ask you to turn it in.

PART 1 - DATA ACQUISITION

1. Find the lake and variable you've been assigned from the "*Your lake and variable*" link on Canvas. Be sure you've got the right lake and variable.

2. Go to the North Temperate Lakes Long Term Ecological Research web page: http://lter.limnology.wisc.edu/

3. Click on the `Data` tab.

    a. Enter the search keyword for your variable in the **Full text search** space
    b. Select the appropriate **LTER Core Areas** from the pulldown menu (except if your variable is Secchi- leave it at -Any-),
    c. Select *NTL Core Datasets* from the **NTL Themes** menu.
    d. Leave the **Dataset Author** menu at -Any-
    e. Hit the Apply button

4. This search will return a list of one or more data set titles. Select the appropriate data set for your variable listed in Table 1.

Table 1. Name of the dataset that contains the data for your assigned variable.

Variable	Dataset
Secchi disk depth	North Temperate Lakes LTER: Secchi Disk Depth; Other Auxiliary Base Crew Sample Data 1981 – current
Specific conductance	North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Major Ions 1981- current
NO3+NO2_N	North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon 1981-current
NH4-N	North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon 1981-current
Total filtered N	North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon 1981-current
Total filtered P	North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon 1981-current
Chlorophyll a	North Temperate Lakes LTER: Chlorophyll- Trout Lake Area 1981- present

5. Selecting the appropriate dataset Title in Step 4 takes you to a page that provides a brief description of the dataset (the Abstract), several links to related datasets, lakes, etc., and a link that says View Full Metadata (Fig. 1)
    a. FYI: This View Full Metadata link will open a new tab for another, similar web page that is the official version of the dataset published by an organization called the Environmental Data Initiative. This is a data repository where scientists freely share their research data. It includes details on when, where, how, and who collected the data. Keep this in mind, as it may be a useful resource for you as you work on part 2 of the assignment.
    b. The Full Metadata page documents the appropriate units for your variable.

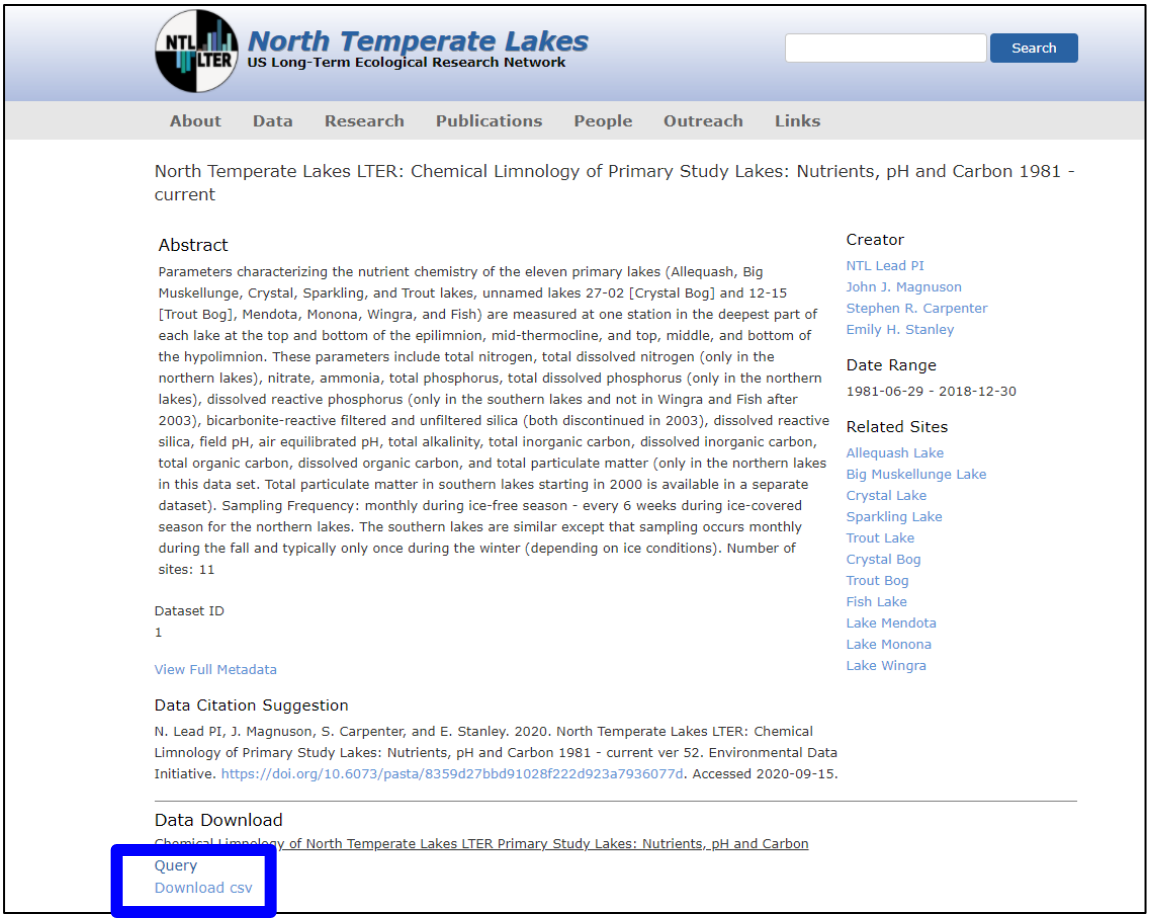
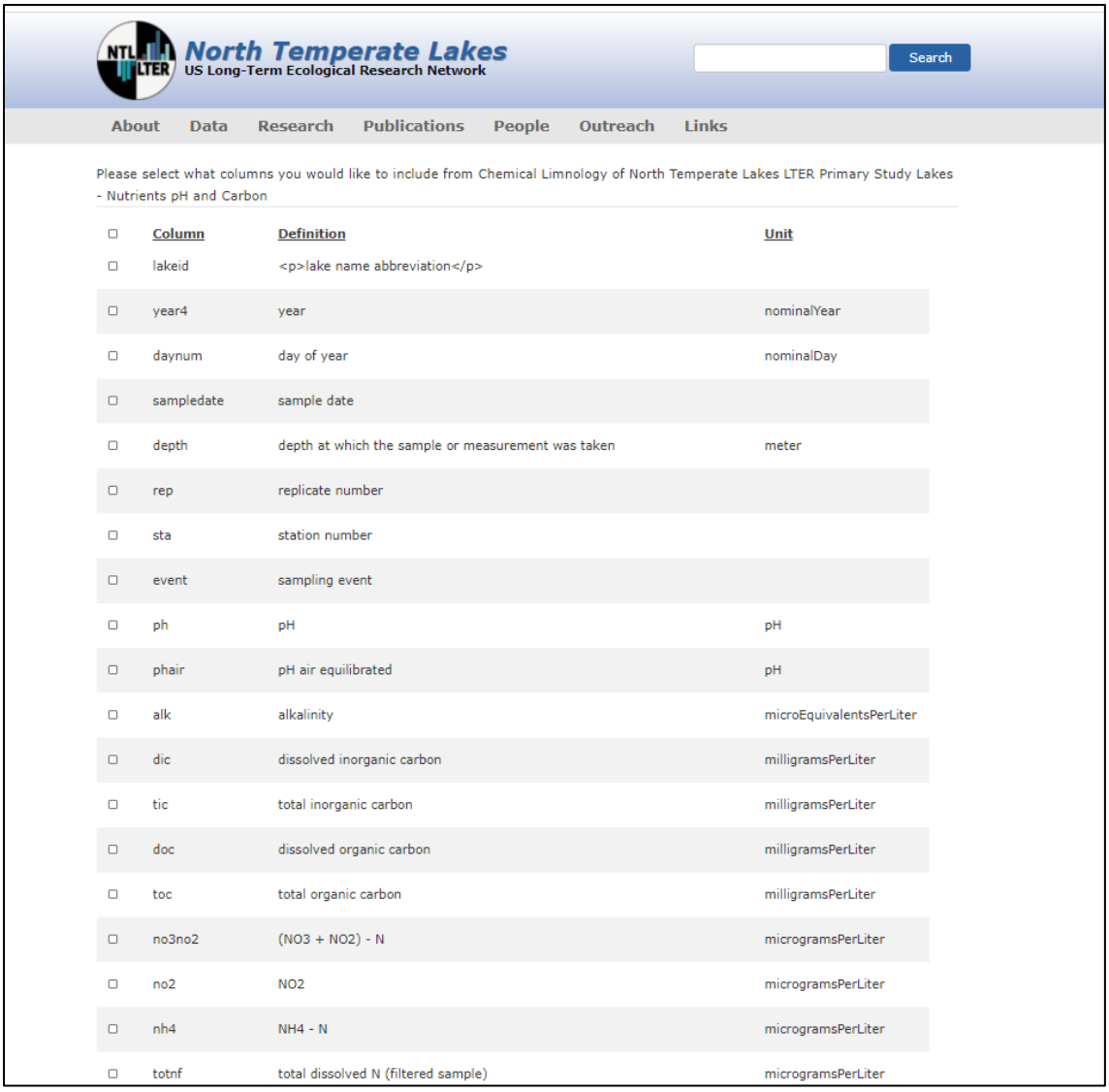


Figure 1. Example of the page where you will land after step 4. The blue box highlights the location of the link needed to get the data you need for this assignment.

As highlighted in the blue box in Fig, 1, select the Query link below the **Data Download** header to get to the 'Query menu' (Fig. 2)

6. Select the following items from your Query menu:
   - lakeid
   - sampledate
   - depth*
   - your assigned variable listed in the "Column name" column in "Your lake and variable"

If your variable is Secchi disk depth, depth will **not show up on this menu, so don't panic when you don't see this option!*



Column	Definition	Unit
lakeid	lake name abbreviation	
year4	year	nominalYear
daynum	day of year	nominalDay
sampledate	sample date	
depth	depth at which the sample or measurement was taken	meter
rep	replicate number	
sta	station number	
event	sampling event	
ph	pH	pH
phair	pH air equilibrated	pH
alk	alkalinity	microEquivalentsPerLiter
dic	dissolved inorganic carbon	milligramsPerLiter
tic	total inorganic carbon	milligramsPerLiter
doc	dissolved organic carbon	milligramsPerLiter
toc	total organic carbon	milligramsPerLiter
no3no2	(NO3 + NO2) - N	microgramsPerLiter
no2	NO2	microgramsPerLiter
nh4	NH4 - N	microgramsPerLiter
totnf	total dissolved N (filtered sample)	microgramsPerLiter

Fig. 2. Page for data acquisition, corresponding to step #6 for any nutrient-related variable. Note the UNITS!

7. Scroll down to the bottom of the page to Please choose any filters you would like to add. (Fig. 3).
   →Select your lake from the **Limit by lakeid:** box.
   →Hit Download

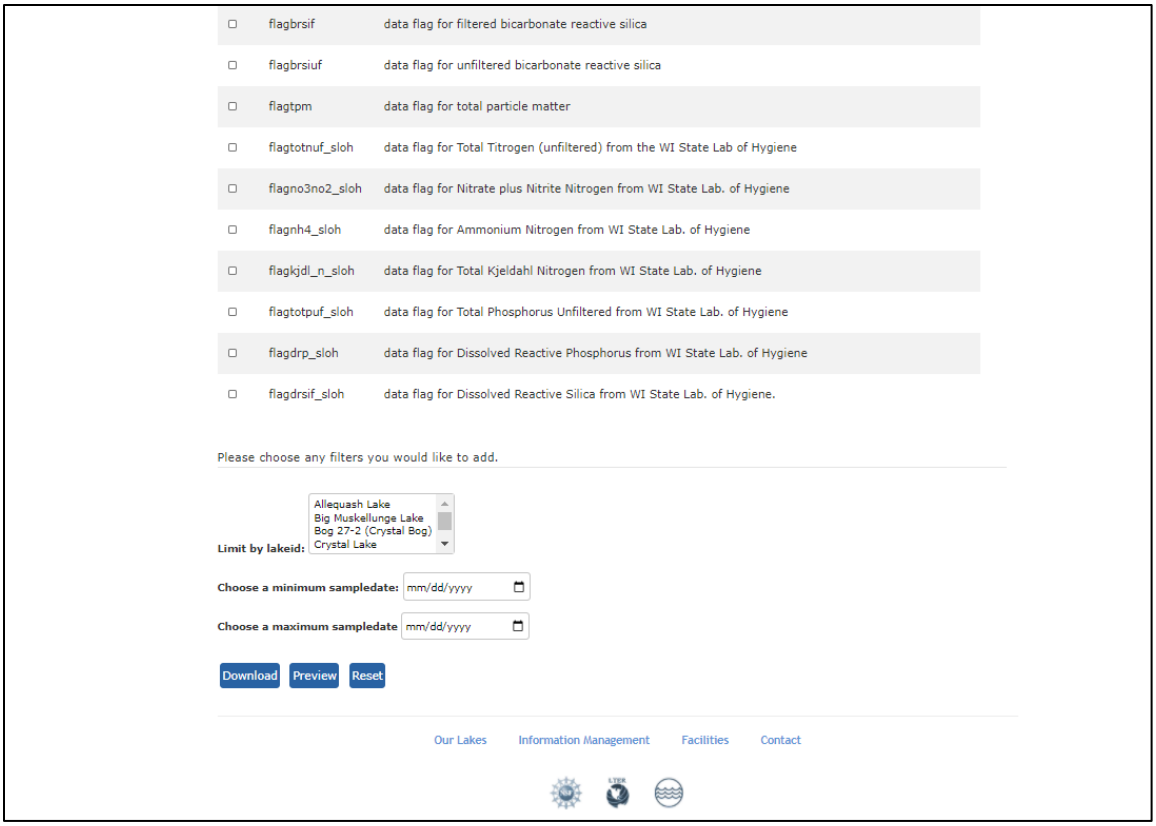


Fig. 3. Filters for data acquisition, corresponding to step #7. The lake being selected is Crystal Bog (CB), and data for the selected variable will be limited to samples collected at the lake surface (0 m). Note that the 'depth' filter may NOT show up for your variable, and you will have to eliminate data from depths you don't need for your report following directions given after you have your data into a google sheet.

8. A .csv file will download on your computer. Do not open this file in Excel! (Why? Because Excel will change the timestamp format).

9. Open googlesheets. You can do this by typing http://sheet.new/ into your browser.

    a. **Make sure you are signed into to google using your UW account**.
    b. Click the *File* dropdown, and click *Import*. Click the option *Upload* from the header menu. Load the .csv you have downloaded.
    c. When the Import file menu appears. Choose the following options.
        i. *Import location = Create new spreadsheet*
        ii. *Separator type = Comma*
        iii. *Convert text to numbers, dates, and formulas = Yes*
        iv. Click *Import data*
    d. Once you have clicked *Import data*, near the top of the Import File box, a message will read appear

File imported successfully. **Open now »**

    e. Click *Open now*

10. Your spreadsheet will open. Rename your spreadsheet in the top left corner. Check to make sure there is data. If there is a blank top row, you can delete it (highlight row, right click, and select delete row).

PART 1- DATA PEPARATION AND GRAPHING

Part 1: to be submitted on Canvas on Oct 18th (6pts)

Now that you have your data, your tasks will be to create graphs of your variables over time.

Before you begin, look over your data sets. First, check to be sure that the lakeid matches your lake and that you have data from your lake only and no other lakes. Next, you will need to do some preliminary work to get your data into shape to make graphing easier.

- Delete rows that have date and depth information but no data for your variable.
- This can be done quickly using the `Create a Filter` function on the Data menu.
- For each column, click the green dropdown filter symbol
- For your variable, you can uncheck any blank values by clicking on (BLANKS) to remove these empty rows.
- For depth, you can uncheck all depth values other than 0 m (if there was no 'depth' filter option for your variable when downloading your data).
- Create a column for month and enter the appropriate number into this column for all dates (e.g., if a sample was collected in October, enter 10 into this new column).
    - As you do this, you may notice that there are often duplicated dates, although the value for your variable may be different. Duplicate samples are often collected for quality assurance purposes and can be left as is for this assignment.
    - For advanced users, you can do this quickly by entering the formula `= month( )` into the first cell of the new column and selecting your *sampledate* value for that row. This formula can then be copied to the entire column.
    - * Filter out data from December. Often there are too few sampling points in December to give a meaningful average. Even limnologists take a break for the holidays!
- Create a new sheet by clicking the + button on the bottom left and selecting `add sheet`
- Rename Sheet2 by double clicking on the tab at the bottom. Name the sheet 'filtered data'
- Select the data in the original sheet, and copy and paste it into the filtered data worksheet.
    - This new sheet will only have data at 0 m depth. Work with this new worksheet for the remainder of this exercise.

Graph #1: Create a graph (googlesheets calls them charts) showing your variable over the entire period of record. There are many graph styles to choose from; try different styles to see which shows the data most clearly. Often, the best style is a simple line graph.

- X and Y axes should be labeled. Date goes on the X axis, and the name of your variable ***and its units*** should be on the Y axis.
- Since you are only graphing one data set, do ***not*** include a legend or a graph title. They take up space and make it harder for you to see patterns in your data. Delete them both!
- You may want to adjust the range for your X and Y axes. Stretching out your data (e.g., if your variable varies between 1 and 10, then a good range for your Y axis would be 0 to 11) make trends or patterns more obvious. Also be sure that all data are included (for the prior example, do not have your Y-axis span from 2 to 8; if you do this, then some of the data will be cut off).
- Axis ranges can be adjusted by clicking on your axes which opens the Vertical Axis options. You can specify Min and Max values.

- Copy your final graph into a text document.
- Instead of a graph legend and a graph title, write a figure caption beneath the graph. This text should be short but should provide enough detail so that someone looking at only the graph and its caption can understand the information being presented. Look at a scientific publication for some examples. Think of captions like alt-text for your graph – if the graph wasn't there, could someone figure out what it is showing?

Outliers:

Does your data have outliers? These are data points far away from the mean/median. Sometimes "bad" data ends up in published data. This can be a result of human error, bad protocols, or a variety of other factors.

- If you feel like a data point should be omitted because it's not realistic, report this and explain why the point(s) was (were) removed in the figure caption.
- Outliers will distort the statistical calculations in Graph 2.
- You can include two graphs — one with all the data, and one with the outlier(s) removed.

Graph #2: Create a graph showing the average monthly concentration at 0 m over the entire period of record. The purpose of this graph is to determine the annual pattern of your variable.

- The quickest way to determine the average value for each month is to use a pivot table. **If you're unfamiliar with what a pivot table is, and why it's used, it's worth finding out more via a google search!**
- The *Pivot Table* function is found under the *Data* dropdown menu.
- In the Create Pivot Table box, select *insert to: New Sheet*. Click create.
- At the moment the Pivot Table is empty. You need to add Rows and Values
- Click *Add* beside Rows, and select your month column. Uncheck *Show totals*
- Click *Add* beside Values, and select your variable column. *Select summarize by AVERAGE*
- Note that you may not have data for all months. December is often missing (why?), and for Cl, SO4, and cond, samples are collected seasonally. If you want to average by season, that would be fine (summer = Jun-Sep; fall = Oct-Dec; winter = Jan-Mar; spring = Apr-May).
- Make a chart of month vs average monthly variable. As with graph 1, label X and Y axis appropriately and do not include a legend or axis title, but do include a figure caption as described above.

Checklist:

- ☐ Did you copy your graphs into a text document and write a figure caption for each graph?
- ☐ Inspect your graphs - do they make sense?
- ☐ Do your graphs look like scientific graphs you have seen in scientific publications and textbooks? See the pdf *10 rules for better figures* on Canvas. There is more information than you need, but some of the 'rules' are very helpful (e.g., Rules 4 and 8).
- ☐ Do you have labels and units on your axes?
- ☐ Did you note any outliers removed?

A note on Figure captions:

A lot of people's intuition is to make Figure captions too complicated. We recommend you look at published scientific papers to get a feel for how figure captions are written.

- Figure captions are placed below the plot.
- The caption should start with "Figure 1: ...."
- The figure caption should only state exactly what is being plotted.
- You do not need to say "This figure shows", as that is the point of having a figure caption.
- The figure caption should include the lake name, the date range, and the depth.
- You do not need to describe the patterns or provide explanations - this will be saved for Part B and written in the text.

In general, the following figure caption will suffice.

Examples:

Figure 1: Variable at 0 m depth in XXX Lake from 1982-2018.
Figure 1: Concentrations of total phosphorus (µg/L) at 0 m depth in Lake Mendota, Wisconsin from 1982-2018.
You can also note any outliers or other data removed in additional text.

Submission:

Please submit on Canvas a pdf document showing both of your graphs.

- ☐ Provide a title that lists your lake name and variable.
- ☐ List the website of the dataset you used.
- ☐ If additional steps were taken in data preparation, please note these in your document.
- ☐ Include the URL to your google sheet. If you coded this assignment in R or Python, include the code at the end of your assignment.

Grading Rubric (6pts):

For each graph:

- ☐ Is the data logical and presented in a visually clean style? (1pts)
- ☐ Is there a figure caption that effectively details the information provided in the graph? (1pt)
- ☐ Are the axes of the graph logical and formatted correctly? (1pt)

Overall:

- ☐ Up to 3 pts subtracted if missing:
  - Title
  - Website of dataset
  - URL to your google sheet or reproducible code

PART 2 - DATA ANALYSIS AND REPORT

Part 2: to be submitted on Canvas on Dec 8th (10pts)

You will have received comments on your graphs from Part 1 on Canvas. Please incorporate any edits required into your final report.

Final Report:

The first page of your report will be your submission from part 1. Make sure your final report starts with your name, your lake, and your variable.

- If you received less than perfect on part 1, please adjust your graphs/captions accordingly, otherwise points may be subtracted.
- If you received 6/6, use your prior submission as is.

Following your graphs, please provide your answers in single spaced text. If you get information from a source other than lecture, please be sure to insert a citation in your text and provide the full citation to your source at the end. *Otherwise, it is considered plagiarism and points will be deducted.* All parts should be submitted as a single pdf to Canvas.

Questions:

1. Provide a description of your lake and the surrounding landscape in which the lake resides. This information is to help provide the context for the next question. Be sure to cite/report any sources that you use. [2 pts]

2. For **graph #2 (monthly averages).** Describe how your variable is changing over the year. [2 pts]
    a. What is the annual pattern for your variable?
    b. When is the annual minimum and maximum?
    c. Are these values high or low relative to other lakes in Wisconsin?

3. Describe the factors that influence your variable and are likely to be creating some of the annual patterns you've described. Think about phenomena in the area around the lake (external) as well as processes that may be occurring within the lake itself (internal). Be sure your explanation makes sense for your specific lake. **Again, be sure to cite/report any sources that you use for this question.** [6 pts]
    a. What is causing this annual pattern?
    b. Does the annual pattern result from external or internal drivers (or both)?
    c. How does the trophic status of the lake affect the annual dynamics? Would you expect something different if the lake was, for example, eutrophic?

Checklist:

- ☐ Part 2 is worth 10 pts and 10% of your grade. Make sure your answers to the questions are thorough and well-supported with citations.
    - Write at least one sentence per point
- ☐ **Units are in metric units**, as is the standard practice in science, and for all but three countries in the world.
- ☐ Single spaced text.

Submission:

- Turn in your report as a pdf (your graphs, answers to questions) on Canvas by the due date.

- Include the URL to your google sheet. If you coded this assignment in R or Python, include the code at the end of your assignment. Two points will be subtracted from the final grade of the assignment if this information is not provided.
- 1 point per day will be deducted for late assignments.

Grading Rubric:

Questions 1-3 are worth 10 points in total (see questions for point values).

Important information and some advice for this report:

- Work alone. This should be an independent effort. Do not borrow a data set from someone else and be sure that you're graphing the correct parameter from the correct lake.

- There is a lot of information about these lakes and about your variable on the NTL-LTER web page including how measurements are made. Information about units can be found on the Query menu (see Fig. 2) You can also get some information about your lake by clicking on the lake's name in the "Related sites" list (see Fig. 1).

- Note that there are many Crystal Lakes in the world; all lakes considered in this assignment are either in Vilas County north of Woodruff/south of Boulder Junction.

- If you have Trout Bog or Crystal Bog as your lake: even though these are called "Bogs," they are not bog wetlands. They are dystrophic lakes.

- Feel free to examine any other variable you want from the NTL-LTER database. This may help you interpret your results. It is also a good idea to consult your class notes or limnology texts or scientific articles to help you interpret your data. And don't forget to cite these sources if you use them. See posted articles about proper scientific citation practices.

- There are some supplemental files posted with this assignment that provide useful information on how and when to include citations in your writing.