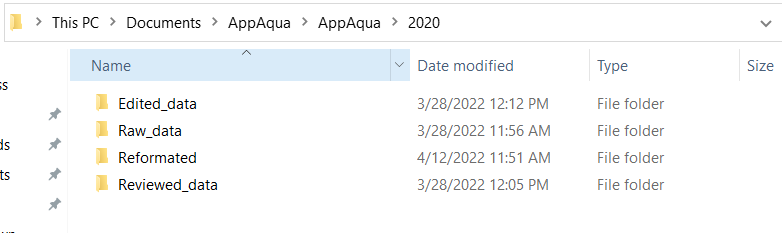
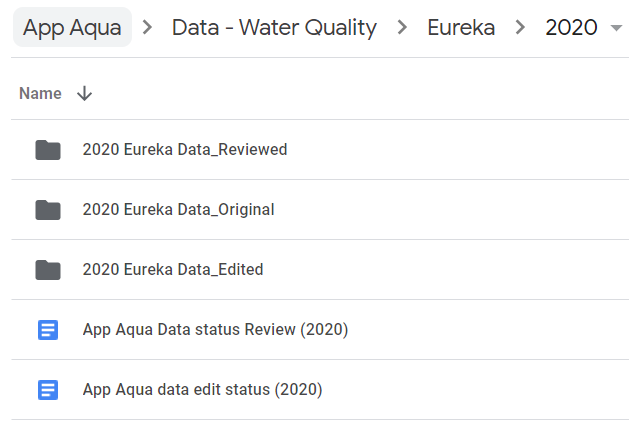
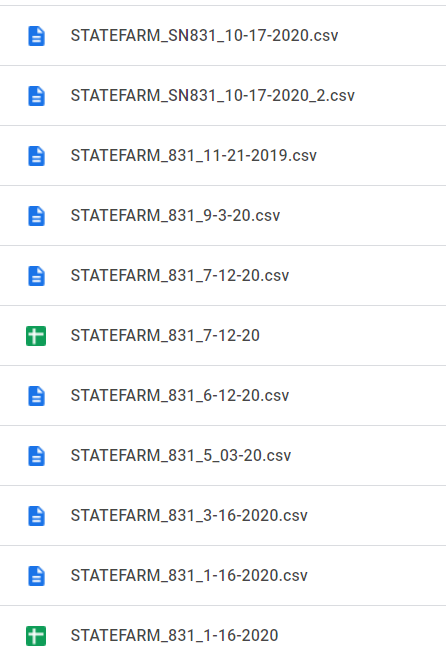
**Before Downloading**

1. Open the folder on your computer/ workspace and create a folder in your Documents named AppAqua if you do not already have one.
2. Create a folder inside of the AppAqua folder named after the year that you will be working with.
3. Create 4 folders named Raw\_data, Reviewed\_data, Reformatted, and Edited\_data. Be mindful of including underscores (“\_”) when you want to separate words, coding programs will think you are naming two separate items if you include a space. Additionally, having a space in a folder or file name is one of the most common reasons documents become corrupted.



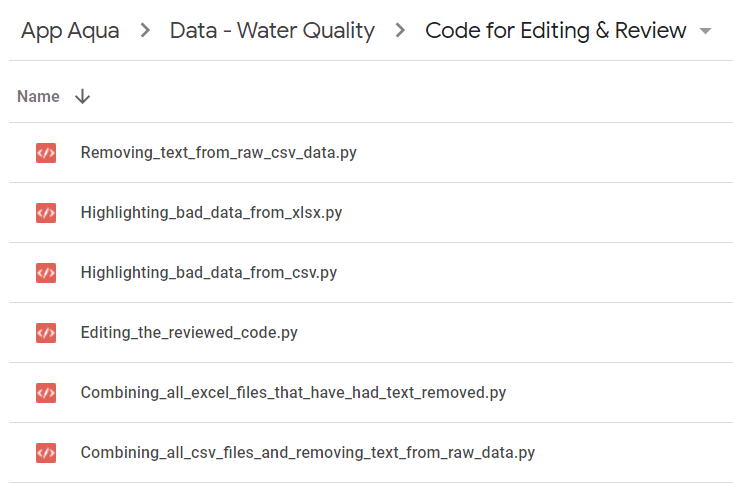
**Downloading AppAqua Data**

1. Go to AppAqua Google Drive.
2. Chose the Data - Water Quality folder
3. Select Eureka
4. Choose the folder with the year that you need to work on
5. Chose the Eureka Data Original folder
6. Chose the creek that you will be working with
7. Download all of the .csv files that you need, be mindful not to download the same time period with different names, this occurs sometimes and will cause issues that you will not see until much later in the process.



Notice how the top two files are the same.

1. Move your downloaded files into the Raw\_data folder that you created earlier.
2. Download all the Python code in the Code for Editing and Reviewing folder inside the main Data - Water Quality folder

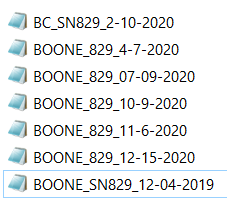
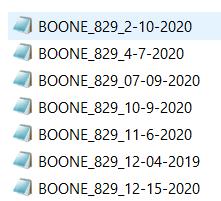


**Working with Python**

1. If you do not already have a Python IDE installed you can download one to your personal computer. I recommend using Spyder from Anaconda’s distribution. I find this to have the best method for installing packages. Another option is if you can use udesk and have no desire to use Python beyond this to use PythonWin, this should also be available on the school computers.
2. You will find that there are a few packages that you will have to install to run all the scripts. In the command line, which has a “>” where you type, depending on what version of Python you are using you will say pip install “your package” or conda install “your package”.
3. I am not going to provide a detailed installation and user guide because there will be more questions than I can think of answers to so my recommendation is to find a YouTube video or Google your specific question. It can be frustrating at times but you can do it.

**Working With Downloaded Data**

1. Arrange the .csv files in order to have the files from the beginning of the year at the top and the ones from the end of the year at the bottom. You might need to rename a few files because the titles will change from time to time. Note that in the example below, there is a file from 2019 with the 2020 data, this happens often and is ok because it will java data from January in most cases.

 =>

1. Open each file and check the first and last row of data, making sure that you do not have two files with overlapping times. Some files will be named incorrectly, for example, it might be named 10\_28\_19 but it has starting values at 09\_09\_19 which is the same starting value as another file. **Using both of these files will cause errors** so one of them needs to be removed before moving forward. To save time later in the process I recommend writing down the date and time of the first and last entry for each file.
2. If you find 2 files with overlapping times, make sure to **remove all repeated times** in one of the files before moving on.
3. Open the download python script, Combining\_all\_csv\_files\_and\_removing\_text\_from\_raw\_data.py in your python. While using the script if you are unsure of what the script is doing there are notes about most of the code within the file to help you.
4. Go to the folder with your raw data, right-click the path, and copy that filename



In the python script paste that pathname within r’ ‘



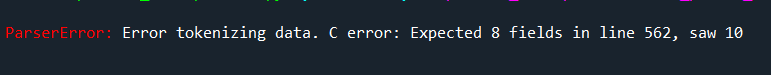
Paste this path within the ( ) of os.chdir() and add an additional \ to every backslash



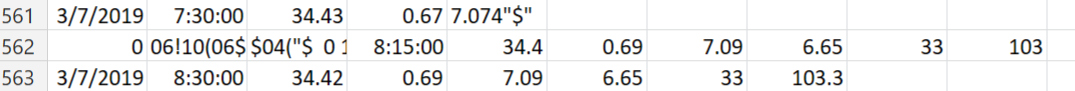
This will set the working directory to this folder and output the combined .csv into this folder.

1. Occasionally there will be errors when you run the code due to there being something wrong in the .csv, sometimes it will be clear which file is messed up, other times it won't be.

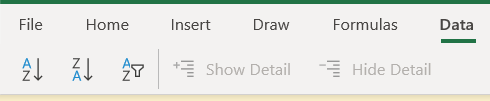
Example issue:

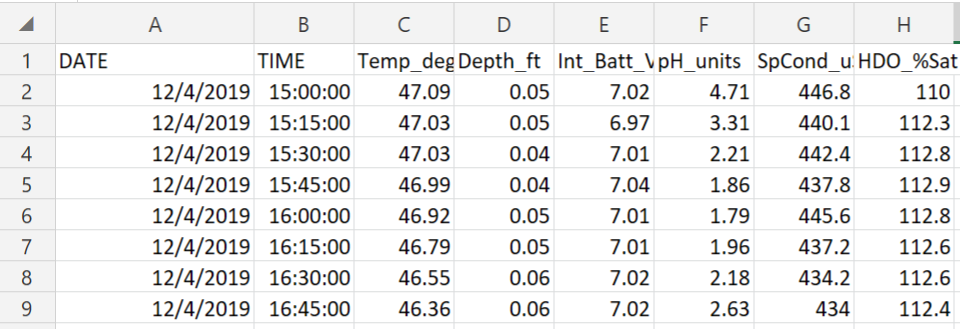


This was the issue:

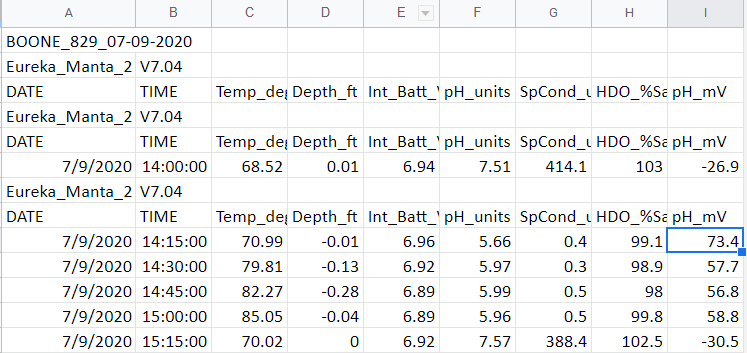


For an issue like this, just delete the two rows that are messed up and make sure that after deleting the information from the rows that the rows are removed as well.

1. After running the code open the combined .csv, located in your raw data “creek” folder, to make sure everything went correctly, if it did it should look like this, you will also want to select the DATE column and sort alphabetically A -> Z.
2. 



If it does not, figure out what went wrong by looking at the original files.



This is an example of a file that caused an error, pH\_mV is never used and adding the extra column through an error in the code.

When you encounter an issue where a column is added that we never use simple delete the whole column and make sure all the other columns are in the correct order. Sometimes you will have data in C instead of F, you will need to make a new column to covert these values and then save as a .csv so that the formula is “broken” but the values remain the same. At this point you will rearrange the columns to the correct order and remove the column in C.

1. For the combined file output add the name of the creek to the start of the file name.



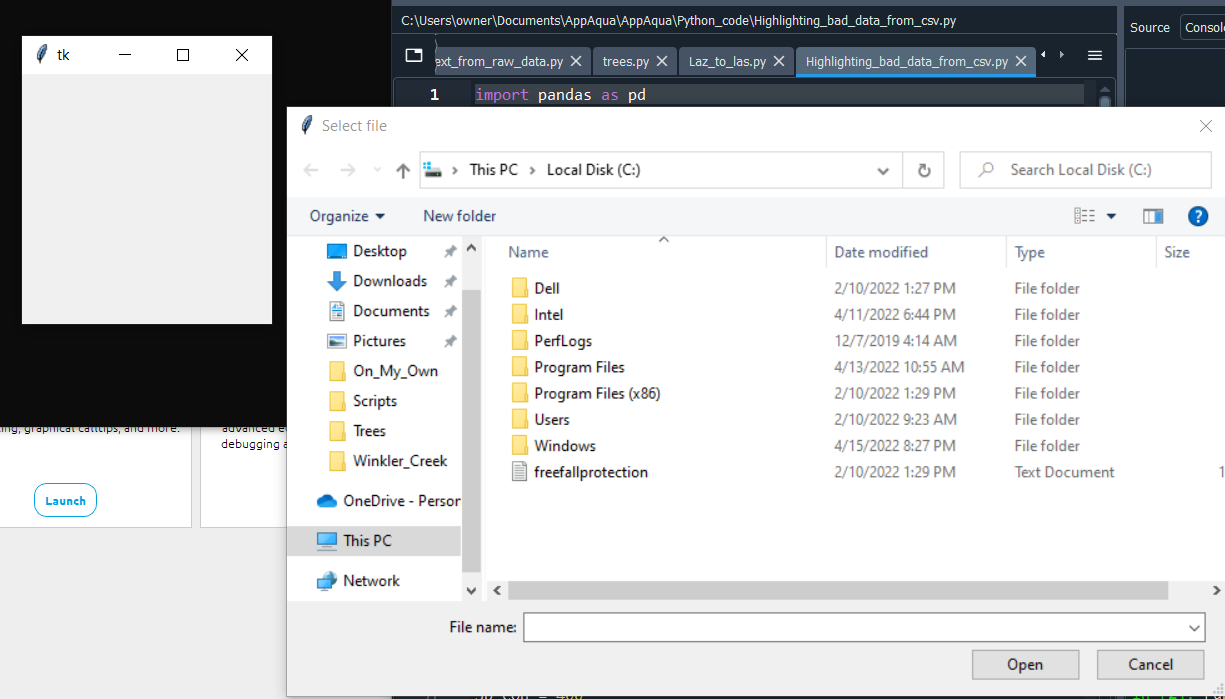
**Highlighting Bad Data**

1. Once this is completed open the Highlighting\_bad\_data\_from\_csv.py
2. Just like you did in the previous portion you need to copy the pathname of the folder of your Reviewed\_data folder and paste that to the rec.to\_excel(), at the bottom of the script, make sure to keep the last section of that line the same, \\reviewed\_””\_Creek.xlsx.



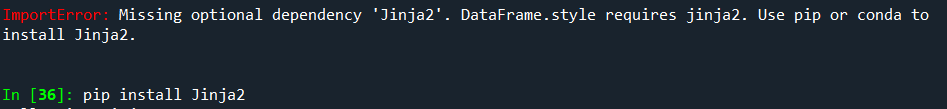


1. You will select the combined.csv file that you are working with when you run the script, make sure to close the tk box that pops up after selecting your file, or else the code will not run.



These windows will pop up when you run the script, sometimes behind your Python environment. The Select file window will allow you to navigate to your file, select it, and this window will close automatically.

1. If this is the first time using python you might get an error message saying you need to install Jinja2, in the command line type pip install Jinja2 and press enter, this should solve that issue.

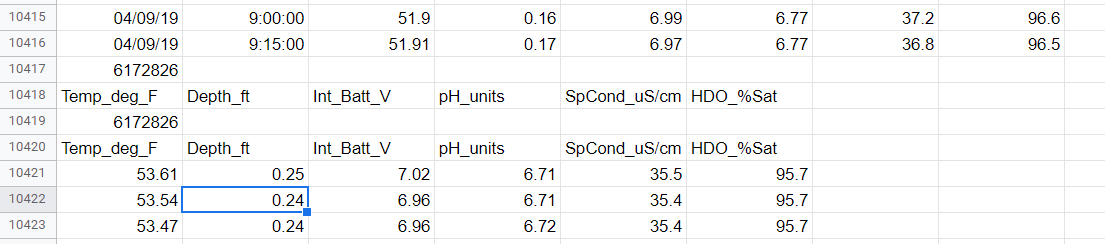


1. Just like combining the .csv files, you will encounter errors here, these tend to be easier to address because its only in one file.

Example issue:



This is the issue:



This issue is actually something that will require you to go back to the original file and fix it there, then combine the .csv again and then rerun the review steps.

Sometimes even though you removed the lines that caused the issue its still not fixed, I find the best way to fix this is to go into that file, sort it so that the text and large numbers are at the top and remove them from the file.

Example issue:



This issue can also be caused by the same date and time occurring on multiple .csv files that were then combined.

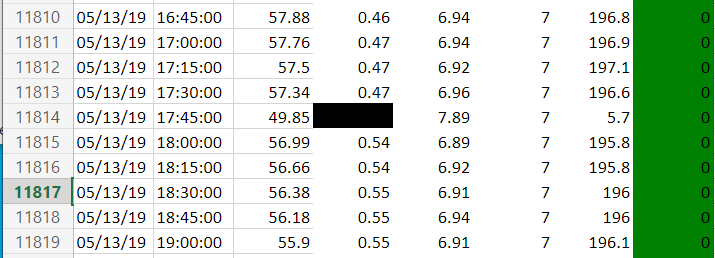
If all else fails you might have to go into the combined file, find what dates are messed up, and copy over those entries from the raw data after you remove lines of text.

1. The code should run and there will be a file created in the Reviewed\_data folder with the file name you chose in the rec.to\_excel command,



When you open it, cells will be highlighted based on the conditions set within the script, these are not “normal” values in the data but that does not mean that they need to be deleted.

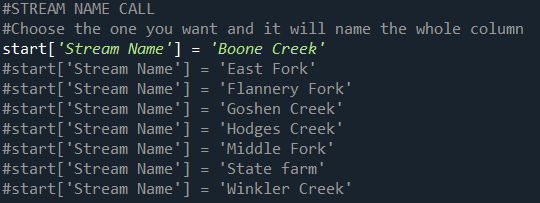
**Manually Reviewing The Data**

1. Open the reviewed .xlsx file in excel.
2. Open your Raw\_data folder for the creek you are reviewing
3. Going down the line in order, open each individual .csv folder to find what time each calibration took place.
4. You want to review the data between each calibration to make sure that the sensors were picking up real information.
5. You are primarily looking for two things, that the data begin to “drift”, increasing or decreasing in value until it reaches the next calibration at which point it will be drastically shifted in value. If so you need to go back up to locate where this drifting begins and remove the cells that are erroneous. The other issue you will see is sometimes the pH sensor will record impossible values, a pH of -10 or 30 can not occur and will need to be removed. The sensors will occasionally have a one-time pocket of very high or low pH or specific conductivity and that is ok as long as the values go back to normal. Sometimes there will be substances moving through the creek that will cause this. But do look at what is going on, does the data make sense? Most times it will go up, plateau, and then move down.
6. When you get to the row where the calibration took place you will be doing two things. Look at the range of values before and after calibration, this is the best time to identify drift. Are the values drastically different on either side of the calibration? Some difference is fine. Also 99% of the time you will need to remove the first point of data from the calibration because the entry values will have taken place outside of the creek, this will be obvious when you see the data. 

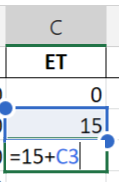
Line 11814 was recorded directly after calibration and the specific conductivity column is drastically different from the ones above and below so we know we need to remove the entire row. This is the only “normal” reason to remove an entire row of data.

**Reformat Reviewed Data**

1. Open editing\_the\_reviewed\_code.py
2. Change the rec.to\_excel() at the bottom of the script to the path of the folder that you are using for Edited\_data.
3. In the Stream Name Call in the script make sure to remove the # from the stream you are working with and make sure all others have a # infront of them. Then run the code. Click the file name under the reviewed data for the stream that you are doing.



1. Open the newly created file, in the newly created ET column you will have the first row will be 0, the second being 15, and the third being =15+C3

This will allow us to notate how much time has passed between calibrations, you will need to drag this to each calibration time and restart the process, in place of C3 it will be the cell that has a value of 15.

1. As you are doing this make sure that the columns look right and nothing seems wrong
2. When you have gone through the whole file save it as Reformated and save it in the Reformated\_data folder and you are ready to move on to the next creek.