WGFP\_Enc\_Summaries\_RShinyApp How-To

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

# Uses

* Combines, cleans, and displays data from Stationary OregonRFID readers, Biomark antennas, mobile runs, Recaptures, release data, pressure transducers, site visits, and USGS in one location.
* Shows fish movement by day in map, table, plot, and custom animation
* Shows dataset of physical “states”
* QAQC of marker tags, lengths and weights, unknown tags, crosstalk, ghost tag movements

Table of Contents

[Uses 1](#_Toc167355620)

[How to Open 4](#_Toc167355621)

[Navigating the App 5](#_Toc167355622)

[About/How to Use Tab 5](#_Toc167355623)

[Individual Datasets Tab 5](#_Toc167355624)

[Encounter Histories Tab 5](#_Toc167355625)

[Encounter Release History Summary Wide 5](#_Toc167355626)

[All Encounter Histories 7](#_Toc167355627)

[Weekly States Tab 13](#_Toc167355628)

[States Dataframe 13](#_Toc167355629)

[States and Days Wide 14](#_Toc167355630)

[Unknown States 14](#_Toc167355631)

[Daily Movements Map, Plots, and Data 14](#_Toc167355632)

[“Movements” Defined 14](#_Toc167355633)

[Map and Table 15](#_Toc167355634)

[Plots 16](#_Toc167355635)

[Pressure Transducer, USGS and Detection Distance 18](#_Toc167355636)

[Time Series 18](#_Toc167355637)

[Movements Overlay 19](#_Toc167355638)

[Variable Correlations 19](#_Toc167355639)

[Detection Distance 20](#_Toc167355640)

[QAQC Tab 20](#_Toc167355641)

[Marker Tags 20](#_Toc167355642)

[Release and Recap Lengths/Weights 22](#_Toc167355643)

[Unknown Tags 22](#_Toc167355644)

[Ghost Tag Movements 22](#_Toc167355645)

[Crosstalk QAQC 23](#_Toc167355646)

[Detection Distance/Water Level 23](#_Toc167355647)

[Adding/Removing Dummy rows in the data 23](#_Toc167355648)

[CSV ReadIns 24](#_Toc167355649)

[Functions 26](#_Toc167355650)

[All\_Combined\_events\_function 26](#_Toc167355651)

[Spatial\_join\_function 31](#_Toc167355652)

[PrepareforStatesMovementsandSummary\_function.R 32](#_Toc167355653)

[Get\_movements\_function 37](#_Toc167355654)

[Get\_States\_function 38](#_Toc167355655)

[Shapefile/Polygon Readins 40](#_Toc167355656)

[Modifying, Updating, Adding New Antenna/Stations 40](#_Toc167355657)

[Adding to the map 40](#_Toc167355658)

[Incorporating the new antenna with the data 41](#_Toc167355659)

[In the All\_Combined\_Events \_function 41](#_Toc167355660)

[PrepareforStatesMovementsandSummary\_function 42](#_Toc167355661)

[In the Ind\_tag\_hist\_summary\_wide\_function: 43](#_Toc167355662)

[Variables.R 44](#_Toc167355663)

[Shiny Modules 44](#_Toc167355664)

[General Notes 44](#_Toc167355665)

[Very Doable tasks for the future 45](#_Toc167355666)

# How to Open

**OPENING FROM U: DRIVE**

1. Navigate to App location and open the WGFP\_dataclean\_vis “R Project” file. Rstudio will open
2. If the “app.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.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.



Note: It typically takes 20-40 seconds to read in all files from the app. If you are frequently opening and closing the app in one session, you can speed this process up by reading in the files to your local environment. These files are located in the “data read ins” section of the app.r file. To read them in, select all the code in this section, then use cntrl+enter to run this code. It takes 20-40 seconds to read in the files but once you do this, when you run the app each subsequent time, it will boot right up in a couple seconds.



# Navigating the App

## About/How to Use Tab

This tab is a concise version of this document. NEED TO EDIT

## Individual Datasets Tab

This Tab shows data incorporated into the app from different sources. (add metadata?)

* Stationary Clean: A clean dataset of all Stationary antenna data with added UTMs, no marker tags, or duplicate rows. All timestamps and dates are in the same format. 900\_ is taken off of the tags. See cleanStationary function in update data app for more info on cleaning.
* Biomark: All Biomark combined detections from Kaibab Park/Granby Diversion and both Windy Gap Sites including Marker Tags
* Mobile: All Detections from all Mobile Runs
* Recaptures: All Recaptured fish
* Release: All release data along with frequency plots for length and weight. Note: entries with no L/W listed won’t show up in the plot
* Ghost Data: a list of ghost tags that also have a “ghost date” associated with them, or the date they turned into a ghost tag
* Aviation predation: list of tags with a “predation date” associated with them when they turned into a predated tag
* Pressure Transducers: a combined dataframe of all pressure transducer data in the app
* USGS 15 Minute Data: USGS Data from their website from the stream gauge at hitching Post. As of 5/22/2024, only water temp and discharge are read in
* Site Visit Data: Combined site visit data from all stationary and biomark sites.

## Encounter Histories Tab

Encounter Release History Summary Wide: shows ENC\_Release\_wide\_summary Dataframe from enc\_hist\_summary\_wide\_function.

* This tab is meant to be a summary file for all fish in “wide” data format. Each tag will have just one entry, along with release info and information about what sort of Events it has experienced over its life.
* Filters are pretty self-explanatory. Click Render Table/Data to display new tables with filters.
* Should be same amount of entries as release file
* Columns and filters to note:
  + SiteCode\_n is the raw number of detections at one antenna, or recaptured, etc.
  + SiteCode/Recapture binary columns are just weather or not a fish was detected or captured by that method
  + TotalEncounters adds the number of unique Events (a recapture, stationary detection, mobile detection, etc). There is a filter for this column in the sidebar.
  + Through\_dam tells weather a fish has gone through the dam in its history, based on release info, recaps, and detections. It does NOT tell if a fish went upstream or downstream through the dam. There is also a filter for this in the sidebar
  + Went\_above\_dam\_no\_channel and subsequent columns are derived from the States function, which tells if a fish went below/above the dam using the channel or not from upstream or downstream. See the Get-states\_function section for more details
  + Sum\_dist is the total number of meters travelled by a fish in its history. There’s a filter for this in the sidebar.
* This is a good tab to see how many fish have swam through CRCC, fish that have swam a long way, or find fish with extensive encounter histories while being able to filter on tag number, species, length/weight, Release Site.

#### Examples

* Finding fish that have swam through the CRCC, specifically down through, that were less than 350mm.
  1. Adjust the filters pictured to desired categories and render table



* 1. Scroll to the column went\_below\_dam\_throughChannel and select “true” from the filters here



* 1. Scroll back to the beginning of the table and view the row numbers. There have been 6 fish of length 350 or below that have swam through the connectivity channel from upstream to downstream (as of 5/14/2024).



* + 1. Note that the A fish registers TRUE in went\_below\_dam\_throughChannel not solely if it is detected in the CRCC (see All Events and plot for that example), but only if the fish was actually detected in CRCC then subsequently detected (whether at RB, HP, Biomark, recapture, or mobile run) downstream of CD or CS.
* Identify Fish to explore as potentially avian predated
  + Fish that have swam abnormally long distances may have been avian predated. Adjust the distance moved filter and render table



* + Scroll to the “sum\_dist” column and click it to sort by ascending



* + Scroll back to the beginning of the table to view the tag names, then manually view the encounter history of each tag in the movements tab or “All Encounter Histories” to see if their encounter history makes spatial and chronological sense. Lots of these tags may already be in the avian predation file, so check that file in the “individual datasets” tab to see if that tag has already been accounted for.

All Encounter Histories: Shows and plots All\_Events Dataframe from All\_combined\_events\_function. These are all detections/events combined, including Statoinary, Biomark, and mobile antennas; recaptures and releases. 230000142723 has over 120k detections on its own at Hitching Post.

* Most filters are self-explanatory, but a the checkboxes are important to note:
  + Display USGS Discharge Filter allows you to filter data based off USGS discharge reading at Hitching Post within 13 hours of the detection (in winter, they record the data every 8-12 hours, 15 minutes in the summer). However, not every detection has an associated discharge reading because of this, so those rows get filtered out of the dataset immediately if you use this filter
  + Display Environmental Data Filter allows you to filter data based off Pressure Transducer readings at specific sites within one hour of the detection. However, only sites that have a transducer at them will get this data associated with their detections and only within one hour, so using this filter will automatically filter out rows of data that do not have any of these readings associated with the detections.
  + Remove Duplicate Days, TAGs, Events and UTMs condenses detections into the first and last detections of the day as well as any other antenna detections with unique UTMs. In this way, it filters out most of the redundant detections across days. For example, using this filter condenses TAG 230000142723 down to 996 entries. This equates to about 259 days of detections from May 6, 2021 to Feb 3, 2022. All at Hitching Post.
    - The df that results of this filter used on all detections is downloaded, brought into GIS, and used to make stations with, which is then subsequently used to make the “movements” df with.
    - This is one of the most “useful” files you can get from this app because of how it shaves off the dataset to the most relevant detections.
  + Remove Duplicate Tags Filter will remove instances where there are multiple detections of the same tag. This is helpful in answering questions involving “How many unique fish?”. Example: on this tab you could answer the question “How many unique rainbow trout over 250 mm hit Windy Gap Biomark and Hitching Post antennas during the day in spring 2021?”
* The plot will plot anything made with the filtered data and the frequency table will display the raw number of detections from that filtered data for each antenna

#### Examples

* How many fish released in 2023 were detected in the CRCC in Fall 2023?
  + In the filters tab, change the following filters:



And make sure the “Remove Duplicate Tags” filter is checked before clicking “render table”



* + The code filters the data based on the date range and events selected, then removes instances where the same tag shows up multiple times under these filters. The result is a df of unique tags that qualify for those filters. Find the number of fish by viewing the row number: 21



* Which antenna had the most raw detections in summer 2022?
  + Select a date range, render the table, and navigate to the “Plot and Table” tab



* + Scroll to the “Raw Detection Frequencies by Event” table to see the distribution of events during that time frame: Looks like HP3 had the most raw detections.



* + These numbers can shoot up if a fish is sitting on an antenna for a long amount of time. If we want to get a better picture of how much activity actually occurred on a daily level between antennas, we can use the “Remove duplicate days, tags, events and UTM’s” filter described above.



* + The result shows a much better representation of what is going on at the sites across this timeframe, showing instances where fish had their first and last detections of the day and all detections in between. Hitching Post saw a lot of action.



* + Why might CF5 have so much less action than CF6? To answer this, there’s a few things to check. My first thought would be that there’s a sculpin hanging out on just CF. We can go back to the “All Events” tab, select just those antennas in the “Event” filter, and scroll through the data so see if this is true



* + We see that there are a lot of different tags on CF6 during this timeframe, so that theory doesn’t work. Next thought is if the detection distance was very low, or the antenna was just off during this time. To check that data, we go to the QAQC tab.



* + Under “Marker Tags”, can check if the antenna was running. Select CF5 and CF6 from the options, select all marker tags, change the timeframe to what we were looking at before, and render the data. Then toggle the plot displays by clicking in the legend.



* + What we learn is that CF5 actually didn’t seem to be working for much of this time period. If you were to turn on CF6 on the map, you’d see marker tag detections pretty consistently across the same timeframe. This explains the discrepancy between CF5 and CF6 during this timeframe.
* Seeing a Fish’s Full Encounter History/Investigate Individual Tag for Avian Predation
  + Using one of the tags found (could be from by sorting on distance moved on the Encounter Histories Summaries Wide tab, seeing something in the states tab, seeing quick movement in the movements tab), we copy and paste that tag into the “All Encounter Histories” Tab. For example, I found 230000143683 (a big rainbow trout) by filtering for fish on Encounter Histories Summaries Wide tab that had 11 or more “events”.
  + Once the table is rendered, you can scroll to look at the fish’s history (it default sorts by datetime). This fish has 630 encounters.



Scrolling through this data, you find yourself seeing a lot of detections on the same antenna across a bunch of days that doesn’t tell you much. However, using the “Remove duplicate days, tags, events and UTM’s” filter here is helpful again, preserving useful detections and removing ones that are unnecessary. In this case, it takes the total encounters down to 104.



* + Now this is easier to view. The Encounter history of the fish started with its release 2022-10-5, hit WG1 until the antenna was taken out of the river 2022-11-02. It hit hitching post and red barn next spring intermittently, including missing hp3 in late May, a time when the marker tag on HP3 wasn’t being registered consistently (seen from the Marker tag QAQC tab). It was recaptured in June at lower Red Barn Fry Site and was registered as having grown 9mm in length. In December 2023, it made its way up the CRCC, hitting CD in December then resuming its journey in March, hitting CU March 12, then hit CF a week later where its more or less stayed since then (as of 5/14/2024). Based on this encounter history, I’d say the fish is not a tag succumbed to avian predation, just well-travelled.

## Weekly States Tab

This is where “States” are displayed. The states are defined on a weekly basis as A (below the dam), B (above the Dam), C (in the Connectivity channel), G (ghost tag detection), or P (predated by bird). A weekly “Unique event” is a filter in the sidebar, and is defined as a detection on a new antenna for that week, recapture, or release. A fish could have 6 different events in a week, but if they are all below the dam, it will just get assigned state “A” for that week.

### States Dataframe

* The States filter in the sidebar can help determine “magic” fish that may have succumbed to avian predation, since they hit multiple weird states in a week. For example, 230000228444 has one week where it transitions at confluence, then red barn, then confluence again.
* Det\_type column is useful when there is only 1 state of the week, but doesn’t accurately capture everything if there are multiple states on a week
* Date is the day that the week starts, not when the actual detection(s) happened.
* C\_number\_of\_detections is the total number of raw detections for the week. Might be 22k, might be 1.
* Weekly\_unique\_events is the total number of events that happened in a week. This can also be helpful in finding “magic” fish.
* These states are used to determine the columns “went\_below\_dam\_throughChannel” etc in “All Encounter Histories Summary Wide” tab, because if a fish started upstream then used the channel, it will have B then a C somewhere in its history

#### Examples

* Here is the fish 230000228709, a sizable brown trout.



* Finding Avian Predated Fish to Investigate
  + Use the “States” filter to view all the different states that fish have been in in a week. Select the most dubious like ABABA, BAB, etc.



* Copy Tag number and investigate in the “All Encounters” or “Movements” tab



### States and Days Wide

This is the same data displayed in States\_dataframe, but in wide format where each unique tag gets a row and each column is a day since the beginning of the study. This may have a bit more rows than the release file because there are some fish in there without release info, but should typically have the same

### Unknown States

This is where events that were unable to be captured by states are displayed. This hopefully should be pretty small...it’s filled with tags with detections before official 'Release' such as in in May 2021 and tags without release info. Mainly tags without an idea where they came from. But if a Tag shows up where release info is known, might have to go into the get\_states\_function.R code to make another case\_when entry to account for the new state. All in all, this is a check to see how well the get\_states\_function is working, and also to see if fish without release info have gotten by the other filters.

## Daily Movements Map, Plots, and Data

Displays data of “movements” (Movement\_table\_no\_transitions from Get\_movements\_function), where a fish has changed position along the river.

### “Movements” Defined

Movements are inferred by encounters of the same tag on different antennas (including mobile runs) or recaptures throughout the river. Each detection along the river is assigned a “station”, breaking up the centerline of the stream into 10 m increments. The stations start at 0 m, starting about 150 m below the Sherriff Ranch Fry Site. They extend to 13480 m on the Upper Colorado River, where the mobile run begins, and to 22850 m (10120 m at confluence + 12730 m from confluence to Fraser River canyon) on the Fraser River (5/23/2024).

These stations, which are made in GIS from stream centerlines, are joined by coordinates to a dataset of all relevant detections for each unique UTM, using the Spatial\_Join\_function. The resulting dataset is passed back to a larger dataset of all detections to assign stations to all UTMs. From there, the data is condensed down to a dataset of daily fish detections, keeping the first and last detections of the day and one detection in between. Wacko fish/avian predation that hit multiple different antenna in between the first and last detection of the day won’t display all of these detections. Example: tag 230000272182. These CAN be seen in the states and All Events dataframes though.

Once there is a dataset of daily movements, the distance moved between days is calculated for each fish when a new detection is recorded, calculated by taking the difference between stations. An upstream movement is a positive difference (IE new detection station = 8000, previous detection station 6000, so 8000-6000 > 0; upstream movement). A downstream movement is a negative difference between current station and previous station (new station = 6000 m, old station is 8000 m, -2000 <0; negative movement). No movement is if there is no difference.

The absolute value of total distance moved is summed for each fish and displayed in the movements tab under “sum\_dist”. It also seen in the Encounter Histories created in the Ind\_tag\_enc\_hist\_summary\_wide function.

* Filters in the sidebar control what is displayed on the map or plot
  + USGS Data is combined with movements on a *daily* level, contrasted with events in All Encounter Histories where PTdata and USGS data is associated with an event within 13 hours for USGS data and 1 hour for PTdata
  + Like the Environmental Filters and USGS data filters in Encounter Histories Tab, using the filters for Release length, USGS WaterTemp, and USGS Discharge will automatically filter out rows that don’t have that data associated with the movements. All the Tiger Musky in the release file for example don’t have release length/weight data so they don’t show up on the map/table/plots if the release length filter is on

### Map and Table

* This tab can be used to answer questions like “what time of year are fish moving?” and is the easiest/funnest way to examine individual tag histories. A method I frequently use is to find fish that have moved a long distance or have weird daily states recorded, copy and paste the tag, and plug it into this tab. Or find an event on the map, click on it, copy the tag number, and plug that in.
* Map is interactive, where you can click and hover to get more info about a movement. Clicking on an event in the map will navigate to the event column in the table, and vice versa.
* Can be used to help find suspect avian predation tags by clicking/sorting on the dist\_moved and MetersPerSecondBetweenDetections columns



* Layer control is available in top right, controlling detections, antennas, Release Sites, Stream centerlines, Stations that were used to make these movements, and mobile reaches. Note that loading the Stations takes about 10 seconds and can also slow map performance.



### Plots

* The plots are also controlled by the sidebar filters
* Fish Movement by Day: simple histogram representation of movements across time, colored by movement type
  + Note: data can be misleading if mobile detections aren’t filtered out, as these cause big spikes in movement activity the day of the mobile run. Activity is also generally higher after release events as well.



* Seasonal Daily Movements: Same plot as above but grouped by day across the years, which is helpful in seeing seasonal movement trends (big spikes here are probably mobile runs)



* Individual Movements: showing all the times a fish has moved (upstream, downstream or changed rivers) upstream and changed rivers are positive integers, downstream is negative. Note the spikes around 2200m due to fish travelling between red barn and hitching post, a distance of 2190m according to the stations.



* Cumulative Movement: the amount each fish has moved in absolute distance, in meters, in total over the course of its life.



* Movement by hour: displays how many movements have been recorded during each hour of the day. Can be skewed towards hour 0 (12:00 AM) especially if a fish is hanging out on an antenna, since the movements df take the first and last activity of a fish during the day as well as unique events in between.



* + If “No Movements” and “initial release” are filtered out from movement type, and “mobile detections”, “release” and “recaptures” are filtered from movement types, you get movements registered by antennas in the river from fish that are not hanging out on antennas for weeks at a time.



* The animation tab lets you render an amination of filtered data from the left sidebar. Click “render” in the left sidebar before clicking the “render animation” button. There is a progress bar in the Rstudio terminal that shows if you minimize the map. It automatically saves a .gif in the parent directory called WindyGapFishMovements.gif

## Pressure Transducer, USGS and Detection Distance

This tab shows graphical representation and overlays of data from Pressure Transducers, USGS and Detection Distance from Site Visits.

* Water\_level\_NoIce is the best variable for water level. Ice readings/observations are displayed as 0.
* USGS data is all from the gauge at Hitching Post. It can be misleading if you select a site and see discharge, but it’s all from the same gauge.

### Time Series

PT data with the option to overlay USGS data. You can select variables to plot and which Y axis to display it on



### Movements Overlay

PT/USGS data overlaid with the same movements df displayed and used in the movements tab. All the same filters are available for movements, and the graph used is the Fish Movement by Day histogram of movements across time. Use the tabs in the sidebar to get to different data filters.



### Variable Correlations

Plot PT/USGS variables against each other. “Summarize by timeframe” is an option used to group data by specified timeframe and can get rid of daily fluctuations. R-squared value in the left graph corner is calculated and found with this function in R: rquared = summary(lm(variableX ~ variableY))$r.squared



### Detection Distance

Detection distance from site visits of Biomark and Staionary Sites with PT data overlaid. You can select which reading you’d like to plot from the sidebar. You can change the opacity of the lines for preferred representation as well.



## QAQC Tab

Used to help verify data.

### Marker Tags

Stationary and Biomark Marker Tag Data set from All\_Combined\_events\_function

* Helpful for finding periods of inconsistent marker tag detections, indicating antenna was not functioning properly during these time.
* In this case, there was a gap: closer examination shows there were not detections between May 21 2022 and August 9 2022 at CF5



* These found gaps are manually recorded in the “Marker Tag Issues” tab in the metadata and are displayed in the “Downtime Periods” tab here



* This is a large dataset to plot with a scatterplot, so it will run quickest when plotting a smaller date range with only one marker tag or site selected.
  + A smaller date range will get better resolution in the plot, making it easier to see downtime periods with the tag.
  + What I do to find these periods is examine a small date range of about 2-8 months, select just one site, then select all marker tags and render the plot.
* If the ARR column is brought in incorrectly from stationary data (which can sometimes happen if the stationary ARR column gets corrupted, rarely happens) then there will be an abundance of detections in the 12:00-13:00 range.

Summarized Marker Tag data

* Raw counts of marker tag detections across selected sites for the duration of the study.
* Helpful for seeing which tags have been deployed at which sites and which sites may have had more downtime periods than others



### Release and Recap Lengths/Weights

* Plot to help ensure that there were no length/weight typos in the release and recapture files.



### Unknown Tags

* List of Tags without release info but started with 900\_ initially and have detections on some sort of antenna
* Display of Enc\_hist\_wide\_summary\_function output “Unknown Tags”



### Ghost Tag Movements

A df made from the “Ghost Tags” csv input. This shows registered ghost tags that have moved >0m after their assigned ghost date. This is helpful in revising the ghost tags file if one erroneously turns out to be a fish.

* Table is automatically sorted by “total\_distmovedAfterGhostDate” so the most problematic-looking fish are first displayed
* antennasDetectedAfterGhostDate is a vector of all antennas that have detected the tag after the ghost date. If the tag was identified as a ghost tag somewhere near upstream of a stationary antenna, it isn’t uncommon to see the tag be detected on that antenna during a high flow event
  + To get more context, you can check USGS discharge in the “Pressure Transducer, USGS, and Detection Distance” tab, find the discharge associated with that specific detection within 13 hours in “Encounter Histories – All Encounter Histories, or find the daily discharge associated with each daily movement in “Daily Movements” tab – Map and Table – “USGSDischargeDaily” column
* If a fish is moving upstream significantly after its ghost date, it might be worth investigating. This is seen by sorting/clicking the “maxUpstreamDistMovedAfterGhost” column.
* Movements of 10-50m ish between mobile runs can typically be attributed to GPS variation



### Crosstalk QAQC

Marker Tag crosstalk, where the tag goes off on one antenna but also registers on the other antenna as well, has been occurring frequently on CD since its installation. This tab is meant to monitor crosstalk instances of *Fish*, where a non-marker tag is registered at both antennas of a site at the exact same time.

Is created from combinedData\_df\_list$All\_Events, originally stemming from All\_combined\_events\_function().

* Summary Table shows the percentage of fish detections across the selected timeframe have been recorded at the exact same time
* Individual site tabs show when those simultaneous detections occurred
* There are instances where the simultaneous detections occurs naturally. In the app, milliseconds are not used, but the data coming off the readers do use milliseconds. Sometimes there are instances where a fish is recorded milliseconds apart on different antennas, but these will show up as the exact same timestamp in the app. However, when CD is doing it far more frequently than the other sites, that probably means something more is going on.



### Detection Distance/Water Level

Shows a color coded table of SiteVisitAndPTData, a joined df of PTdata and Site Visit Data by time to the nearest 13 hours.

* Using the variable “Water\_Level\_NoIce\_ft”, this table colors red the 32mm readings that are *below* the observed water level and green *above or at* the observed water level
* Helpful for identifying periods where detections may have been missed because of high flows.
* Readings that were entered as “TOUCHING” are changed here to .001



## Adding/Removing Dummy rows in the data

Dummy rows were added to the stationary, biomark, and release data using the dummy\_rows.R script to ensure the framework for new antennas was in place.



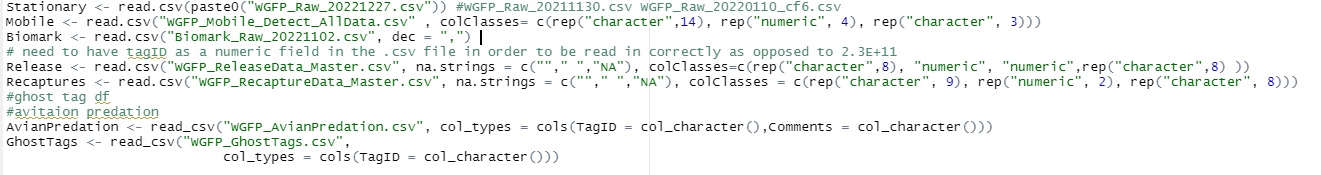
The rows remain while the function is ran to set the framework then the rows are removed in the data so this dummy data isn’t used. The rows are removed….

* In the get\_movements\_function
* 
* In the reactive for all events in app.r: 
* In get\_states\_function: 
* After the functions are ran in the following individual datasets in app.r: 
* In Ind\_tag\_enc\_hist\_wide\_summary function: 

To put the dummy rows back in, comment out the code that filters out the dummy rows above. Once there is actual data using cd7-cu12 antennas, you can delete these lines.

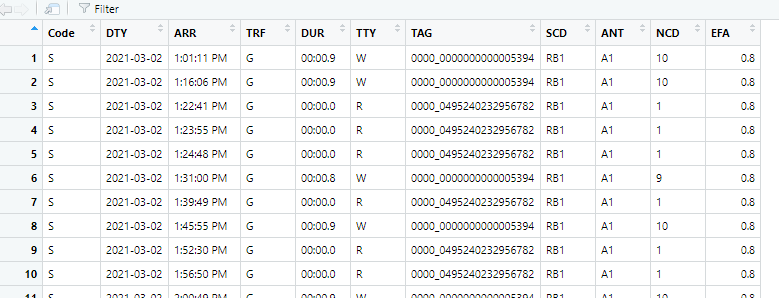
# CSV ReadIns

The app takes the following .csv files to run. If the column names of any important columns change in the csv’s, it will require a bit of easy tweaking to the code.

When a new version of these files comes out, simply change the first argument of the read\_ins to the new file name. For example, WGFP\_Raw\_20220203.csv would become WGFP\_Raw\_20220303.csv.Put the old file in an archive folder.

**Column names and order matter**. The read-ins and code could be tweaked so that at least column order won’t matter (use read\_csv and specify column types instead of read.csv), but names will still matter.

Column names outlined below are how they appear when brought into R at the beginning of the app using the code above. In Excel, names will appear slightly differently.

* Stationary
  + Combined file of all stationary detections. Obtained from the “Combine data RShiny app” found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\CodingDetections\WGFP\_CombiningData\_RShinyApp.
  + Most recent file is found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\All\_Stationary
  + This is a sample of what the data should look like. The most important columns are DTY, ARR, TAG, and SCD. The ARR column can sometimes get corrupted and brought in incorrectly for whatever reason. This is easiest seen when using the CombiningData\_RShinyApp and outlined in that How-To
* Mobile
  + A combined file of all mobile detections found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\MobileRaftAntenna\Mobile\_Detections
  + Usually Eric R prepares this
  + Column names and order: "Num”, "River" , "MobileSite", "Date", "Time" , "T109\_C”, "UTM\_X", "UTM\_Y" ,"TagType" , "TagID", "Event", "Ant", "Pass", "Species" , "Length" , "Weight" , "TagSize" , "RS\_Num", "ReleaseSite", “Survey", "Notes"
  + Of these, TagID, Date, Time, UTM\_X, UTM\_Y, and Ant are the most important columns
* Biomark
  + Combined file of all Biomark detections, found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Detections\Biomark
  + Made from the “Combine New Data RShiny App”
  + Column names and order: "Scan.Date" , "Scan.Time" , "Download.Date", "Download.Time", "Reader.ID" , "Antenna.ID", "HEX.Tag.ID" , "DEC.Tag.ID", "Temperature.C", ”Signal.mV" , "Is.Duplicate", "Latitude", "Longitude", "File.Name"
  + Most important columns are Reader.ID, Scan.Date, Scan.Time, DEC.Tag.ID.
* Release
  + Master Release file of all tagged fish. Found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Tagging
  + Column names and order: "RS\_Num", "River", "ReleaseSite", "Date" , "Time" , "UTM\_X", "UTM\_Y", "Species", "Length" , "Weight", "TagType", "TagID", “QAQC", "TagSize", "Ant" , “Event", "FinClip", "Mortality", "Comments"
* Recaptures
  + File of all fish that were recaptured found at U:\Projects\Colorado\_River\Windy\_Gap\_FishMovementStudy\Data\RFID\Recaptures
  + Column names and order: "Num" , "RS\_Num", "River" , "RecaptureSite" ,"Date", "Time", "UTM\_X" , "UTM\_Y", "Species" , "Length", "Weight", "TagType", "TagID", "QAQC" , "TagSize" , "Ant", "Event", "FinClip", "Mortality" , "Comments"
* Avian Predation
  + csv of tags succumbed to avian predation with a date of predation
  + used in the “States” function to assign a ghost or predated state
* Ghost tags
  + csv of ghost tags with date of ghost tag
  + used in the “States” function to assign a ghost or predated state
  + Column names: RS\_Num, River, ReleaseSite, ReleaseDate, Species, Length, Weight, TagID, TagSize, Event, GhostDate, UTM\_X, UTM\_Y, Comments

This section of the app adds dummy rows with fake data from CD, CU, B5, and B6. It was added to make sure the functionality for these new antennas was working, and is not currently in use.



# Functions

Here are some of the main functions of the app.

## All\_Combined\_events\_function

* + Main function that combines and cleans detections and release files
  + Takes Stationary, Mobile, Biomark, Release, and Recapture files as arguments
  + Returns:
    - WGFP Clean: a clean dataset of all stationary detections, no weird tags or marker tags, or duplicate rows. All timestamps and dates are in the same format. 900\_ is taken off of the tags
    - All\_Events: cleaned dataset of all 5 input datasets, containing every detection/event from each tag that hit an antenna, complete with release info.
    - Marker\_Tag\_data: file of just marker tags, to be used in QAQC
    - Recaps\_detections: a file of all antenna detections and recaptures, but no release data. This is just to be used in enc\_hist\_wide\_summary\_function

About the Code

The app first splits Stationary dataset into a dataset with no marker tags, and one with only marker tags. It also removes known test tags



The Marker tags are cleaned, first putting them into a mdy date format.

Then it goes through a timestamp cleaning process that is necessary because there are timestamps that have AM/PM, and others that do not. So we need to get them to military time.



Then the rest of the stationary data without marker tags or test tags is cleaned, including changing antenna names for the upcoming multiplexor, formatting dates, assigning UTMs for each antenna, and removing duplicate rows if there are any



Biomark dataset is then cleaned, including changing the reader ID since we want it in B3-B6 format, changing date format, filtering out test tags and marker tags, assigning UTMs, and removing duplicate rows



Then Mobile, Stationary, and Biomark columns are all renamed and then they are joined together. Mobile data requires a little Tag cleaning but not much.



Scan time is now cleaned like it was for marker tags, but with all the data



Data is filtered to only include detections past the study start date, and datetime is put in a good format



Release and recapture files are brought in and timestamps are cleaned. There is need for this because sometimes the way that times are entered in the spreadsheet are not uniform. Columns are renamed for joining.



Reformatting the all detections file to also be ready to join, then binding release and recap data to the df. Binding the release df to the main df makes it so there will be an event for “release”, and then left joining will then give release info for each fish for each detection.



Then there’s some minor formatting/renaming for display purposes in the app, and this is the final df that is used in the All Encounters Tab!



This line makes sure that the tags displayed are from the release file. There is a tab in the QAQC tab to see unknown tags histories.



This data frame condenses detections down to their daily summary; so it will take the first detection a fish had that day, and the last detection a fish had that day, and all unique detections in between. This reduces a fish to the “most relevant” detections. This df is ultimately used in the movements df, but first in joining with the Stations Data, which is the next function. WARNING: It’s important to note that if a fish has a day where the movement sequence is like “RB1, RB2, HP4, HP3, HP4, RB2, HP3, HP4”, then for that day, the sequence will register as “RB1, RB2, HP4, HP3, HP4”.



## Spatial\_join\_function

* + Joins detections and events to the shapefile of stations in order to later help calculate states and distance moved for each fish.
  + Takes condensed events (arg 1), made from all\_combined\_events\_function (all\_events\_most\_relevant). Also takes simple\_stations (arg2), which is usually read in as simple\_stations2 in the polygon\_read\_ins.r file
  + Returns a dataframe of condensed\_events with all station data attached.

About the Code

This function first converts the utm’s to numeric, assigns a projection/zone/datum, converts to lat/long, then to a spatial points object, then to sf object, then joins with the stations shapefile, also converted to an sf object. The joining is by the nearest feature, so if a detection is not quite on top of the station, it will get assigned the closest one.



## PrepareforStatesMovementsandSummary\_function.R

* + Puts stations onto a condensed All\_events dataframe
  + Takes StationData returned from spatial\_join function.
  + Returns:
    - All\_events\_days1: a condensed dataframe of all pertinent daily detection info for each tag with stations attached.
      * Row entries for tags with multiple detections on the same UTM\_X, UTM\_Y, at the same antenna, on the same day are filtered out. Leaving the first and last detections of the day, and all detections on unique antennas and UTM’s in between.

About the Code

The columns are renamed and the River is assigned for specific antennas



This part accounts for stations up the Fraser River, because those stations start at 0. This adds the station where the Colorado/Fraser confluence to the Fraser stationing. Then duplicate rows are taken out (shouldn’t be any) and relevant columns are selected



This part gets the number of daily unique events and detections for a fish, lets you know if a detection was above and below the dam. The type of movement is condensed to a “detection type” field. This is helpful in the movements map where it doesn’t necessarily matter which antenna specifically is hit, only if the stationing is different between events.



Ind\_tag\_enc\_hist\_wide\_summary\_function

* + Makes a summary dataframe of all released/tagged fish with over 60 columns of summary info, like “total number of antenna encounters”, “total distance travelled”, “moved through dam or not”
  + Takes recaps\_and\_all\_detections from All\_Combined\_events\_function, Release data, combined\_events\_stationsfrom PrepareforStatesMovementsandSummary\_function, and States\_summarized from the Get\_states\_function
  + It doesn’t take All Events from All\_combined\_Events\_function because we want to do some operations on the events that aren’t Releases, before bringing the release data back in for joining.
  + Returns:
    - ENC\_Release\_wide\_summary : Summary “wide” dataframe of each tagged fish and summary info
    - Unknown\_Tags: dataframe of tags that have encounters, but have no release info

About the Code

This part counts the number of Events for each fish. It then pivots the data to wide format, so each fish/tag has a row. Columns are then renamed. This is the start of the summary file.



The release data is then joined to the summary file. This is also when the unknown tags df is made (displayed later in qaqc). These tags all start with 900. They are probably cormorants or mergansers that have swallowed other PIT tagged trout and chubs from glenwood ;)



If there wasn’t a count for a fish, that NA gets changed to 0. True/falsecolumns are made on whether a fish hit a specific antenna or was recaptured.



Here some summary statistics based on the previous columns are calculated. It tells the number of detections a fish may have at a paired antenna. Also a T/F column to see if a fish was detected at a paired antenna. The fish with unknown release data are also filtered out here.



This part brings the condensed df with stations back in for summaries. It first counts the number of encounters each fish had above and below the dam. Then describes that event (eg “Mobile Detection above the dam”). Then if pivots the data wider to get each fish its own row and get in a format for joining back to the original release summary file. It also changes the NA’s to FALSE, because when you use count(), you only sum fish that have had encounters.



The data is joined with the original summary file. A column is made based on those new joined columns describing if a fish went through the dam or not.



The states\_summarized dataframe is then brought in and added. This has data about how a fish got through the dam, if it used the connectivity channel or not.



The distance a fished moved is then calculated from the condensed df detections (see the All Events Combined description to see how this is formed). Here, some data might be lost since it is using the condensed “most relevant” daily rather than every single detection, but it’s pretty rare that that will matter anyway, because most fish that that warning/edge case would apply to have actually been predated.

The total movement is calculated for each fish in the sequence of its encounter history. This is then joined to the summary dataframe as well.

Dummy rows are also taken out



## Get\_movements\_function

* + Makes a condensed dataframe of fish “movements”; see “movements defined” for more info
    - A movement is defined as a change in stationing (assigned to detections in the spatial\_join function). A fish that continuously hits RB1 and RB2 for example would register as a “No Movement”, but a fish that goes from RB2 to HP3 would register as an “Upstream Movement” on HP3. All data is incorporated in this calculation, so if a fish was released below red barn, hit the red barn stationary antenna, then was detected by a mobile run upstream of red barn, then was recaptured downstream of that mobile detection, it will register as a “Initial release”, “Upstream Movement”, “Upstream Movement”, “Downstream Movement”. The absolute total of this distance is also summed and displayed as a column in the Encounter Histories Summaries Wide tab.
  + Takes all\_events\_days1 from PrepareforStatesMovementsandSummary\_function
  + Returns:
    - Movement\_table\_no\_trans: dataframe where only daily movements are included per tagged fish on unique antennas and UTM’s. See “Movements defined” for about them

About the Code

The code first groups by each tag and arranges each tag in order by their first, then subsequent events. It then calculates the distance moved between each event for each fish, based on the stationing. The code accounts for Fraser/Colorado River transitions by first calculating the distance to the confluence from the last event, then finding the distance from the confluence to the present event, then adding those together. Otherwise, it’s just the difference between the last event and the current one.

The sum distance is then calculated as an absolute value of each movement that was calculated for a fish.

Marker Colors and icon colors are then assigned for later mapping.

Lastly in this code chunk, X and Y are defined as columns in preparation for changing the utms to lat/longs. Note: there are a couple places in the app where UTMs are converted to lat/longs, and the code would be a little more elegant if these were just all made to sf objects, but this stuff also all works fine.



The UTM’s are fully converted to Lat/longs, then the df is parred down with the distinct() function like it has before, but this time, it takes “movement” into account instead of “first/last”. Movement\_table\_notransitions1 is what is mapped and filtered in the app.



## Get\_States\_function

* + Makes dataframe of weekly “states” either above or below the dam. Below the dam is state A and above is B. State C is a detection in the connectivity channel. Ghost state is G, and Predated state is P
    - Any detection, recapture, or release is registered as a state
    - A fish is inferred to remain in that state until it becomes a ghost tag or is detected/recaptured again.
  + Takes combined\_events\_stations from PrepareforStatesMovementsandSummary\_function, the Ghost tags df, and the aviation predation df.
  + Returns:
    - All\_States: Dataframe of daily States of fish. Certain assumptions are made to infer transitions, such as if a fish hits 1/2 antennas at a site then hits an upstream antenna, it is said that an upstream transition originally occurred and it just missed 1 antenna. For more assumptions see thee Daily States Tab section
    - Days\_and\_states\_wide: same data as all\_states but in wide format where days are the columns, TAG is the first column, and values are states
    - Flagged Movements: States that couldn’t be accounted for in the code

About the Code

The 3 dataframes are joined then the df is cut down to the most succinct: rows with a distinct combo of Tag, Event, Datetime, UTM’s, and first and last event of day.



The states are then assigned based on above/below the dam, ghost tag, and predation date



The states are then kinda “squished” together into one row on a weekly basis, and duplicated letters in a row are removed, because a fish doesn’t need a weekly states history that says “AAAA”, it just needs “A”. A column is also calculated to get the amount of weekly unique events.

The df then gets rid of duplicated rows, since each fish has multiple rows of states in the same week at this point. Relevant columns are selected and renamed

Some summary information for the fish are then calculated based off these states. In the states\_summarized df, the states are “squished” together for all states the fish has. From there, you can see if a fish went above/below the dam, and used the channel, which is what the new columns check. These columns are eventually joined back into the encounter histories summaries wide dataframe.



This is an idea to put all the wonky states in place so it would be easy to see, but it isn’t fully compatible with the app yet and therefore isn’t used. My idea is that instead they would be in one tab on the states page so less one manual parsing is needed.



# Shapefile/Polygon Readins

The file “map\_polygon\_readins.R” reads in all shapefiles for the movements map. Also contains some graphics options for release sites, stationary antenna sites, and labels for stations.

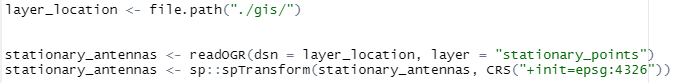


# Modifying, Updating, Adding New Antenna/Stations

Need to update this to new function: prepare for states, movements, summaries

## Adding to the map

* + Head to the GIS database and export desired layer as shapefile, saving it in “gis” folder within the app directory
    - Right click on layer in left column, scroll down to data -> export
  + Using the rgdal package, read in the .shp file,
    - If you are adding an antenna or making new stations, you can change the “layer name” argument to the name of the new .shp or .rds file file, and it will automatically get added to the map for that layer.



* + The espg:4326 converts the projection to something usable
  + If the layer is very large, it might be good to convert the .shp file to a .rds file, which decreases resolution. For example, the stations file is a .rds file and was converted below. It keeps 10% of the original resolution



* + It is read in like so and is much faster than when read in as a shapefile



* + If you make a new layer that you want to put on the map, you’ll have to add it to the map.



* + Add\_polylines is used to bring in lines like stream\_centerline, addMarkers is used for SpatialPoints, addPolygons is used for polygons.
  + Add a Group argument to the data and add it in the addLayersControl function to make sure it shows up

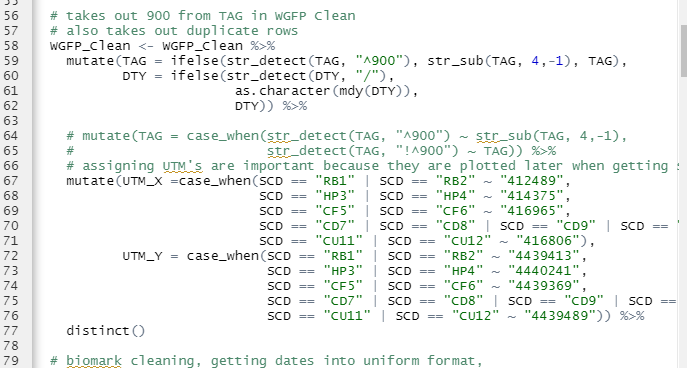
## Incorporating the new antenna with the data

First, go to the dummy\_rows.r function and add rows with the desired antenna (see adding/removing dummy rows). Then you need to modify the following functions.

### In the All\_Combined\_Events \_function

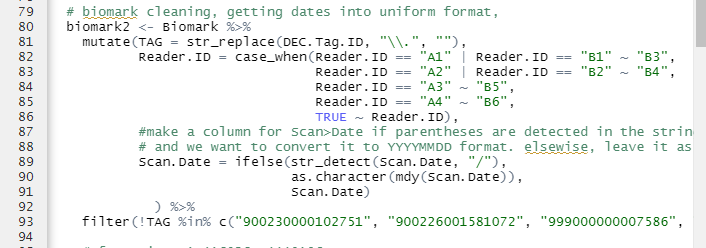
Antennas are assigned specific UTM’s during this section. These are specifically used later when each detection is assigned to a station in the “spatial join” function.

For stationary antennas, you need to assign UTM’s based off the site code field

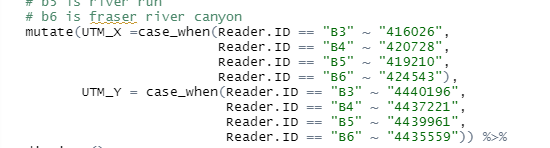


For biomark antennas, you need to add in the Reader.ID field, based off what the antenna is called in the field. Since Windy gap biomark has been called A1 or B1 at times in the past, so this changes all those entries to B3.

That last line also removes marker tags. There’s another line later that also removes all biomark tags that start with anything other than 900, which would include marker tags.

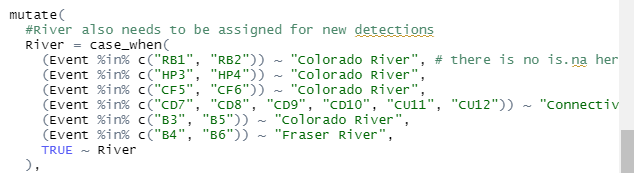


This is the section that assigns UTM’s to biomark antennas, after making the correct reader.id

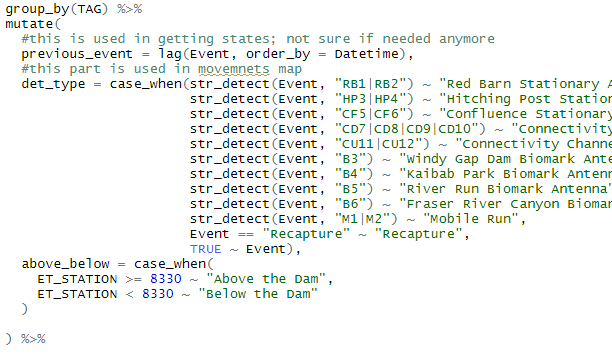


### PrepareforStatesMovementsandSummary\_function

The River is assigned to the new antennas in this part.

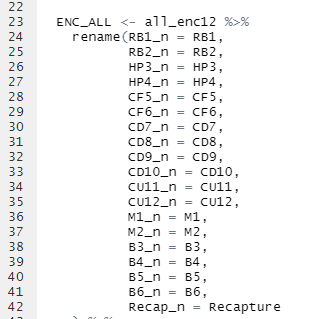


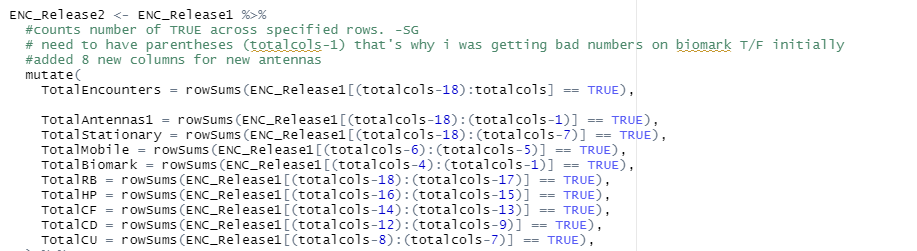
The type of movement is condensed to a “detection type” field. This is helpful in the movements map where it doesn’t necessarily matter which antenna specifically is hit.



### In the Ind\_tag\_hist\_summary\_wide\_function:

You need to assign column names for the new antennas quite a bit in this function, starting with this bit of code. It should be pretty intuitive what to add/where to add if you examine the code.



This bit is the tricky part: Where sums of rows are totaled and it’s based off the total number of columns in the dataframe. Add the desired columns here, but be sure to check your work to make sure you’re counting the columns right and getting the results you expect.

# Variables.R

There is a variables.R folder that is meant to put important info in one place. This only has a couple variables implemented right now, but it would be good to put the UTM’s for assigning in one place as well.

# Shiny Modules

Each tab is split up into its own module. This makes the code reusable (can plug the ui and server into any app easily with the right arguments) and organized. This is where the data processing occurs from the sidebar filters and how graphs and tables are rendered. Read more about Shiny Modules [here](https://mastering-shiny.org/scaling-modules.html).

# General Notes

* All plots are interactive. Hover over plot for more info and toggle what is displayed by clicking and double clicking items in the legend. Zoom in by clicking and dragging within the plot.
* Datatables are also interactive. Click the column titles of any column to sort in ascending/descending order. You can use the Search bar on the right to search/filter for any specific values. Some of the individual filters work, some don’t, depending on column type I believe.
* It’s important when adding a new updated file to the app that it had the exact same column names and order as the older file
* UTM’s for stationary and biomark stations are assigned within the app
* Keep a folder of CSV backups in case the time column goes wonky
* Currently there is a Release site called “Shefiff Ranch Middle Field”; if this typo is corrected in Release File, it will have to be changed in the get\_states\_function code as well. Otherwise all fish will show up in the “unknown states” tab.
* This information is in the Combining Data How To as well, but 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 Kaibab Park, 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*.
* The above error message has also shown up for me when loading in data where the full TAG ID has been deleted/cut off. Double check that the “DEC TAG ID” column has actual full values, if not, they were lost somewhere along the way and you’ll likely have to re-combine files. Still unsure why it sometimes cuts them off, but I’ve found that changing the column format to “Number” with 12 decimal places before making edits fixes the problem.

Let me know if there are other questions and concerns, I’m happy to add to this if needed. Let me know [sfigraf@gmail.com](mailto:sfigraf@gmail.com)

# Very Doable tasks for the future

* Expand on the “variables.R” folder. My idea is to put important data in this about gps coordinates of antennas, antenna names, and other things that come up. Then we don’t have to go on a goose chase hunting down the variables in the code when we just change it in the variables
* Use the package “minicharts” to make fun bar graph animations on the movements map. This would be a good way to visualize the quantity of detections over time.
* Streamline the data cleaning process so that the data, especially Stationary data, is clean when it is read in to the app. This would help with the time it takes to load the app, since it would make most of the All combined events function obsolete. My idea is to take out the pieces of that function and clean each individual dataset, save as a csv’s, then modify it so the AllCombinedEvents function will only basically combine all the DF’s.
* Have the data be read in from a database instead of CSV’s. Maybe a good step would actually be to use .rds files instead of csv’s at the least. Sometimes the time columns of biomark and stationary get corrupted and I have no idea why, and I can’t help but think this wouldn’t happen if they were being read in from a DB instead of CSV’s.
* Convert and use spatial data as sf objects, not shapefiles. Data wrangling is a bit more intuitive and the spatial join function could be cleaned up because you can more easily convert from utm’s with st\_transform
* Change all the plots to the rainbow trout color pallete
* Make all “upstream/downstream movements” colors on the animation map the same as the ones on the movements map