WGFP\_CombiningData\_RShinyApp How-To

Location: U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\CodingDetections\WGFP\_CombiningData\_RShinyApp

Web Location: <https://sfigraf.shinyapps.io/WGFP_CombiningData_RShinyApp/>

NOTE: Text files downloaded from CU, CD, CS are not in the proper format. Include spaces in the appropriate places to combine data in the app. Use Edit\_ Replace to edit. Examples below:

* 00G to 00 G
* SCDANT to SCD ANT
* CD1A1 to CD1 A1

# Uses

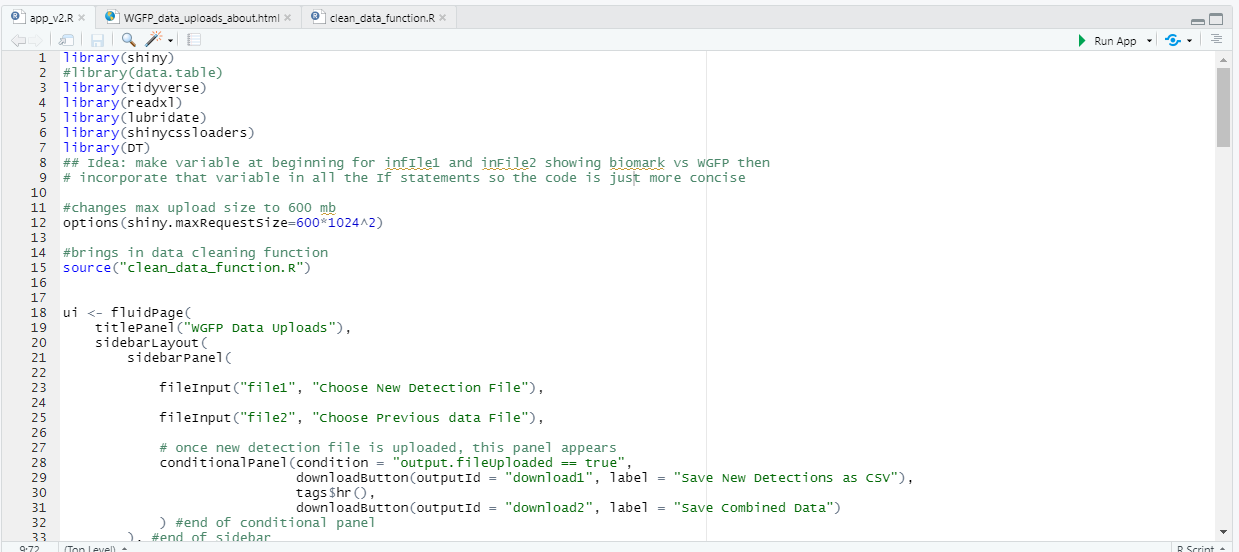
* Converting Stationary .txt files from ORSR readers to usable format and downloading it as .csv
* Combining Stationary .txt files from ORSR readers with Master Stationary .csv file
* Combining Biomark .xlsx files to master Biomark .csv file
* Combining other .csv files that have already been made/converted from .txt or .xlsx files

# How to Open

Go to <https://sfigraf.shinyapps.io/WGFP_CombiningData_RShinyApp/>

**IF OPENING FROM U: DRIVE**

1. Navigate to App location and open the WGFP\_updatedata\_app “R Project” file. Rstudio will open
2. If the “app\_v2.R” file isn’t already open in the top left panel: in the bottom right panel, navigate to the “files” tab and open the app\_v2.R file.
3. Click “Run App” in the top right of the screen. You may be required to download some packages if this is the first time running the app.

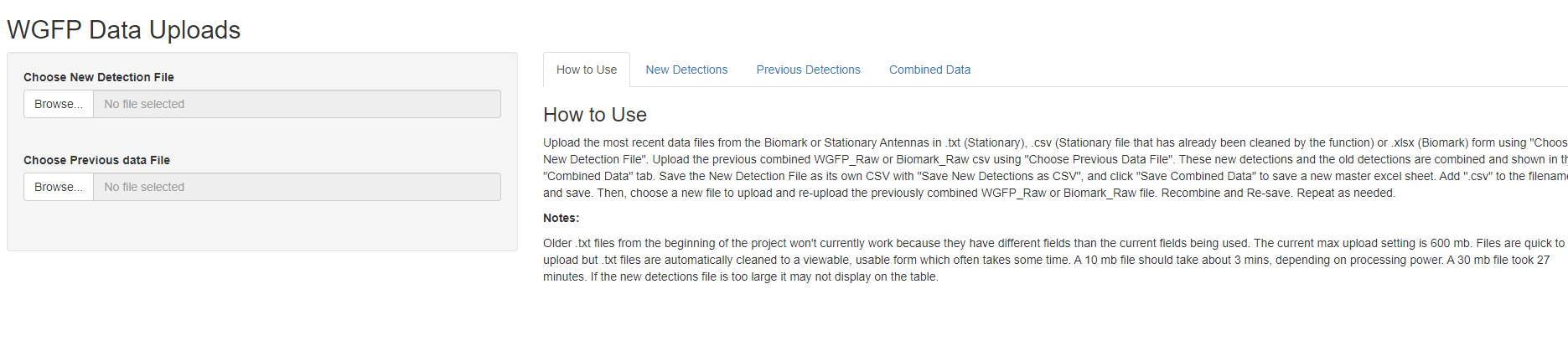


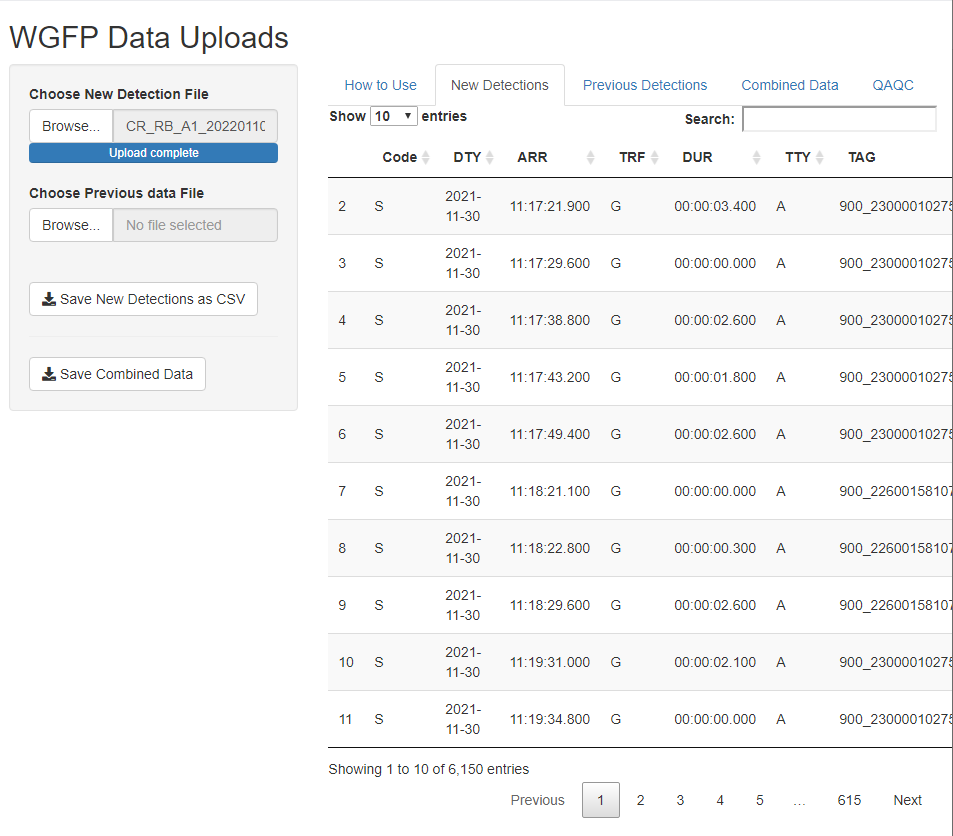
# How to Use

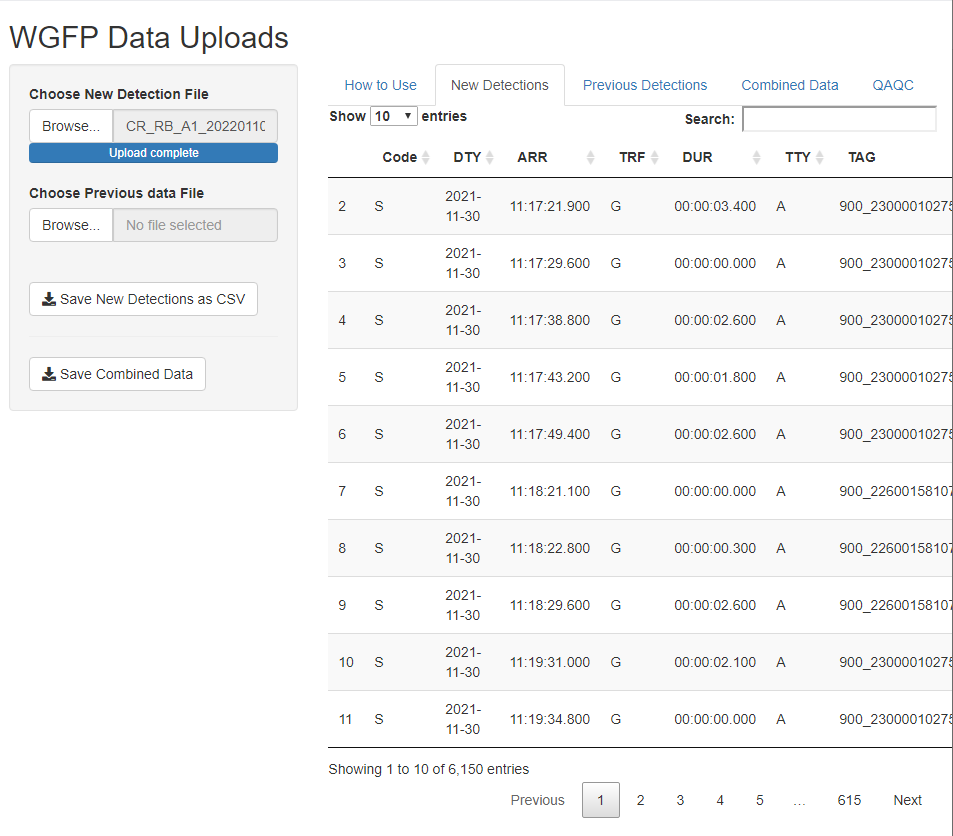
**Overview:** Upload new detections, upload previous/master file of detections, QAQC both those files, then download new detections and new combined file as CSV’s in appropriate directories. Then repeat as needed for each new detection file.

1. Click “Browse” under “Choose New Detection File and choose the most recent detection file from the U: Drive, located under U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\(STATIONARY SITE)

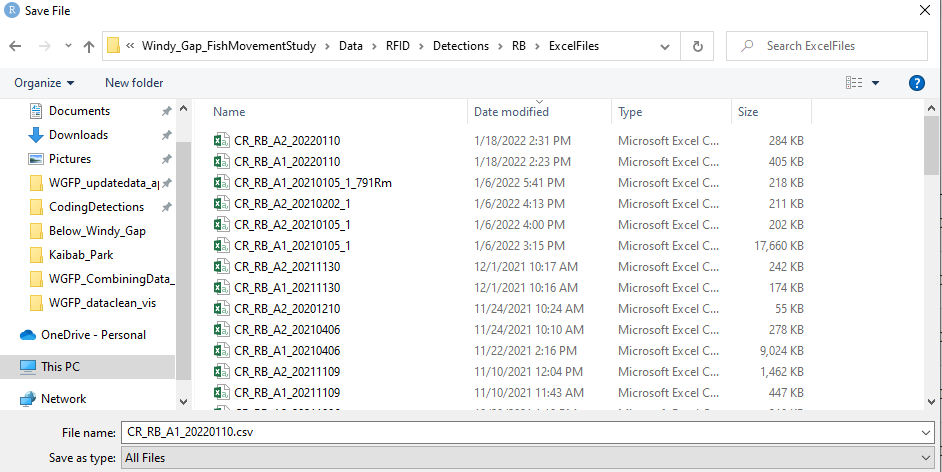
For this example, we’ll use a .txt file as obtained directly from the Oregon ORSR readers around Windy Gap.



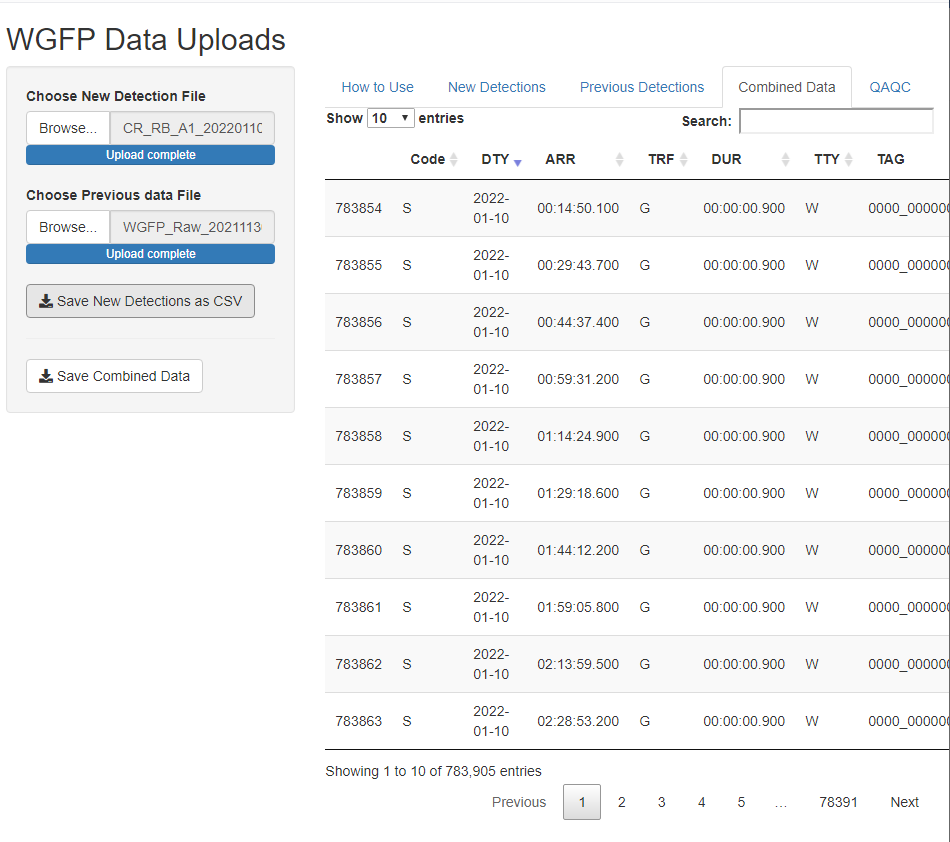
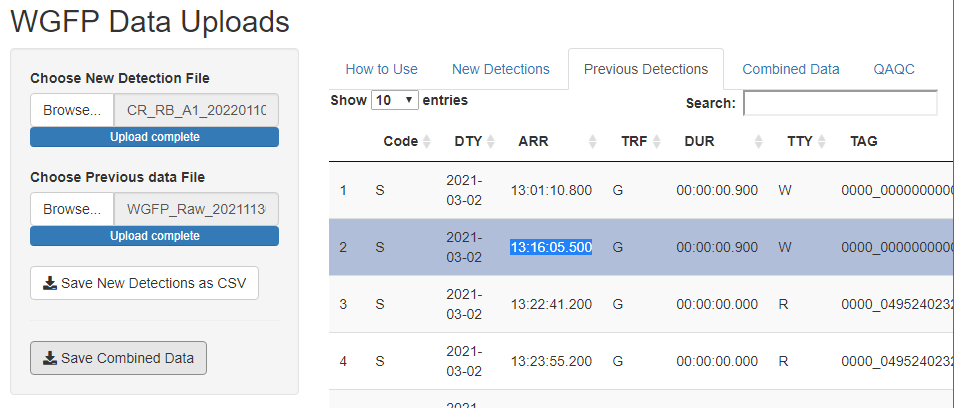
1. Once the upload is complete, the data will be cleaned into a viewable format. Depending on how big the new .txt detection file is, this can take quite some time. A .5 mb file usually takes 20-30 seconds, 10 mb file can take about 3 minutes, and a 30 mb file took 27 minutes. After the data is cleaned, it will show up in the “New detections” tab.
2. After the file is uploaded, do some quick QAQC, including checking that marker tags were detected normally, the “ARR” column was read in correctly, “SCD” column is correct, there are about as many entries in the table as you’d expect to see, and the Dates are from the last date the sites were downloaded to the most recent date. For more info, see the QAQC section.
3. Upload previous detections using the “Choose Previous data File” input. U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\All\_Stationary. File starts WGFP\_Raw.Once this is uploaded, it will show up in the Previous Detections Tab



1. Do some QAQC here too, particularly making sure the “ARR” column is read in correctly. When read in correctly, it will contain hours, minutes, seconds, and tenths of seconds. Example: “13:16:05.500”. See QAQC section for more details.
2. Click “Save New Detections as CSV”. Navigate to U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\(STATIONARY SITE)\ExcelFiles, add “.csv” to the end of the filename, and save.



1. Go to the “Combined Data” Tab, which combines what is uploaded as “New Detections” with “Previous Detections”. Check that ARR column looks Ok, the number of entries look ok, and DTY ends with the most recent date. In this case, there were 6150 entries in the new\_detections file, and 777,755 entries in the “Previous Detections” file. So you would expect to see 783,905 entries in the Combined Data tab.
2. Click Save Combined Data and Save as “WGFP\_Raw\_YYYYMMDD\_RB1.rds” in U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\All\_Stationary. This will become your next “Previous Detections” file to add the next “New Detections” file onto.



1. Repeat the process over with the next .TXT file, this time using “WGFP\_Raw\_YYYYMMDD\_RB1.csv” as the “Previous Detections” file. If using the app online, refresh the app to clear any caching.
2. Repeat the process over again, saving each new Combined file suffixed with \_RB2, \_HP3, \_HP4, \_CF5, \_CF6 until you’re finished with all the new files. Each of these files should be a bit larger than the previous file.
3. Rename “WGFP\_Raw\_YYYYMMDD\_CF6.csv” (assuming this file now contains all new detections from RB1 through CF6) to WGFP\_Raw\_YYYYMMDD.csv. Delete or archive previous iterations ending in \_SiteCode. Move previous month’s Master Combined file to Archive.

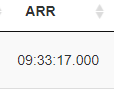
# QAQC

When combining the data, there are a few things to check for to make sure a good dataset is being made.

* .TXT files being uploaded are > 1 kb
  + This means that the data that was downloaded from the readers has actual detections.



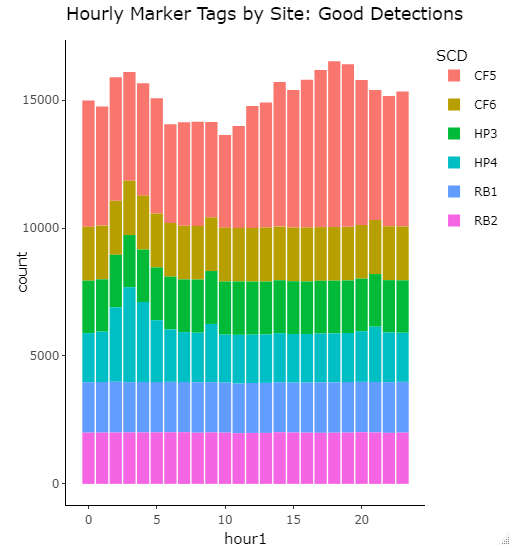
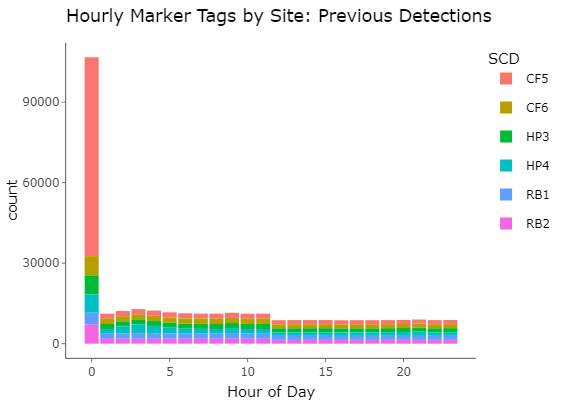
* Check that the ARR column of new detections and previous detections shows up in the format HH:MM:SS.XXX.
  + This is when the ARR column gets corrupted and can happen sometimes seemingly randomly, mainly in the previous detections file. This screened for in a couple different ways. The first way is that you can see it show up in the datatable.



Correct

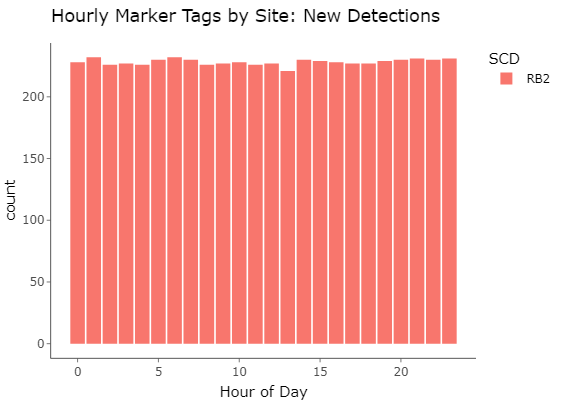
Incorrect

* + Another way is the bar plots and datatable that are found in the QAQC Tab panel. ARR is eventually cleaned into a Datetime format YYYYMMDD HH:MM:SS, but when ARR is brought in incorrectly, it will get translated with hour 0. In the incorrect example above, the time would get translated as 00:01:10, or 12:01 AM and 10 seconds. This error will then show up in the frequency hourly bar plots as shown below with an excessive number of detections occurring between 12 and 1 AM. Marker tags only are plotted to make the plot rendering process quicker.



Correct ARR column read in

Incorrect ARR column read in



If the ARR field (or in Biomark case, Scan TIme) isn't corrupted, this is what the frequency plot of marker tags should look like: pretty equal distribution of detections across all hours of the day

* The datatable in the QAQC Tab keeps track of “problem times”, where ARR was read in incorrectly. If all times are read in correctly, this table will be blank.
* The ARR column could be brought in as blank. Check Previous Detections and New Detections files to make sure this is not the case.
* If you find problems with the ARR column, the easiest way to “fix” this is to return to a previous copy of the Master Stationary file without problem times and start over from there.
* Check that when adding each new site to the master file, the new combined file is bigger in size than the previous combined file. If it’s smaller, the most likely culprit is that the ARR column got corrupted and you need to re-combine that previous detection.
* With the Biomark antennas, sometimes one file downloaded can contain duplicated info from the last Biomark file that was downloaded. This is checked for in excel by seeing if there are multiple ReaderID entries in a new biomark file.
  + Check to make sure that this in fact is duplicated data by opening the other file where you suspect contains the source of the duplicated data and cross-check to see if this is the case.
  + This isn’t a big deal because duplicated data gets cleaned later before analysis, but it is helpful to be aware of rows that will be deleted in the Master File.
* Check that you add “.csv” to the names of the new and combined files that you save.
  + If you forget, you can go back and add “.csv” in file explorer and that will enable you to open it in excel.

Things to look for: new file is smaller size than previous

Bad timestamps/not showing up

ARR is < 8 characters

Bar plot of marker tags: when ARR is cleaned up, it should have a good barplot of the marker tags hours. Otherwise there will be a bunch of entries in the 0 hour when making a datetime

When entering Biomark data, ensure the Reader ID column is correct in the original excel files. While the data will read in as B1 for Windy Gap and B2 for Kemp Breeze, they need to change to A1 and A2 respectively (just change B to A) in order to read in properly in the Get\_Movements\_Function. If they are not adjusted, you will get this error message: *Error in convUL(movement\_table\_notrans, km = FALSE, southern = NULL) : Invalid X/Y data 'xydata'. One or more columns (where NAs are not allowed) contains NAs. Columns that cannot contain NAs: X, Y.*

1. If you forget to add “.csv” and it shows up without file extension, can just change the name later adding “.csv” and you won’t have to do it again.