**OBSERVER DATA PREPERATION**

**WEIGHT DATA**

**Creating master sheet:**

1. Unmerge all cells.
2. Insert missing data on species category and species names. Where species is unidentified mark this with the name “un\_id”. Where category names are missing or misplaced insert or shift to correct place.
3. Create a new sheet in Excel
4. Copy all headings from old sheets first to row one of the master sheet which contains all drags i.e. the meta data and the catch data.
5. Create new variables:
   1. Trip number (trip\_no)
   2. Drag number (drag\_no)
   3. Sample weight (sample\_weight)
   4. Total weight (total\_weight)
6. Copy data over for every drag under the respective variables.

**Tidying master sheets with all drags for the respective trips:**

1. Give master sheet different colour and name
2. Ensure all cells contains values only and no formula
3. Create a new variable called “groups” which will contain the following values:
   1. Discarded Bycatch (discarded\_bycatch)
   2. Retained Bycatch (retained\_bycatch)
   3. Target Catch (target\_catch)
4. Insert three added rows under each drag for retained catch and bycatch for every drag. Copy and paste this data from the individual sheets.

**Merging of drags master sheets in trips master sheet and more detailed cleaning:**

1. Double check raised total and sample weights (summative values) with the variables.
2. Create a new workbook and copy over the master sheets. In total 8 trips (sheets).
3. Check for consistency with variables i.e. each column is name similarly before deleting extra copied columns/variables. Freeze first row and scroll down to check b4 deleteing. These were copied over to avoid copy-paste errors which could jepordise the analysis. When finished only the first row should contain names of variables and all others should contain data for the trips.
4. Simplify variable names (coding data) using three-digit codes and create a table for records. Simplify some named values in like manner and create and save a similar table. Create a meta data sheet.
5. Format final master sheet. Formatting was done here to avoid repeating 8x. 😊 Formatting include looping through/over every row and making the following changes (turn on filter to get quick view of variable values):
   1. Create a new drag variable which sequentially counts all 48 drags together despite trips i.e. 1,2,3,4,5,6,7,8,9,10.... This was done by creating a temp dataset with each trip and drag. A merger row of the two was created in the temp dataset using the CONCATANTE function. This step was repeated in the master sheet. Both columns were copied as value only to remove formulas. A new variable was created in the master sheet named “drag\_all”, and in the first cell of the column the VLOOKUP function was used to find the corresponding trip-drag number in the temp data. The first 15 rows were checked to ensure that the formula was working. When confirmed the formula was copied down and then the column formatted as values only. .
   2. All dates were formatted to short dates.
   3. Abbreviations were used for long observations. Those with multiple words were joined by “\_” to remove spaces.
   4. Time variables were formatted to time. New variable was created “time\_fished”. This was done by subtracting time drag start from time drag ends and which returned the elapsed time i.e. in AM time. This was changed by right clicking and select Format Cells and change it to h:mm. Trip 2 drag 6 time drag end was corrected from AM to PM. Trip\_8 times seem erroneous. Should be reviewed.-Fixed
   5. Different spelling for the same observations were corrected e.g. data entered by D. Spellen, D-Husband, spellen…or Bonny fish, Boney Fish, Bony Fish
   6. Scientific and common name mismatches were fixed. The scientific names were assumed to be correct. Closer attention to species photo ID sheet will reduce these errors in future.
   7. ~150 cases of cases of species were unidentified. Some were already named in the earlier trips and some were common species 😊.
   8. Advanced filtering done using seabob (mentioned once per trip) was applied to create a new sheet with no repeated rows. This sheet was used to check for accuracy of e.g. dates
   9. Remove total weights and sample weights column as they were calculated.
   10. Check to match categories to species to make sure they are accurate
   11. Reorganise variables logically.
   12. Double check days at sea.
   13. Check drag date that it is equal to or more than dep date and equal to or less than arrival date.
   14. For species where the common name was unknown the first letter each of the Genus and Species names will be used, separated by “\_”.
   15. Fill all blank cells with intuitive names.
   16. Remove zero values
   17. Checked that all trip days, times, and dates are in sequence. E.g. trip 2 date must be after trip 1.
   18. 26 rows with duplicate scientific names were fixed. Common name was ticked unidentified but there was a scientifc name given?
   19. Get hauls done per trip from log sheets (google drive shared files).

**Questions**

1. How do you get estimate of seabob for sale and prawn for sale? There are one or two similar weights back to back, is that accurate? T
2. Is there underreporting of prawns. Figures seem small. What about the trip where you sent photo?
3. Comment…I think what we have been making a distinction all the time with sorted and unsorted weight can be simply referred to as total\_discards and sample\_weights. Lol
4. Do you take pictures of every sampling haul? If so we may be able to retrieve co-ordinates from those pictures.

**LENGTH DATA**

1. Separate sheet was created.
2. The lengths were removed by trips and drags individually.
3. The “=” sign was first removed from cells.
4. The scientific name and length column were copied to a separate sheet
5. The left and right function was used to removed “(“ and “)/30)” from the around the lengths.
6. Next the text to column was used to remove the “+” sign and place each length in a separate column.
7. This data was then copy over to the main sheet where the data was stored. The transpose function was used to change data from wide to long format.

**HELP SOURCES:**

|  |  |
| --- | --- |
| **Description of issue** | **Source of help** |
| Calculating time | <https://www.youtube.com/watch?v=hiuY1PGL_I4&ab_channel=DannyRocks> |
| Selecting the first of repeated rows with Advanced filtering. | <https://www.youtube.com/watch?v=UV-SPoVkDaU&ab_channel=TutorialsPoint%28India%29Ltd>. |
| Change first letter only to capital | <https://www.extendoffice.com/documents/excel/2771-excel-capitalize-first-letter-of-cell.html> |
| Replace “\_” with “ “ | <https://excelribbon.tips.net/T012488_Using_a_Formula_to_Replace_Spaces_with_Dashes.html> |
|  |  |

**SHORTCUTS USED:**

1. Copy here as values only: Ctrl + Alt + VV (Opens Paste Special Dialogue Box)
2. Paste special transpose: Ctrl + Alt + VE
3. Text to columns: Alt + AE
4. Fill blank with value above: Ctrl + G then Alt + S then K (Opens GoToSpecial dialogue box) then =above then Ctrl + Enter then #2
5. Move to top of data or bottom, left or right: Ctrl + direction

**TIPS**

1. Upload CSV file in Google Drive, open with google sheets and download CSV to get rid of “?” in r.

**SCALED UP FISHERY DATA**

1. The total trip hauls were obtained from vessel logs.
2. Data on target catch, prawn bycatch and discarded bycatch were used from observer hauls and raised to trip level. Data on retained finfish were used from the fisheries department reports.
3. To compensate for missing catch data the following approach was taken:
   1. Missing total hauls: It was assumed that the vessels sailed to the fishing ground on day 1 and on the last day of the voyage it sailed to port. The assumption is therefore that there is zero fishing on those days. It is known that the average haul time is 3-4 hrs per haul. Therefore, the assumption is that an average 6 hauls per day was used to calculate the total hauls per trip where this data was absent. Therefore, if the vessels spent 10 days at sea the calculation of total hauls made will be raise the total hauls made was (10days-2days)\*5hauls/day = 40 hauls.
   2. There was one instance where the total finfish landed (trip 7) was absent. The eight trips was divided equally amongst nhs and psi (4 each). It was a trip for nhs that the data was missing for. Therefore, the average for the three previous trips was used.

**CHANGES MADE TO CSV FILES DURING ANALYSIS**

File names were changed as follows (meaning – “filename”):

* 1. Observer data scaled up to level of the fishery – “observer\_fishery”
  2. Length measurements by species – “observer\_lengths”
  3. Observer data scaled up to level of the trips - “observer\_trips”
  4. Weight measurements by species - “observer\_weights” (Main file)

**File A**

**File B**

**File C**

**File D**

1. Two duplicate rows were removed.
2. The new variables were added. Namely (meaning – “variable\_name”)
   1. Species identification status i.e. yes or no– “id\_status”
   2. Species commercial status i.e. yes or no – “com\_status”. Where species were unidentified the com\_status was assumed to be “no”.
   3. Species order – “order”
   4. Species family – “family”
   5. Species class – “class”
   6. Fao 3-alpha code – “alpha\_code”
   7. Animal group i.e. Vertebrate (yes) or invertebrate (no) – “a\_group”

Where it was not possible to group and unidentified species it was marked by “uk” meaning unknown.

1. A lookup table using species idendentified in unu report was used to find order and family with vlookup. Other sources of information were:
   1. Tomas 2018 report Guyana
   2. WoRMS, fishbase and sealife species databases. Tip: when unable to find species using scientific name, google name and google will suggest correct name.
2. Cases where species were grouped and not identified individually the classification columns were marked “not\_applicable”. Where the species were otherwise just unidentified i.e. “un\_id” or on the species name was known and not the Genus, these were marked as “unknown”.
3. Some scientific names were misspelt or have since been deemed unaccepted. These were corrected and updated, respectively.