This outlines the entire workflow for the SLS dataset:

1. The overall outline in terms of the 3 major scripts:
   1. Read relational data from Access to R: SLS.R
   2. Output list of flagged outliers: SLS\_QAQC.R
   3. Create the joined dataframe: SLSIntegrate.R
2. Detailed workflow:
   1. Reading in the relational tables from Access into R
      1. Prerequisites:
         1. Same program architecture for R as your Access database, i.e., 32-bit R if you only have 32-bit Access and 64-bit R if you only have Access.
         2. Two R packages:
            1. DBI
            2. odbc
         3. Access to the UDrive folder where the Access file is kept
            1. If remote, you can use the VPN

Note that this may cause the code to run slightly slower based on the connection speed of the VPN. This should not be too bad though.

* + 1. Specify what tables do you want:
       1. Currently, 6 relational tables right now:
          1. Catch
          2. Lengths
          3. MeterCorrections
          4. TowInfo
          5. WaterInfo
          6. Station\_Lookup
       2. For these relational tables, only select the columns that are provided in the current public facing Access database on the FTP website
          1. Complication: there are different names for some of the columns between the public database and the back-end database. If this is the case, change the name to what is currently on the public facing FTP website: two columns

EntryOrder == entryorder

YolkSacOrOilPresent = YolkSacorOilPresent

* 1. Fix column names of the various
     1. For these relational tables, only select the columns that are provided in the current public facing Access database on the FTP website
        1. Complication: there are different names for some of the columns between the public database and the back-end database. If this is the case, change the name to what is currently on the public facing FTP website: two columns
           1. EntryOrder == entryorder
           2. YolkSacOrOilPresent = YolkSacorOilPresent
        2. The procedure to pick these databases from the list of dataframes read from Access is convoluted. It involves searching for unique columns in each data table. I do not simply use the name of the table because the table names between the FTP version and the back-end version are different...
           1. Going forward, this can be simplified to just the name of the table if the option to download from the FTP website is removed
  2. Fix float issue across all data tables
     1. Find columns that may have this issue. To identify those columns, print 14 digits of the values in each column. Do the values all end with 0 or something that is not 0? If not 0, then there is a float issue for that column. These problematic columns are printed out in the console before the code fixes them.
     2. The fix involves rounding the values to the 7th decimal point. This value is due to the difference between the “single” and “double” field sizes in Access.
        1. This float issue occurs because the column should be entered as a “double” field size in Access but is instead entered as “single”. This causes a rounding issue when you read the data into R or Excel, which converts the “single” values into a “double” format. Since the two formats display the decimal values differently, the “ghost data” issue occurs due to the conversion.
        2. After rounding to the 7th decimal point, round again to the 2nd decimal point as the data is only ever entered as so (other than the kfactor).
        3. I cannot think of a way that this rounding would change the underlying data. We are rounding to such a small decimal point that it should not matter for that entry.
  3. Export the relational tables as csv files and as an rds object
  4. Import the export files into the QAQC script
     1. Specify the current season of interest
  5. Run QAQC steps:
     1. GPS coordinates
        1. Manual process
     2. CableDepth
     3. MeterReading
     4. NetMeterSerial
     5. TowDuration
     6. BottomDepth
        1. BottomDepthMonth
     7. Temp
        1. TempMonth
     8. TopEC
        1. TopECMonth
     9. BottomEC
        1. BottomECMonth
     10. Secchi
         1. SecchiMonth
     11. Turbidity
         1. TurbidityMonth
     12. Each of these QAQC steps have been checked multiple times to see if they replicate the operations in Access:
         1. Reread the actual coding for both R and Access SQL
         2. Manually checked between the R and Access outputs across all queries
         3. Automatically checked between the R and Access outputs across all queries
            1. There is a section named “Technical coding section” in which the query operations are compared. There are quirks in this section that will require manual understanding of the process.
  6. Export the outlier excel datasheet
     1. From here, the lead ES will have to use their professional judgement to fix the flagged datapoints
     2. Update the back-end as appropriate and then export the tables again (so steps 1-4 here)
  7. Join the outputted relational tables
     1. Read in the relational tables
     2. Manipulate various variables to the correct units:
        1. Secchi from cm to meter
        2. Salinity calculated from conductivity
        3. Tide into a descriptive categorical variable
        4. Bottom depth from ft to meter
     3. Create a Length\_NA\_flag column: this flags tows during which 0 fish were caught
     4. Calculate expanded length frequencies
     5. To QAQC the joins:
        1. Use ONLY full\_joins
        2. After every join, attempt to create the two relational tables used to join the tables from the joined dataset and document differences