1. File from MM emailed on 2024-02-26 “Copy of Guana\_masterdata\_2024-02-26\_revised by\_MM.xlsx” was corrupted and therefore not able to be used. Was able to copy all the data into a new spreadsheet to complete the following work.
   1. Separated all years into their own “sheet”.
   2. Added data dictionary as a sheet at the beginning of the workbook.
      1. Added GTMOLNUT\_dup information
      2. Added lat and long for GTMDNNUT as it was missing
      3. Added lat and long for GTMDSNUT as it was missing
   3. Added a “Notes” sheet and “QAQC” sheet to describe QAQC and procedures for working with this datafile.
2. Column adjustments
   1. Removed “AdjustedDate” and “PrepDate” columns
      1. What is `PrepDate` and is it important?
         1. Resolution: Removed “PrepDate” column. Looks like it is something that has been used with AEL lab. Can always be added back into the data but seems to be entered in a variety of formats when it was entered. Again, removing from all years and data sheets.
   2. Renamed “SampleType” column back to “ActivityType” as this column is named directly for WIN purposes.
   3. Added `F\_Record` column to all sheets to allow for QAQC flags and codes that apply to the entire sample to be included in the datafile
      1. 2017 has examples on how this can be performed, specifically how the codes should look.
3. UNID column stopped being used in 2019 as did entering the `Lat` and `Long` columns of information.
   1. Resolution: Removed the `Lat` and `Long` columns from all sheets since this information is in the `data\_dictionary` tab and can be added when needed for analysis later on.
4. *What do to about Secchi readings greater than 1.2 m? In data listed as `>1.2` which is not able to be analyzed*.
   1. Resolution: Created new error code [SAL] which means sample above limit. Can flag with 0, 1, or -3 depending on results.
5. QAQC of 2017
   1. ‘GTMOLNUT\_dup’ data:
      1. Field notes on 2017-10-18 indicate that the Overlook Duplicate samples (first collected this month) were taken by Jimmy and Kaitlyn in the middle of the lake to test for the representativeness of the OL site. The lat and long of this sample was 30.08302, -81.34286 which is the location of GTMLMNUT which began sampling “officially” in December 2017.
      2. Added GTMOLNUT\_dup data back into the data file and rewrote data dictionary to reflect that these data are most similar to Lake Middle than the original GTMOLNUT station code.
      3. Added {CSM} code in `F\_Record` column for GTMOLNUT\_dup data in 2017-10-18 to note why it was collected. (notes in previous bullet point)
      4. Added {CSM} code in `F\_Record` column for GTMOLNUT\_dup data in 2017-11-02 to note why it was collected. (notes in previous bullet point)
   2. Laboratory information “Analysis Method, Lab ID, Lab Accreditation, Laboratory Name” information is incorrect.
      1. Resolution: Revisited laboratory documents to complete missing information and check data entry. Date Received and Date Analyzed information also missing. PQL, Dil also added.
   3. Flagged bacteria data in Aug that had a B remark as <1> (CSM) because Results based on colony counts outside acceptable range
   4. Incorrect result entered for CHLa\_UnC in September GTMLSNUT 09/20/2017 09:47. Corrected for 34.3
   5. Added {CWE} {CSM} to `F\_Record`column for all sites on 2017-10-18 due to Field notes indicating “Nor’easter and water flowing north to lake”.
   6. Added {CRE} {CSM} to `F\_Record` column for all sites on 2017-12-13 due to Field notes indicating “that samples had to be taken at a lower tide due to duck hunting in Guana Lake that lasted until 12:00”.
6. `SampleFraction` column is only used in 2022-2023 data. If using, need to retroactively complete this information in all previous years to be consistent.
7. Need Lat and Long of the Overlook station for the few months it was collected. The previous Lat and Long was the LM site, not the station that was accessed via the trail.

Started building code to read in all data for analysis. While doing so, QAQC addressed in Guana\_MASTER.xlsx are noted below:

* Station Code
  + Found station code spelled “GTMGMKNUT” in data for 2021-02-24.
    - Resolution: confirmed it was an entry error for field data. Corrected station code to “GTMMKNUT”
* Parameter names
  + CHLa\_C entered as CHLA\_C in 2021-12-01 for all sites.
    - Resolution: corrected to CHLa\_C which is used in all other entries
  + CHLa\_UnC entered as CHLa\_Unc in 2021-11-03 for GTMGL2NUT and 2022-05-12 for GTMGL1NUT.
    - Resolution: corrected to CHLa\_UnC for these two entries.
  + TOC written as ComponentShort instead of W-TOC
    - Resolution: corrected to W-TOC
      * All sites 2021-11-03
      * All sites 2021-12-01
      * All sites 2022-02-14
      * All sites 2022-05-12
      * All sites 2022-07-11
      * All sites 2022-09-07
      * All sites 2022-11-07
  + PHEA entered as PHEa
    - Resolution: corrected to PHEA
      * 2022-02-14 GTMGL1NUT
      * 2022-09-07 GTMGR3NUT
      * All sites 2022-10-06
      * All sites 2022-11-07
      * All sites 2022-12-05
      * All sites all sampling events in 2023
* Flags and Codes
  + All ammonium data flagged as <1> (CUF) rather than <0> (CUF) from 2018-2021 (inconsistent sampling days). Need to determine whether they are *actually* suspect or if they are good data, but unfiltered. <0> (CUF) was used in the earlier data when it passed QAQC but was unfiltered.
  + Secchi
    - All values >1.2 replaced with 1.2 result and flagged <0> [SAL]
    - >.6 value on 2018-09-06 10:48 GTMGL2NUT replaced with 0.6 and flagged <0> [SAL]
    - > 2.0 value on 2018-12-05 14:22 GTMMKNUT replaced with 2.0 and flagged <0> [SAL]
    - >.6 value on 2019-12-16 GTMLSNUT replaced with 0.6 and flagged <0> [SAL]
* Results
  + NAs
    - 2019 all NA entries removed from Result column and left blank and flagged <-2> [GDM]
  + Blank results
    - Flagged <-2>
    - 2020-09-14 10:08 FECCOL sample at GTMGL4NUT changed from <-3> flag to <-2> [GCM] since the data is missing, rather than having a value that is rejected.
    - 2021-04-07 GTMGRNUT OD664b/OD665a sample since missing flag was changed from <-4> [SBL] to <-2> [SBL] \*this entry should be revisited from laboratory results
    - 2023-12-11 field data for GTMGL2NUT all flagged <-2> (CSM) for now. If metadata documentation that “water level too shallow to collect WQ readings” is something that should be included in metadata. Otherwise, <-2> alone would be fine. \*revisit this data when doing metadata\*