Running the Hui Reports

This procedure assumes that you have done the “Getting Started with Hui Reports” steps and have checked out a copy of the “hui-reports” repository on your computer. As the Getting Started document stated, you should be comfortable running commands on the UNIX command line, either on a machine running MacOS or Linux.

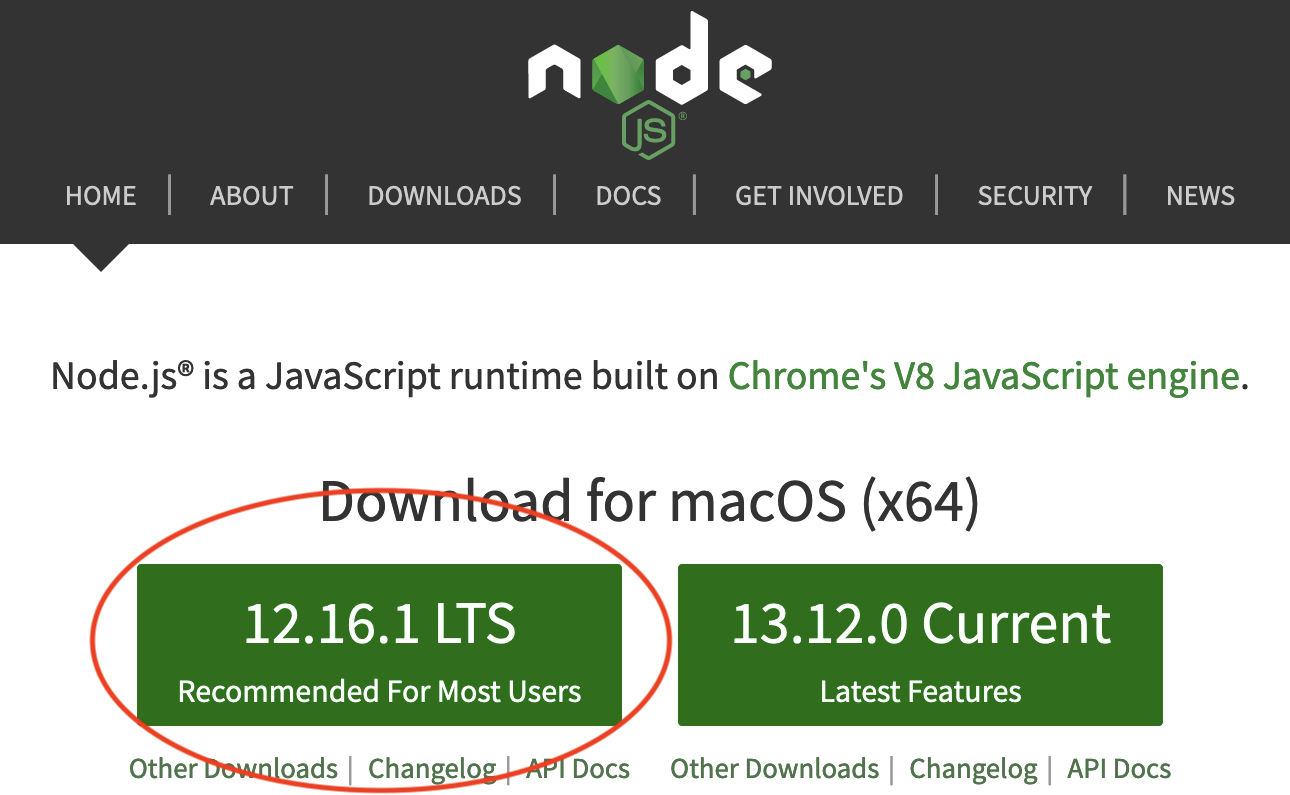
# Installing Nodejs

The only special software needed to run the reports is a package called nodejs. It installs two commands: “npm” and “node”. To see if you have them installed on your computer (you probably don’t), type in “node” on the command line in a terminal:

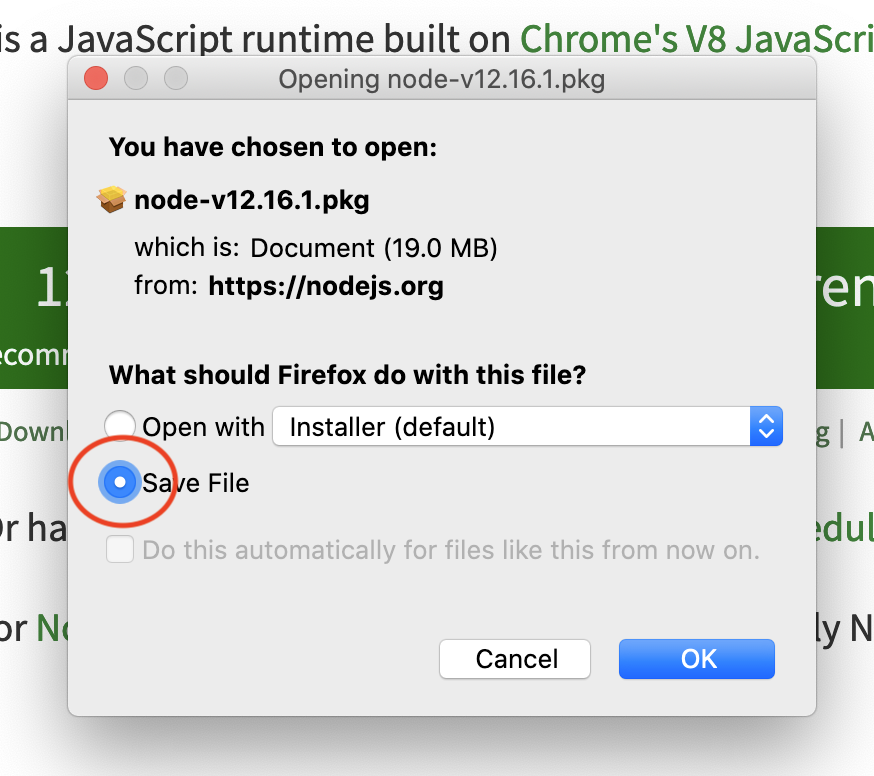
% which node

node not found

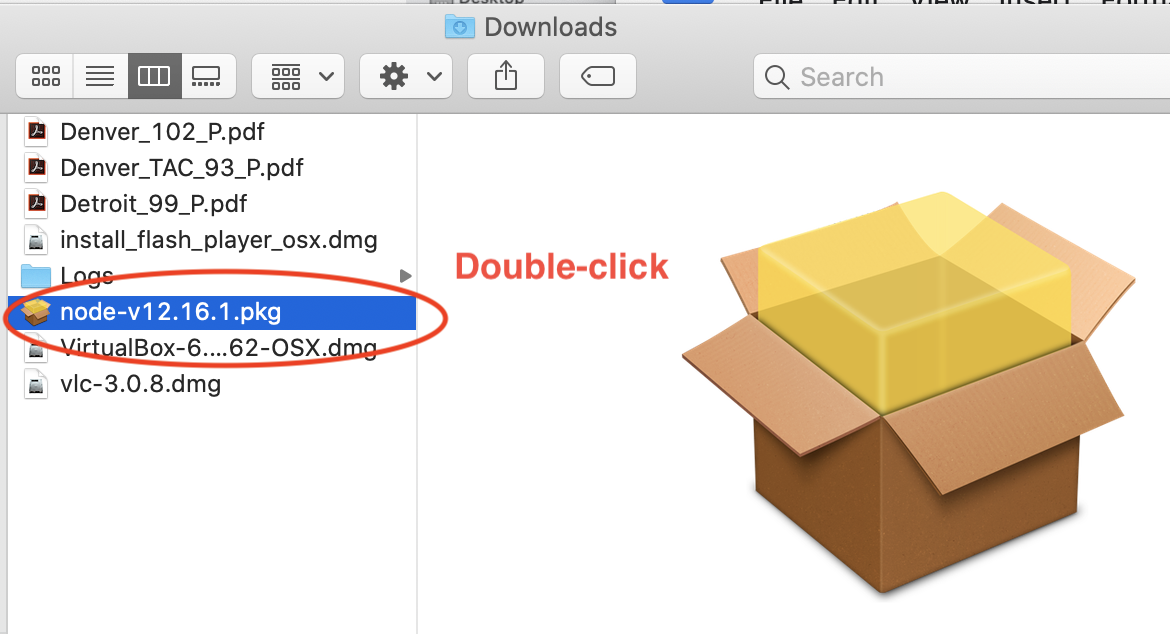
To install nodejs, go to the nodejs.org web site and get the current recommended version of nodejs. When this was written, this was the recommended version:



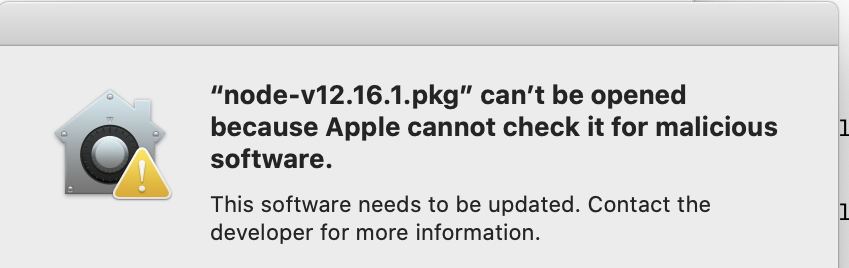
Select the “Save File” button and click OK



This will download the pkg file into your Downloads folder. In the Finder, double-click on the pkg file and go through the steps to install the software.



On later versions of MacOS, installing software has become more complicated due to increased security. If you get a message like this below, follow the steps below in “[Installing on MacOS Catalina 10.15](#_hbznpmmmbhwh)”. Not to worry: nodejs if safe to install on your computer. It is used to power the apps of Netflix, Walmart, Uber, to name a few. This installation problem is supposed to be fixed in Nodejs version 13.12.0, so you may not encounter it.



When you are done, you should be able to do these “which” commands successfully and

% which node

/usr/local/bin/node

% which npm

/usr/local/bin/npm

% node --version

V12.16.1

Now install the one nodejs node library with the “npm install” command.

npm install

# Updating the Insitu Data (Data from Google Drive)

If you are going to create some web export Excel spreadsheets for distribution, you need to download the latest data that is stored in Google Drive spreadsheet “Hui o ka Wai Ola Data Entry” and storing it in the hui-reports git repository as tab separated text files. This is the spreadsheet where the field data is entered by the team leads, commonly referred to as insitu data.

It may be in preparation to create a quarterly web export file for distributing the data to the Hui web site and other interested parties, or just updating the git repository after a two week session of collecting data in both West Maui and South Maui. It is preferable to do this process after each 2 week session to keep the amount of data to review to a smaller amount, but not required.

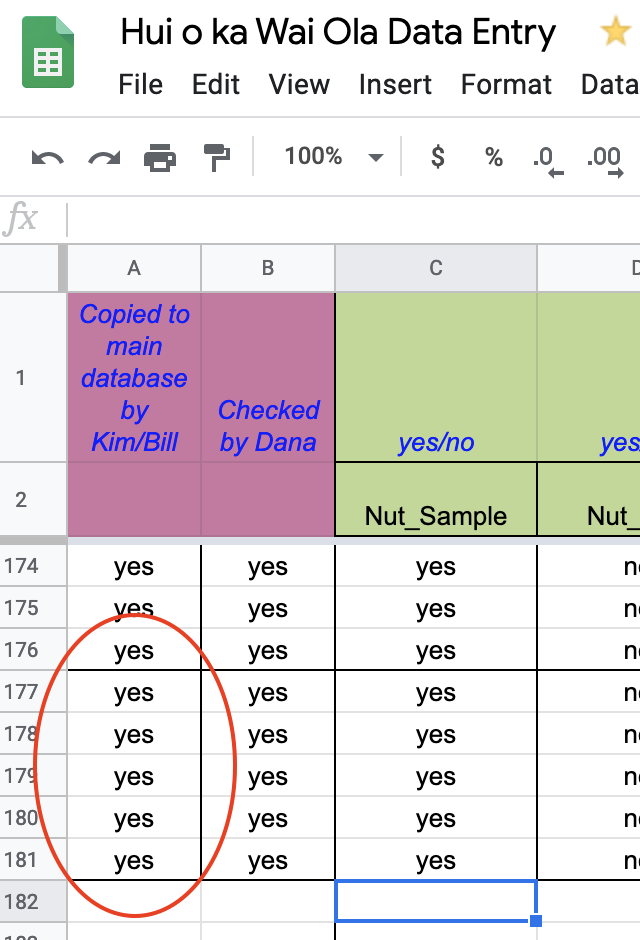
This process should be done after the QA officer reviews the data, comparing the field data sheets to the data entered in gDrive. There is a one week period between sessions where neither lab is collecting, and this is a good time for the QA officer to review things and the Data Manager to do this procedure.

It is expected that you are comfortable using the UNIX command line. The examples are done on a Macintosh computer. They should be very similar for Linux, but who knows for a Windows PC. I don’t.

The example may be from different sessions of processing data, so the date and session names for the files, etc. may not be consistent, but you should be consistent in how you name various files, using the names here as examples.

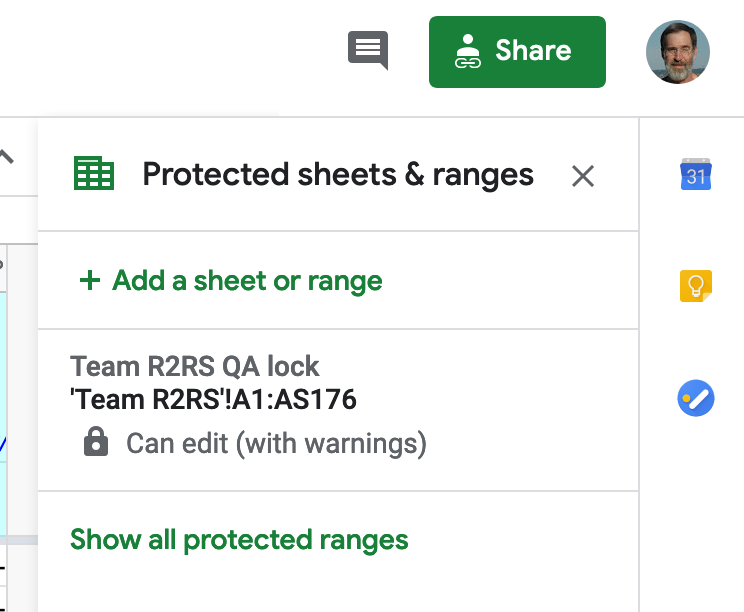
# Downloading new insitu data

Open “Hui o ka Wai Ola Data Entry” spreadsheet in gDrive. Each sheet will be saved as a tab separated file (.tsv). When it is a team sheet of data collected by that team, fill the “Copied to main database by Kim/Bill” to “yes” for each sheet. Cut and Paste from the “Checked by Dana” working nicely, and you should not be doing this unless Dana checked it.

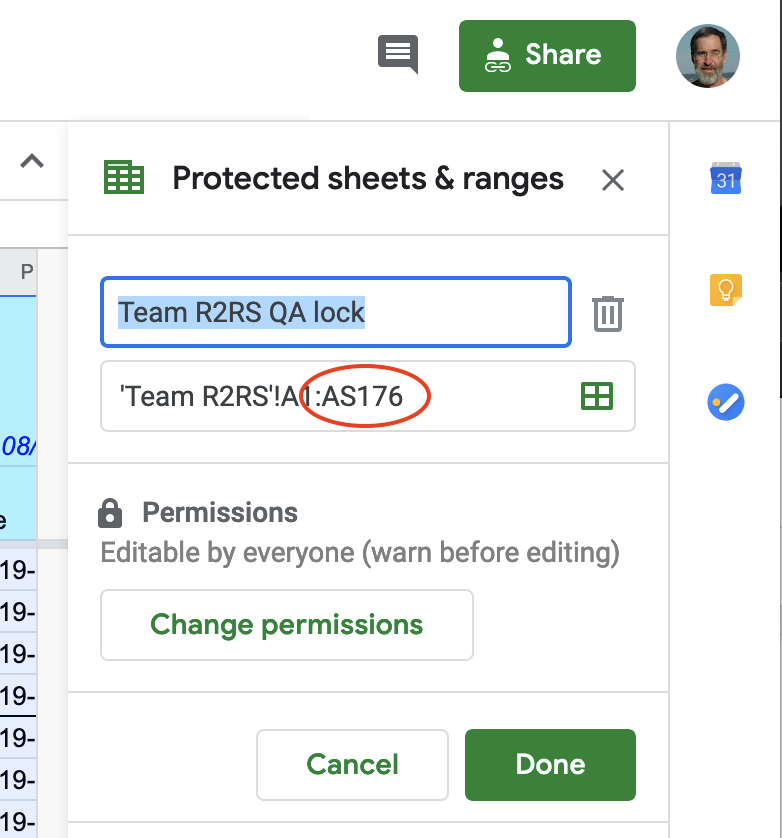


Also, extend the “Protected sheets and ranges” to include the new data. You can come back and fix things if you find a problem later. You will be warned about your change, though.

Under the Data menu, select “Protected Sheets and Data”. You will see this in the top right of the page.

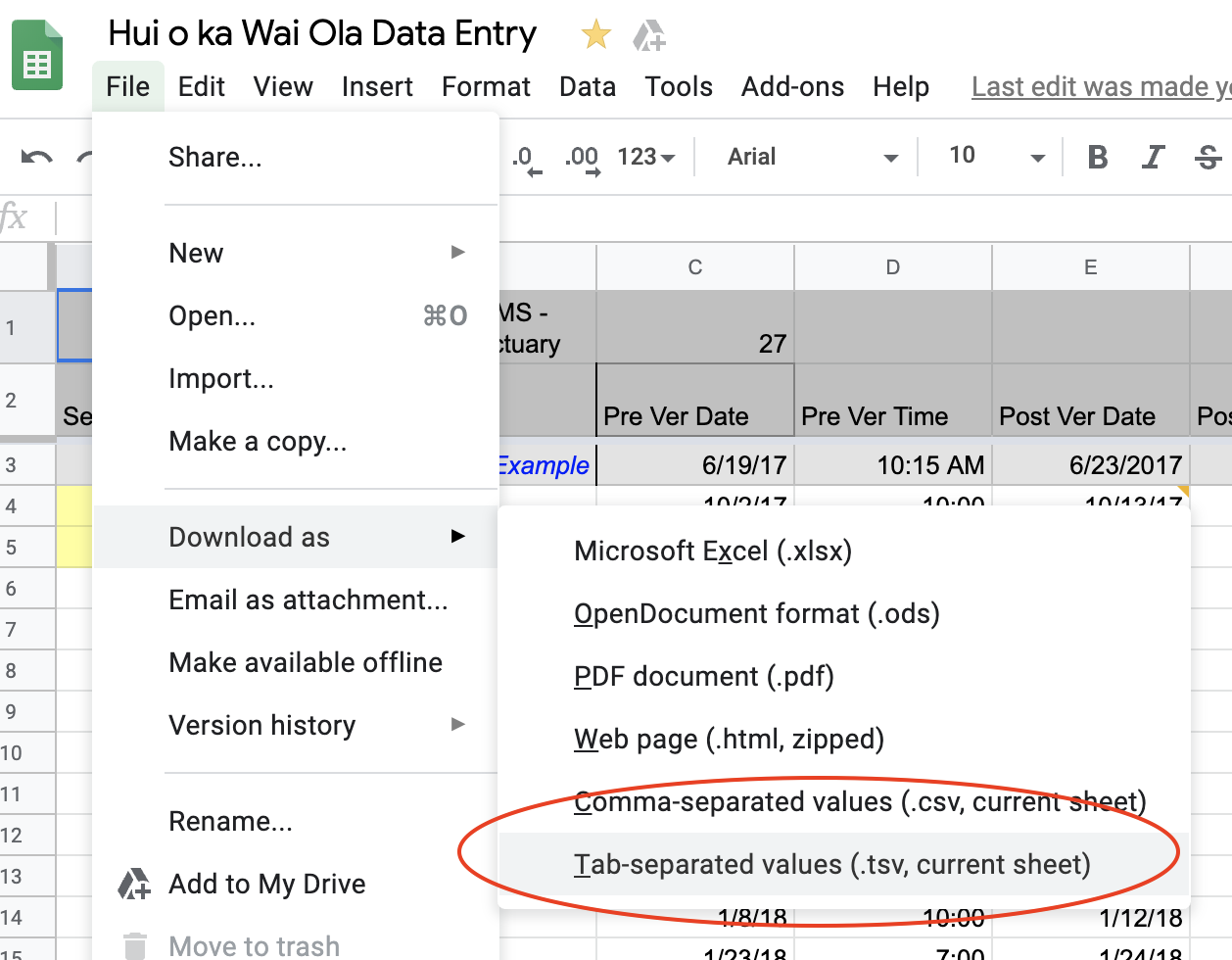


Select the lock that matches the sheet you are editing. Change the row number. In this example, it is row 181, so change it from 176 to 181, and select “Done”. After making the change, select it again and verify it is covering all the rows of the sheet. Select “Cancel” or “Done”.

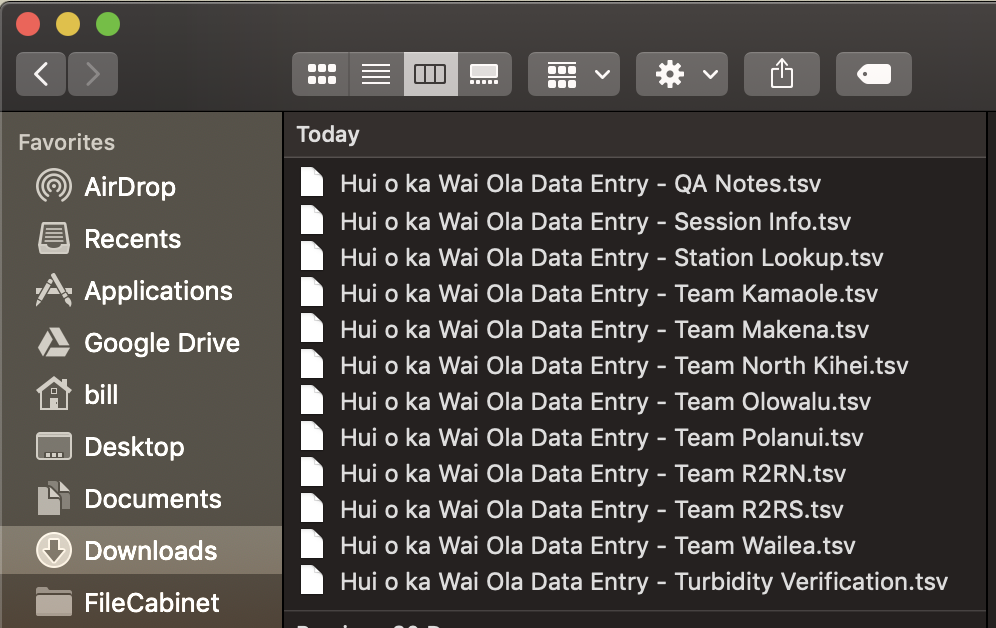


Go to “Hui o ka Wai Ola Data Entry” spreadsheet and export each sheet as tab delimited.

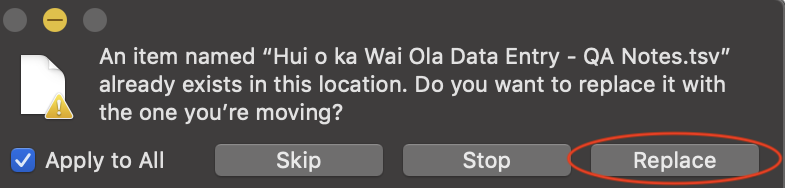
File / Download As -> / Tab Separated Values (.tsv current, current sheet). Do this for all the sheets in the spreadsheet, including Session Info, Turbidity Verification, etc. so there is a complete set. The “Protected sheets and ranges” steps need only be done for the Team sheets where the insitu data is recorded.



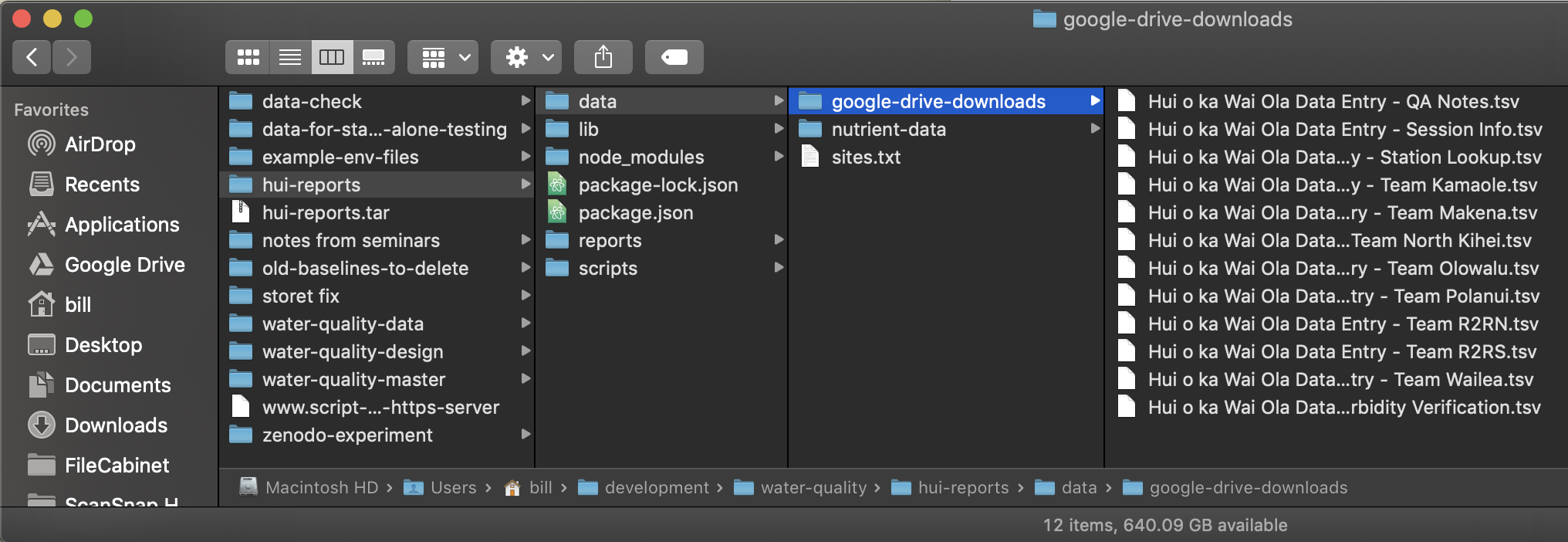
When you are done, you should have 12 files in Downloads or wherever you have your browser save your files.



Copy or move the downloaded files into the hui-reports/data/google-drive-downloads folder Should be 12 files. You will be asked to replace the files, say “Replace” with “Apply to All” button selected



You should have all new files now in the hui-reports/data/google-drive-downloads folder.



## 

# Updating the Nutrient Data

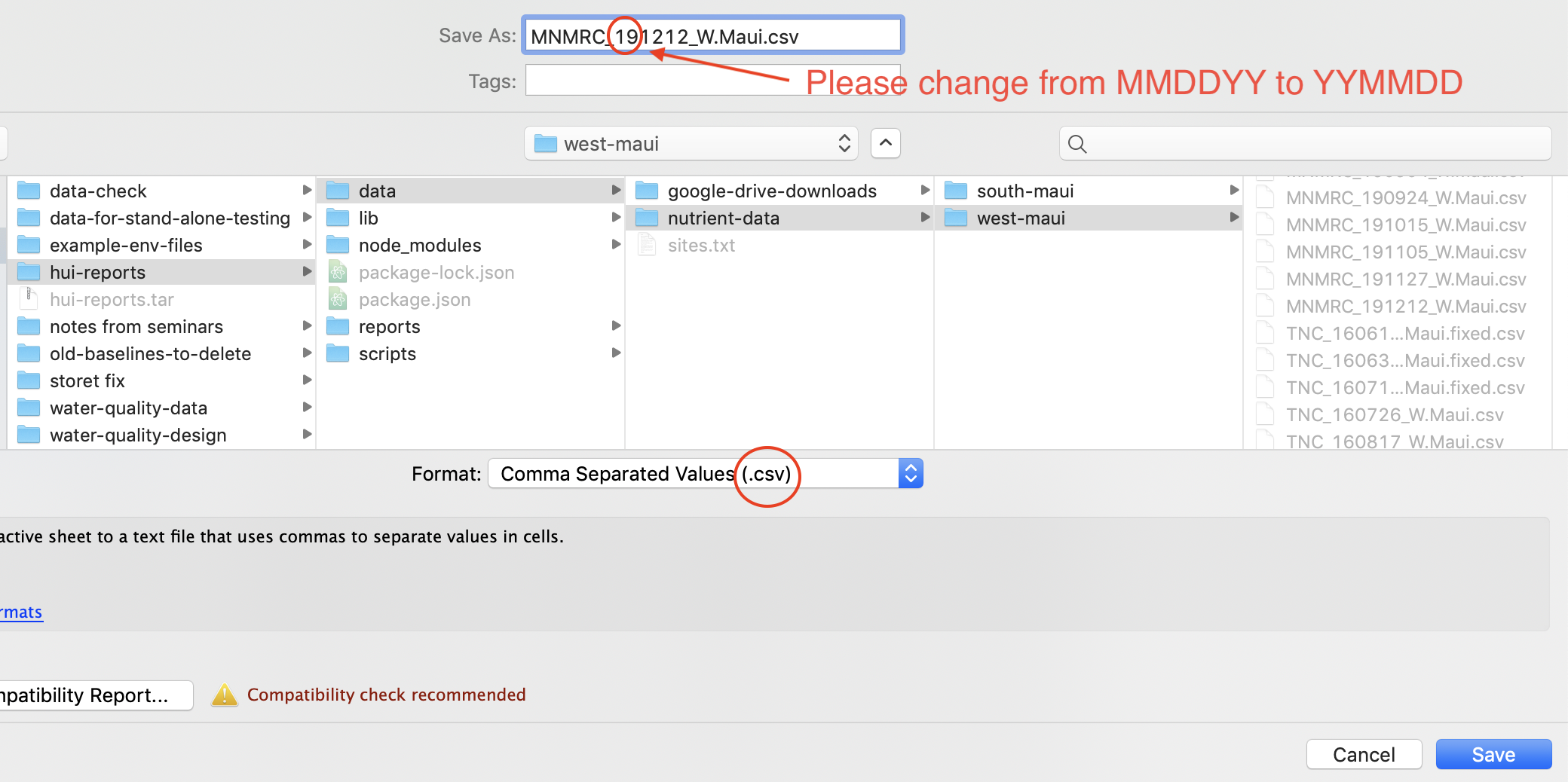
These are the steps to take when receiving new nutrient data. They currently work for data received from the SOEST lab at University of Hawaii - Manoa. They are sent as Microsoft Excel spreadsheets which are saved as CSV files (comma separated variable files) and then ingested by the script that recreates the web export files.

It is assumed that you are comfortable working on the command line in a terminal on the Mac or a Linux host.

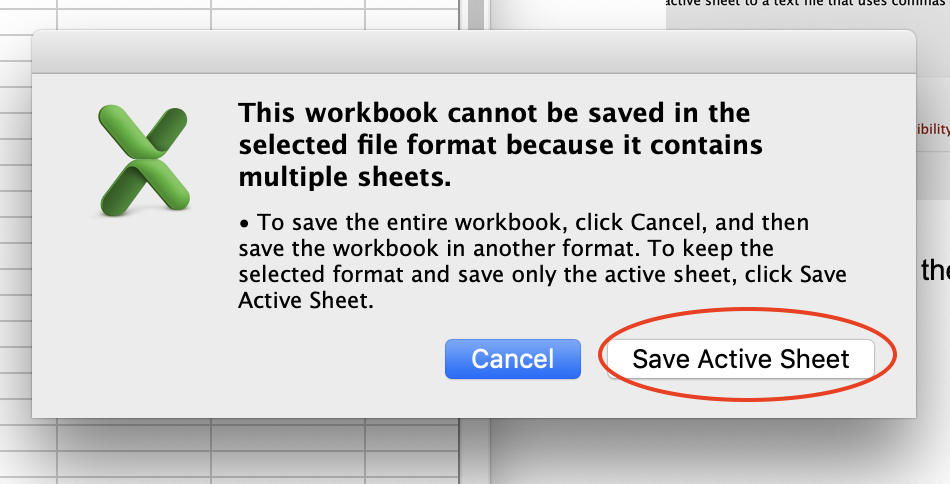
If you are also updating insitu data at the same time, you should have added that first: [Updating the Insitu Data (Data from Google Drive)](#_tdgjbyqbvwx4)

If you have new spreadsheets from the SOEST lab, open them up one at a time and save as comma separated files (.csv) into either hui-reports/data/nutrient-data/south-maui or hui-reports/data/nutrient-data/west-maui, depending on the lab.

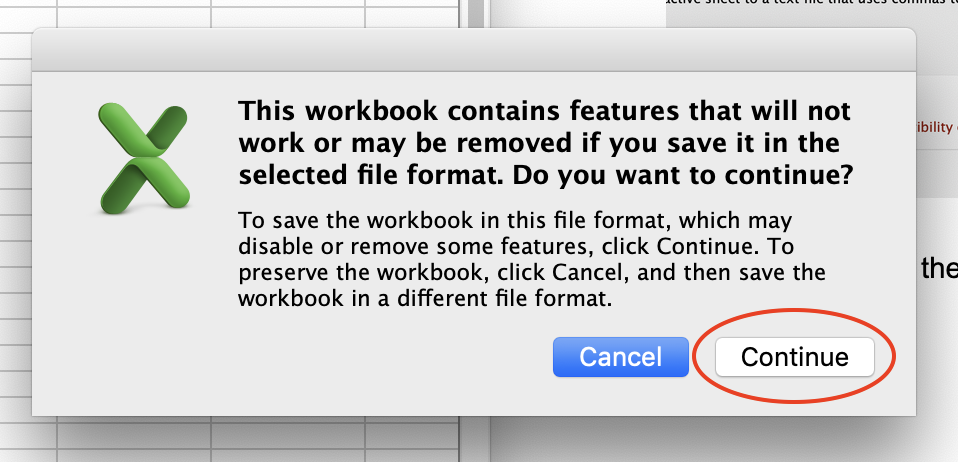
Open the received spreadsheet with Excel and convert the “Final Data” worksheet into a csv file with “save as -> Comma Separated Values (.csv)” into the csv folder for that lab (see screen capture below). As you do, please convert the name of the csv file from MMDDYY to YYMMDD. Thank you.



Don’t worry about the “This workbook cannot be saved in the selected file format because it contains multiple sheets” error. Only the “Final Data” sheet is needed. Select the “Save Active Sheet” button to continue.



Also select “continue” when prompted about “This workbook contains features that will not work or may be removed ……”



Close the worksheet. Always select “Don’t save” when prompted by Excel when you close the original spreadsheet.

Note: you will see some of the existing files with part of their names including “fixed”. Example: TNC\_170718\_W.Maui.fixed.csv. There are sometimes problems with the files and it takes some fixing by the QA person. Usually it is a problem with the sample ID, like the date is wrong. These errors need to be corrected or the nutrient data will not get associated with the insitu data when the web export file is created. The script looks for files in the directory that end in .csv and processes them. The rest of the name does not matter. Therefore, you can make modifications to the name if it helps you track problems and fixes, but it must end in “.csv” . Just don’t leave two files for the same session in the folder.

# Creating the Quarterly Web Export Files

These are the steps for creating the quarterly web export files, which are Excel spreadsheets, to distribute the Hui’s data. This is usually done on the 15th of the month after the quarter is done, so:

|  |  |
| --- | --- |
| Quarter | Date Published |
| 1st quarter | April 15th |
| 2nd quarter | July 15th |
| 3rd quarter | October 15th |
| 4th quarter | January 15th |

If you haven’t done so, make sure you have the latest data downloaded in your data folder from Google Drive [Updating the Insitu Data (Data from Google Drive)](#_tdgjbyqbvwx4) and any new nutrient data from the SOEST lab [Updating the Nutrient Data](#_qp11qwoqq9jy).

In the terminal, cd to hui-reports/scripts

cd hui-reports/scripts

Run the quarterly-run-file.sh. If you just type in the command with no arguments, you will see the usage for the script:

% ./quarterly-run-file.sh

Usage: quarterly-run-file.sh <basename for file. ex: 2019-4th-quarter.0>

Pick a basename for the script. The script is going to create 3 files: one for each lab and one that contains all labs. For example: the basename of **2019-4th-quarter.0** creates files:

**2019-4th-quarter.0**.LLHS-lab.tsv

**2019-4th-quarter.0**.NMS-lab.tsv

**2019-4th-quarter.0**.all-labs.tsv

The form YYYY-xxx-quarter.0 is suggested so that an “ls” (listing) of the directory or viewing it in the Finder by name will have them in descending order for each year and quarter. The “.0” on the end indicates it is the first version for the quarter. Usually it is the first and only, but if problems are found, you might need to re-run the reports and it will help you make sure it is the newer version. Most of the time the SOEST lab is slow getting the nutrient data delivered, so you might also do a second distribution half way through the quarter.

The files are written into the hui-reports/reports/web-export-quarterly-reports folder. It is best to direct the output to a log file for documenting the process and for comparing to the last run. The log file should start with the form YYYYMMDD followed by a short descriptive name. Ex: 20200112-4th-quarter-web-export.txt

Run the script:

% ./quarterly-run-file.sh 2019-4th-quarter.0 > logs/20200112-4th-quarter-web-export.txt

You will see messages coming to the screen. Most of these are known warnings or debug messages from the script. They are also written into the log. This is where it is helpful to have a log file each time you run it. You can “diff” the new log file against the last file and see if there are any new warnings or errors with the UNIX diff command (see [Example of diffs you might see in when diffing logs](#_xn2m9ukzzh1q)). You can see the current known warnings and errors here: [Current known warnings when running the scripts](#_11ef621uq9ll)

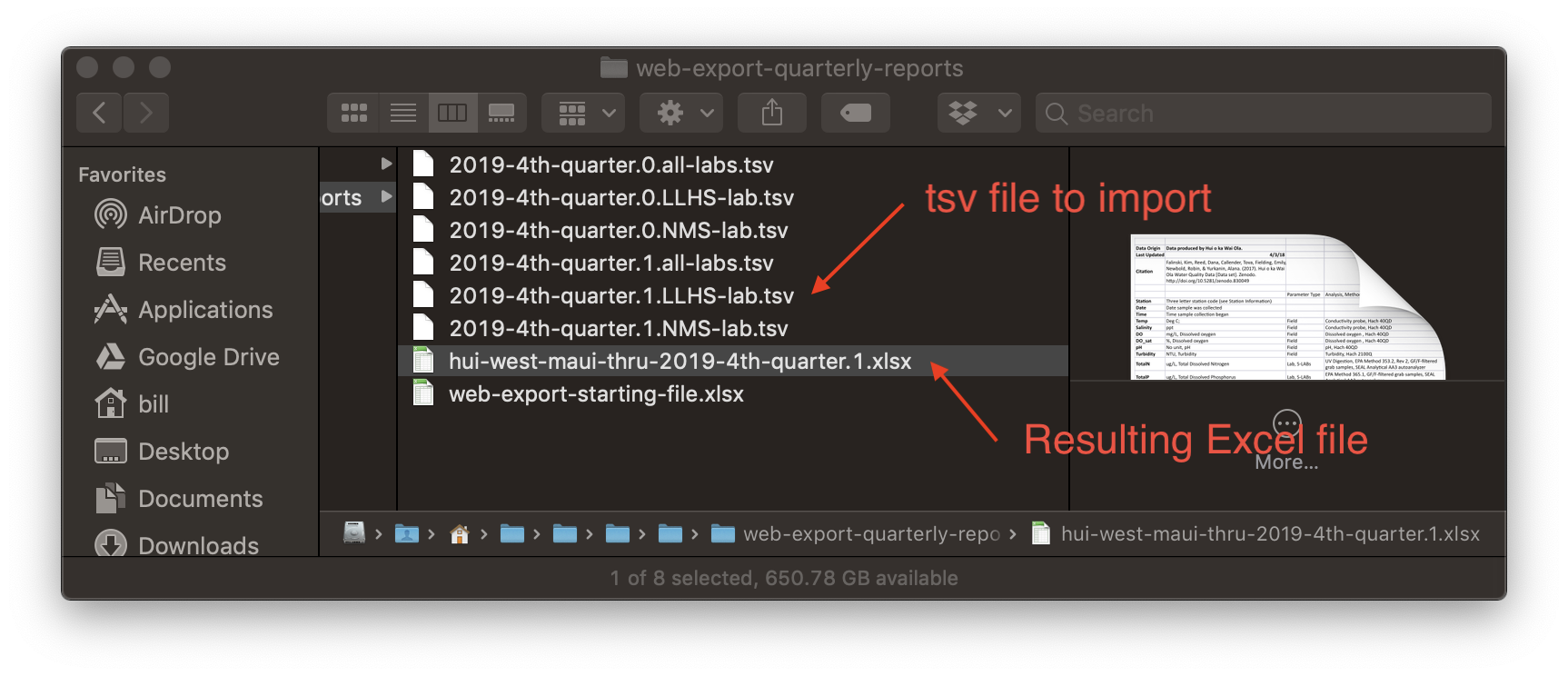
# Preparing the EXCEL File

This example is using Microsoft Office 2011 on a Mac. Your experience may vary given the version of Excel you have.

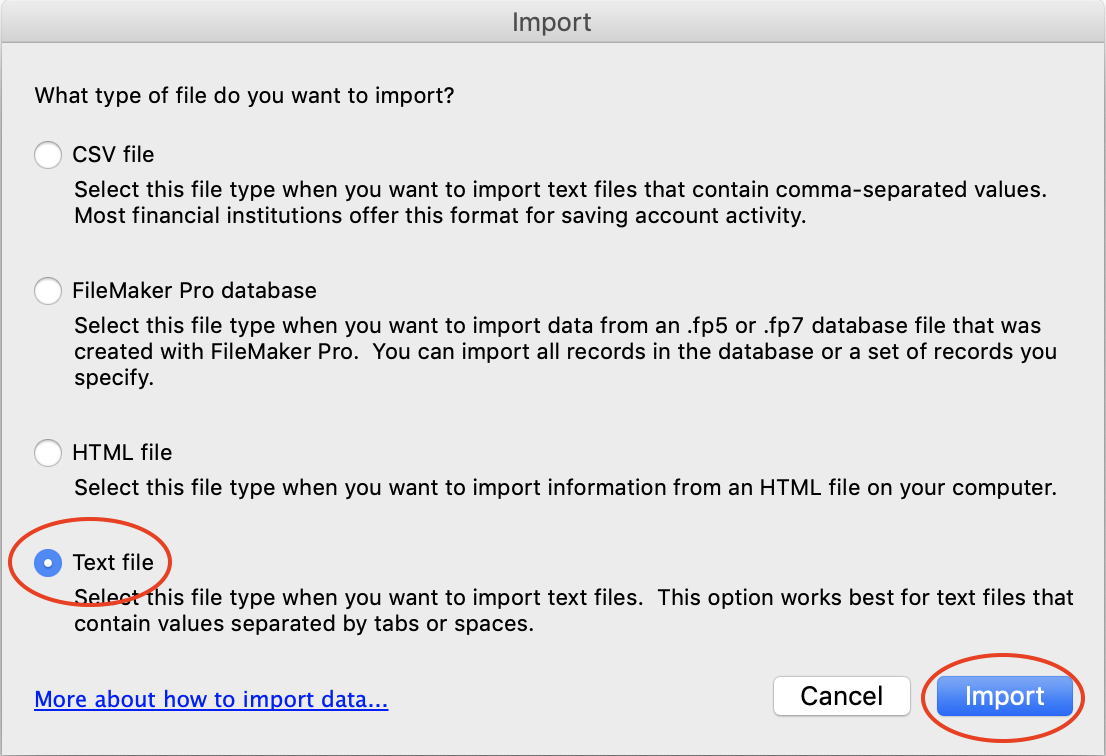
Open the web-export-starting-file.xlsx in the hui-reports/reports/web-export-quarterly-reports.

This file is important because if the spreadsheet has “use 1904” on a Mac, it will screw up the dates. Check this in Preferences -> Formulas and List -> Calculation and make sure “Workbook and Options -> Use the 1904 Date System” is **unchecked.**

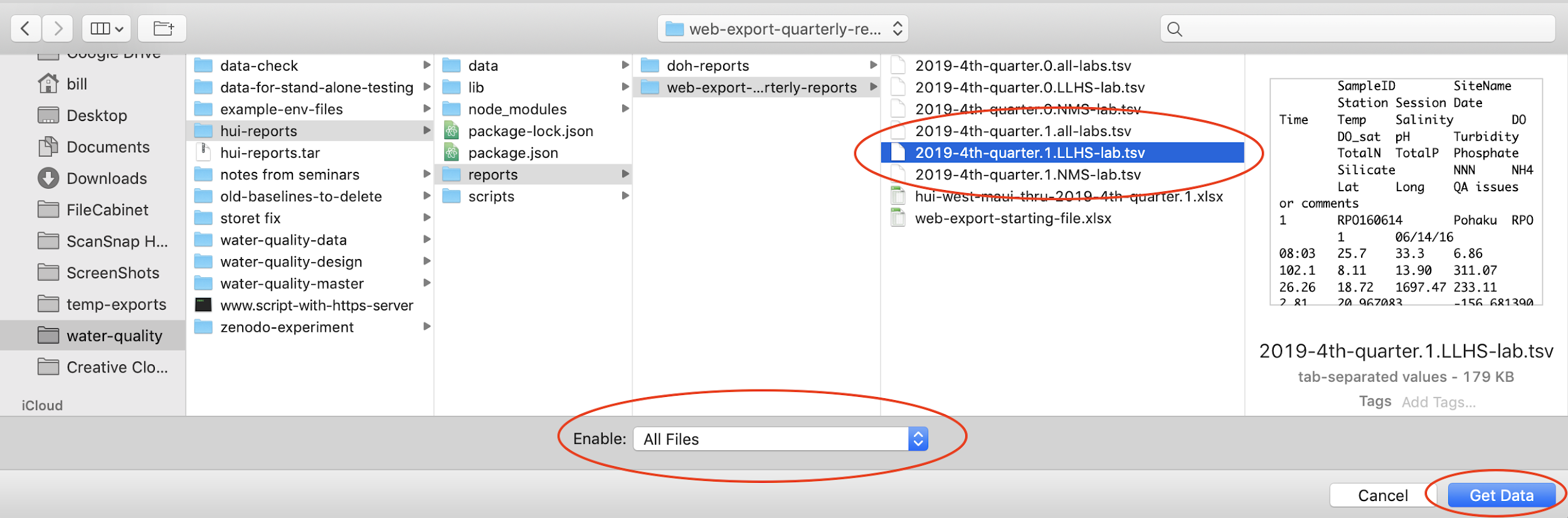
In the same folder as the tab delimited files, save as xlsx file with the name of the form: hui-<lab>-maui-thru-YYYY-<1st|2nd|3rd|4th>-quarter.<version #>.xlsx (examples: hui-south-maui-thru-2019-4th-quarter.1.xlsx or hui-west-maui-thru-2020-1st-quarter.0.xlsx).



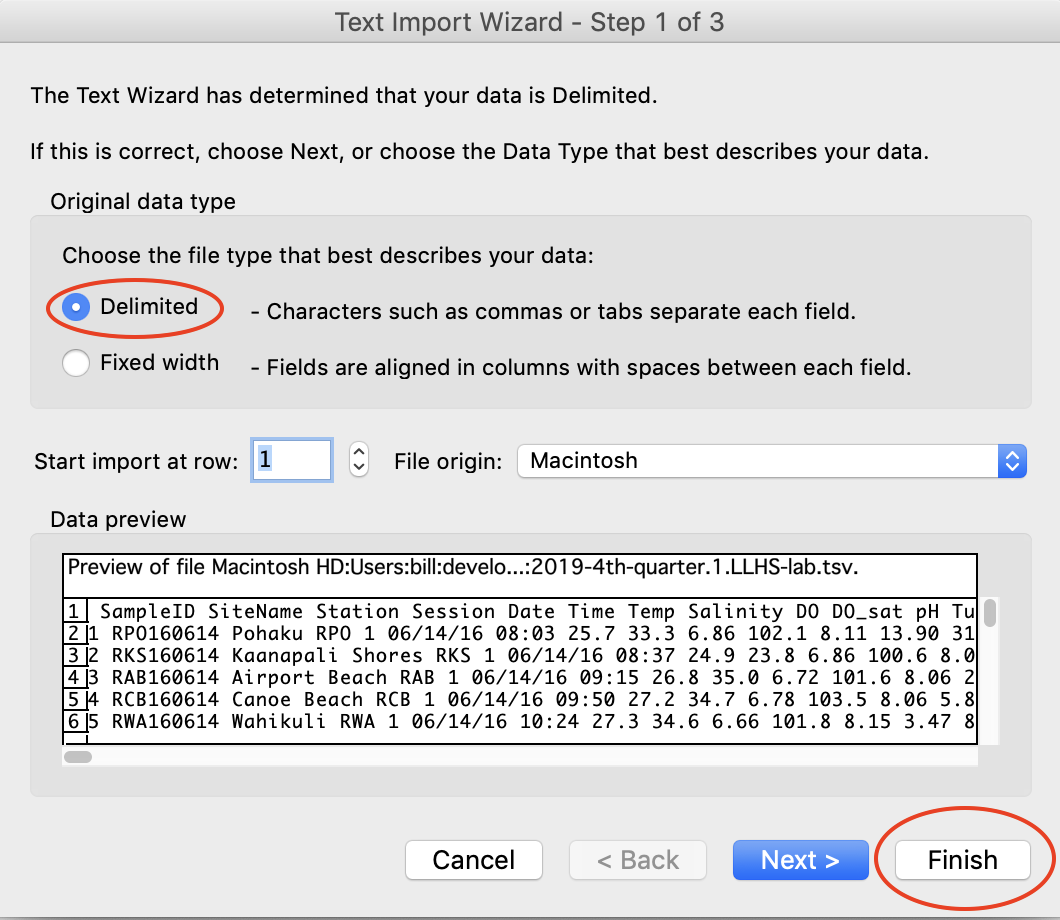
Select File --> Import… --> .In the Import popup, select the “Text File” button and select the “Import”

.

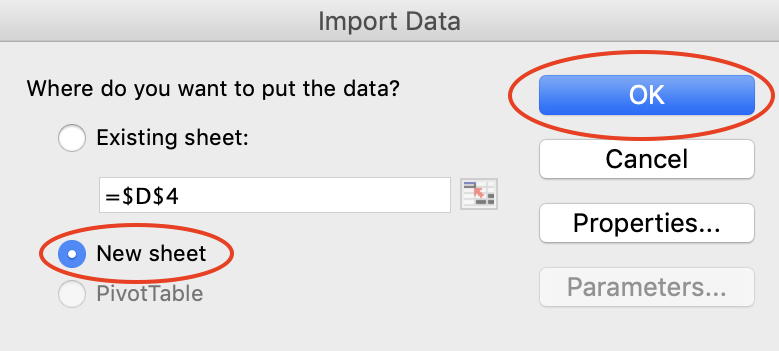
Use the file chooser window to select the \*.tsv file for the lab. You will need to “Enable all files” in the file chooser. Select the file and push the “Get Data” button.



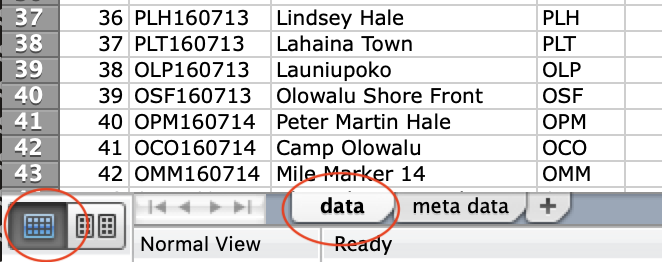
Use default of “delimited” button, and select the “Finish” button.



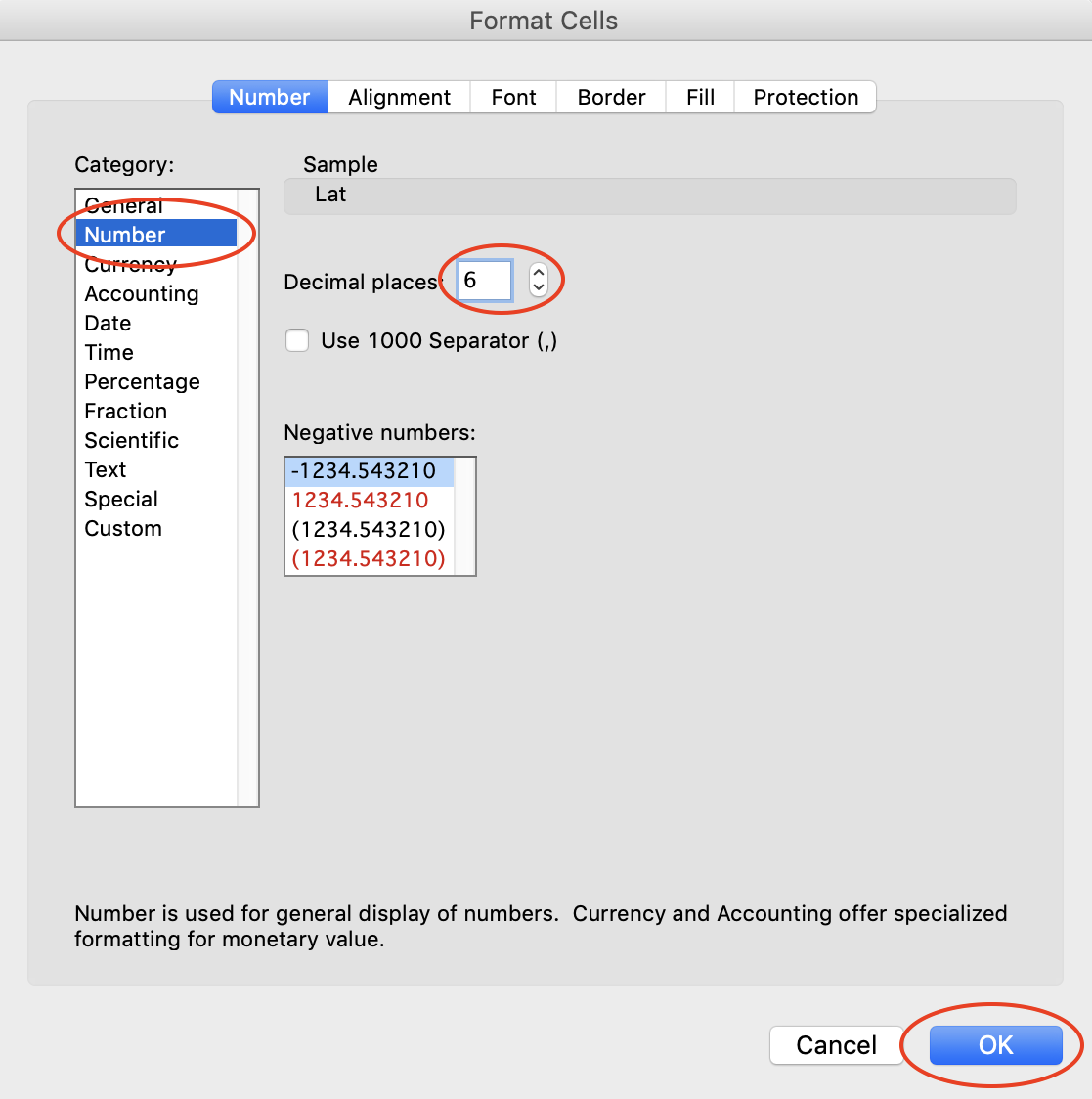
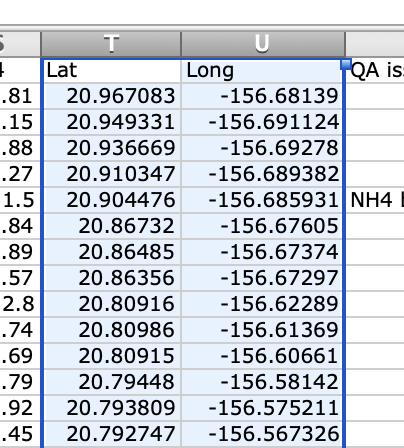
Select “New Sheet” button and OK.



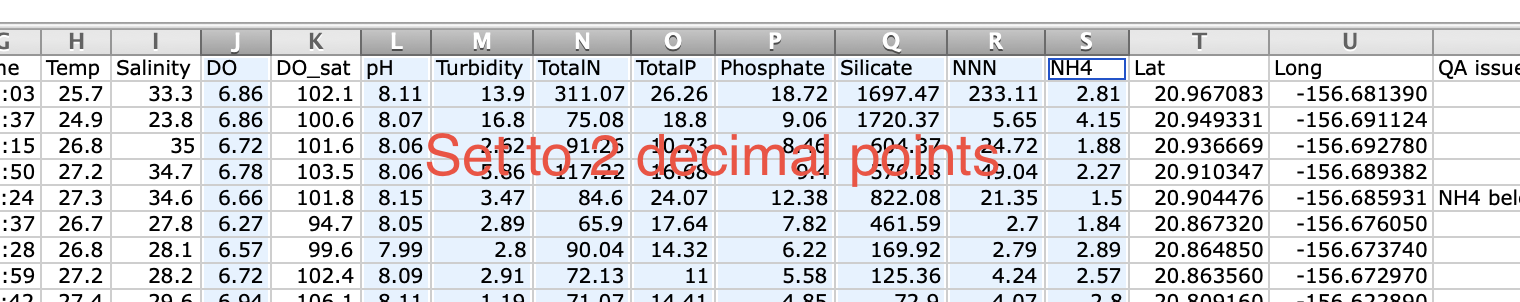
Rename the sheet “data” and change it to the “Normal” view in the lower left corner.



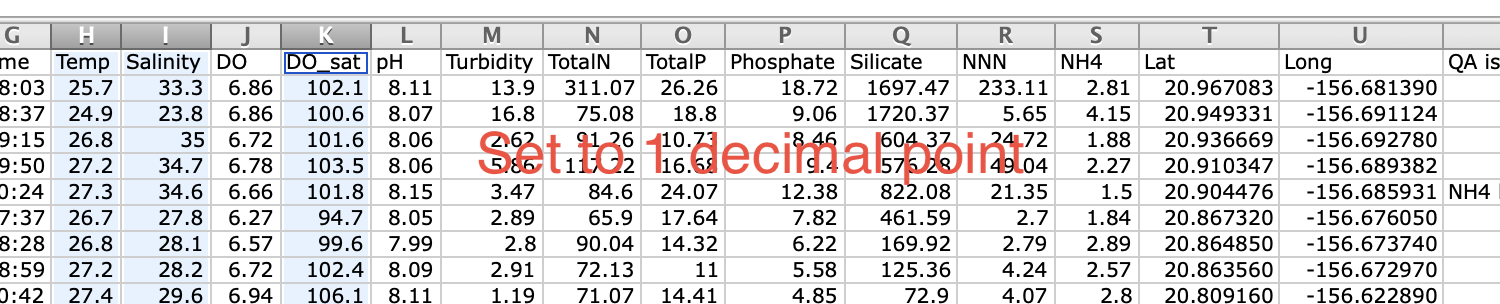
Select lat and long columns and and format them to be number, decimal places 6. (command 1 on Mac after selection).



Select DO, pH, turbidity, TotalN, TotalP, Phosphate, Silicate, NNN, NH4 (9 total) to be formatted as a number, decimal places 2.

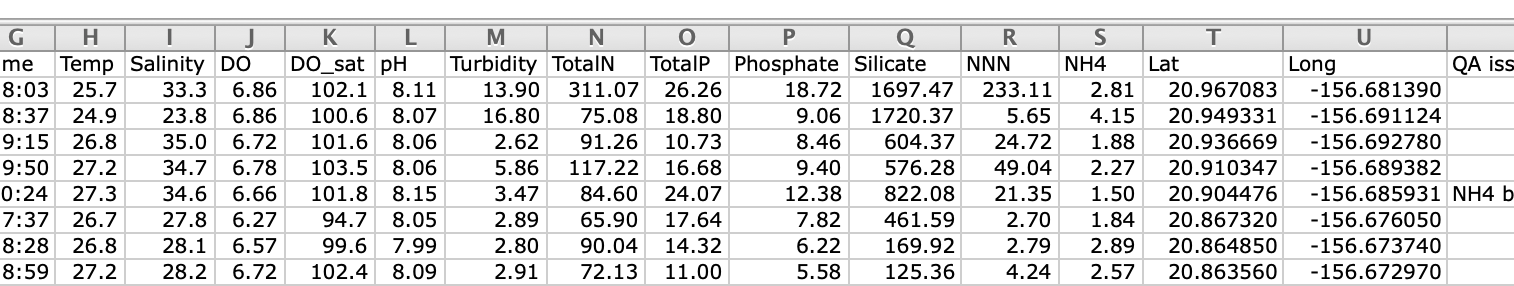


Select Temp, Salinity, and DO\_sat and format to be a number, decimal places 1.



Widen the “date” field a bit. Go to the bottom and observe the “nutrient data pending” comments and widen comments field for them if needed.

Finished decimal point adjustments should look like this:



Open the “meta data” sheet and update the “Last Updated” field.

Select data sheet again, make sure multiple columns are not selected, and save.

# Commit the changes to the repository

If you are controlling the hui-reports directory in git, add and commit your changes

In the top level directory (hui-reports)

git add .

git commit -m "first quarter report without nutrient data for sessions 53 west and 28-29 south"

Push, or upload, the changes to github so they are saved on the cloud

git push

## Tag the git repository

If this is a delivery of new reports, tag the repository

git tag -a v19.1.0 -m "first quarter report without nutrient data for sessions 53 west and 28-29 south"

git tag -l -n

To send to GitHub:

git push --tags

# Distribute the Web Export Data

## Dropbox

Place a copy of each file in Dropbox

Dropbox -> Technical Team\_HuiokaWai Ola Share -> Data\_Water Quality - TechTeamShare

On Mac - drag and drop with alt/option key selected to make sure it is a copy, not move.

## Hui Website

Email a copy to Alana Yurkanin to add to Hui website.

Email a copy to PacIOS via Fiona Langenburger: [fional@hawaii.edu](mailto:fional@hawaii.edu)

## Zenodo

Log into Zenodo as Kim

Select New version button

Delete the last quarter’s xlsx files from new version

Drag and drop new files to add them to the new version

Edit dates in the description of the Hui data

Also, edit publication date

Push the upload button

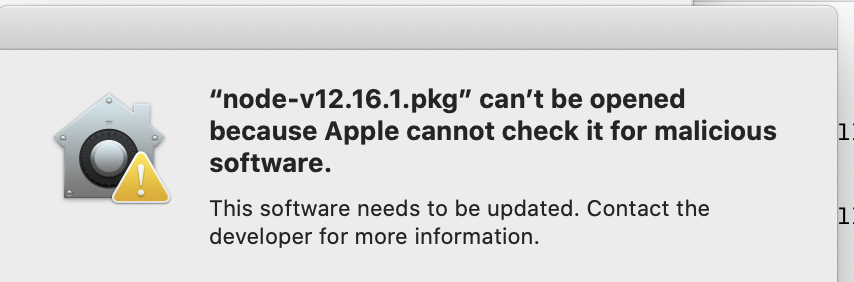
Push the save button

Push publish button

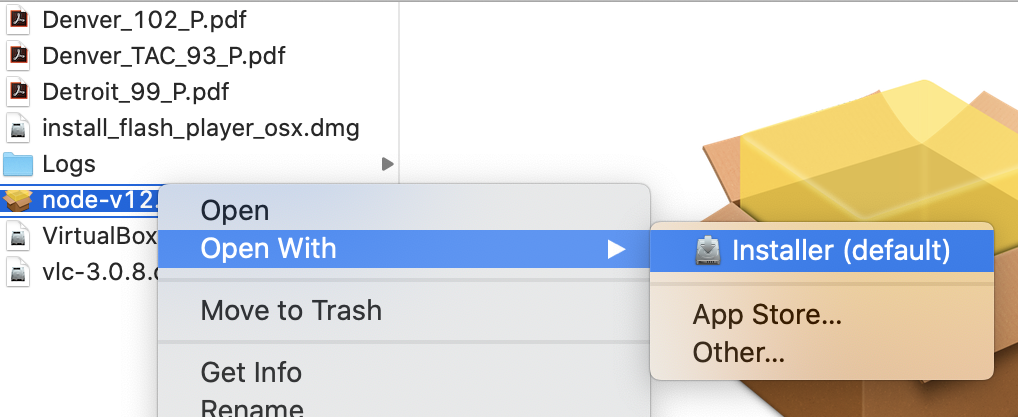
# Addendum

## Installing on MacOS Catalina 10.15

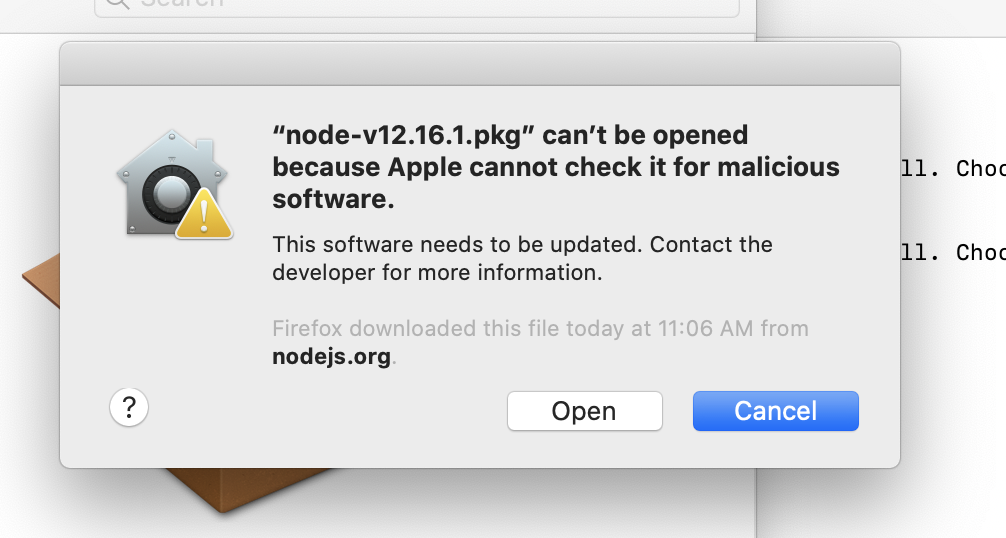
When installing nodejs on later versions of MacOS, you may get this prompt after double-clicking on the pkg file to open it.



A workaround is to use the “open” option on the pkg file. Select the node pkg in the Finder and right-mouse (or control-click) it. It will pop up a menu and use the “Open With / Installer (default) option.



You will be prompted to continue the installation. Select “open” and follow the instructions.



## Current known warnings when running the scripts

These are the know warnings and errors as of 3/30/2020

These are all sample events where the samples were actually not taken due to conditions at the site. Sometimes the samplers put in the time they arrived, sometimes they don’t. The script warns about this in case the time was accidentally left blank on a good sample set.

'' time is blank, which may be a uncollected sample, setting to NULL sample: MML180912 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MMB180912 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MON180912 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MAN180912 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MAS180912 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MML181205 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MMB181205 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MON181205 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MAN181205 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MAS181205 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: MAS190522 in Hui o ka Wai Ola Data Entry - Team Makena.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NHP180823 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NKP180823 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NSB180823 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NKC180823 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NMP180823 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NHP190117 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NKP190117 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NSB190117 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NKC190117 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: NMP190117 in Hui o ka Wai Ola Data Entry - Team North Kihei.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: OSF190807 in Hui o ka Wai Ola Data Entry - Team Polanui.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: RKO180720 in Hui o ka Wai Ola Data Entry - Team R2RN.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: RKV190607 in Hui o ka Wai Ola Data Entry - Team R2RN.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: WKD180824 in Hui o ka Wai Ola Data Entry - Team Wailea.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: WKB180824 in Hui o ka Wai Ola Data Entry - Team Wailea.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: WUL180824 in Hui o ka Wai Ola Data Entry - Team Wailea.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: WPL180824 in Hui o ka Wai Ola Data Entry - Team Wailea.tsv

'' time is blank, which may be a uncollected sample, setting to NULL sample: WPO180824 in Hui o ka Wai Ola Data Entry - Team Wailea.tsv

These are warnings about TotalP being exceeded by phosphate, which should be less than TotalP. They are close enough to allow.

-- WARNING: Phosphate of 6.34 > TotalP of 6.31 from line OCO181101-N-1,11 TNC WEST,84.35,6.31,6.34,407.53,8.89,2.93,,

-- WARNING: Phosphate of 3.54 > TotalP of 3.53 from line PPU161116-N-1,10 TNC,64.32,3.53,3.54,191.28,4.45,<1.5

-- WARNING: Phosphate of 9.83 > TotalP of 9.68 from line RPO180220-N-1,20 TNC WEST,114.19,9.68,9.83,544.20,64.31,2.28

-- WARNING: Phosphate of 12.12 > TotalP of 11.89 from line KWP181204-N-1,MNMRC 2,180.16,11.89,12.12,973.81,68.82,10.36,,

These are malformed Sample IDs, but they are malformed intentionally. They are one time samples at sites that are not in the site list, so they get flagged. This data is not to be included in the reports.

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* ERROR ERROR ERROR ERROR \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* in MNMRC\_190712\_S.Maui.csv line 35

-- ERROR: found unknown site name of NHC. from line NHC190703-N-1,21,424.22,22.03,20.89,1347.58,343.69,6.59,,

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* ERROR ERROR ERROR ERROR \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* in MNMRC\_191022\_S.Maui.fixed.csv line 36

-- ERROR: found invalid ID BADWPO191018-N-1 Line -> |BADWPO191018-N-1,22,269.11,15.94,11.65,39.65,64.52,45.73,Nutrient Data likely contaminated - bottle cracked when S-lab received sample,,|

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* ERROR ERROR ERROR ERROR \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* in TNC\_171114\_S.Maui.fixed.csv line 25

-- ERROR: found invalid ID DR171104-N-1 Line -> |DR171104-N-1,11 TNC SOUTH,640.08,215.91,184.48,\*,22.08,164.34,|

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* ERROR ERROR ERROR ERROR \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* in TNC\_180227\_S.Maui.csv line 38

-- ERROR: found invalid ID \*\*WPO180223-N-1 no sample received Line -> |\*\*WPO180223-N-1 no sample received ,,,,,,,|

Same as the TotalP warnings above.

-- WARNING: NNN of 445.89 + NH4 of 5.54 > TotalN of 346.50 from line KCP180515-N-1,4 TNC SOUTH,346.50,31.68,25.54,1581.70,445.89,5.54,

## Example of diffs you might see in when diffing logs

This is an example of diffing logs of two quarterly report runs. The “<” points to a unique line in the first file and the “>” points to a unique line in the second file. Notice there are no new errors in the differences like in the section above ([Current known warnings when running the scripts](#_11ef621uq9ll)).

diff 20200207-4th-quarter-addendum-web-export.txt temp-test.log

2c2

< 'arguments passed in { \_: [],\n odir: \'../reports/web-export-quarterly-reports\',\n bname: \'4th-quarter.1\',\n gsdir:\n \'/Users/bill/development/water-quality/water-quality-data/gdrive-downloads/data-sessions-28-66w-1-42s\',\n lfile:\n \'/Users/bill/development/water-quality/water-quality-data/legacy-data/west-maui-legacy-data-sessions-1-27.tsv\',\n sfile: \'./sites.txt\',\n ndir:\n \'/Users/bill/development/water-quality/water-quality-data/soest/csv-files\' }'

---

> 'arguments passed in { \_: [],\n odir: \'../reports/web-export-quarterly-reports\',\n bname: \'temp-test\',\n gsdir: \'../data/google-drive-downloads\',\n sfile: \'../data/sites.txt\',\n ndir: \'../data/nutrient-data\' }'

4d3

< In readLegacyFile

194c193

< Number of samples: 2112

---

> Number of samples: 2254

877c876

< Report Percision:

---

> Report Precision:

895c894

< Writing file to ../reports/web-export-quarterly-reports/4th-quarter.1.all-labs.tsv

---

> Writing file to ../reports/web-export-quarterly-reports/temp-test.all-labs.tsv

898c897

< Writing file to ../reports/web-export-quarterly-reports/4th-quarter.1.LLHS-lab.tsv

---

> Writing file to ../reports/web-export-quarterly-reports/temp-test.LLHS-lab.tsv

901c900

< Writing file to ../reports/web-export-quarterly-reports/4th-quarter.1.NMS-lab.tsv

---

> Writing file to ../reports/web-export-quarterly-reports/temp-test.NMS-lab.tsv