**README\_LW**

**October 7, 2019**

The LW\_Manuscript\_Oct2019 folder contains materials related to the UWL-REU project on Large Wood, specifically on the distribution of large wood in space and time. Files are organized into two main folders, described as follows:

LW\_Manuscript\_Oct2019/Manuscript

* CWD\_distribution\_paper\_draft1\_KJJ\_Oct2019.docx
  + the current working version of the manuscript
* CWD\_distribution\_paper\_SupplementaryMaterial.docx
  + the current working version of the manuscript’s supplementary material (appendices)

LW\_Manuscript\_Oct2019/Analysis

* /data/
  + all\_reduced\_clean.Rda
    - cleaned data from pools 4, 8 and 13; used as input to descriptive\_stats\_clean\_updated\_Oct2019.R
  + all2\_reduced\_clean.Rda
    - cleaned data from pools 26, la grange, and open river; used as input to descriptive\_stats\_clean\_updated\_Oct2019.R
  + data1.zip
    - a zipped folder containing raw data, ancillary data from GIS analyses, and misc data related to the generation of the two allX\_reduced\_clean.Rda files listed previously. It might be a good resource for revising the datacleaning steps.
* /R\_scripts/
  + cwd.datacleaning.updated\_deprecated.R [Deprecated as of 11/26/19, but we’ll keep it as a reference]
    - a script to pre-process raw data for analysis. This is probably the best starting point for reviewing the workflow.
  + cwd.datacleaning.updated.Rmd
    - R Markdown version of the above, created by Kaija, Oct/Nov 2019, with code cleaned up and annotated.
  + descriptive\_stats\_clean\_updated\_Oct2019.R
    - current working script that generates results to populate tables, create figures, and produce results contained in the manuscript and supplementary materials. This uses a cleaned datafile as input to produce results for the manuscript.
  + libraries.R
    - list of packages to be called in analysis. See [this blog entry](https://thewoodpeckr.wordpress.com/2019/10/12/loading-packages-efficiently/) for information about how this works.
    - referenced in:
      * descriptive\_stats\_clean\_updated\_Oct2019.R
      * cwd.datacleaning.updated.Rmd
  + ownfunctions.R
    - referenced in cwd.datacleaning.updated.R
  + data\_cleaning\_misc
    - a folder of misc data cleaning scripts. They are included in case you need to backtrack further.
  + Environmental\_analysis
    - A folder of misc scripts related to developing predictive models of wood occurrence based on environmental attributes.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Some further notes:

1. **Identifying pseudo-shoreline observations**

Two fish database files are referenced in cwd.datacleaning.updated.R files, “ltrm\_fish\_data\_new.Rda” (called in line 176) and “fish\_data\_EF.Rda” (called in line 23). Both have a column name of “summary” which conveys information about the quality of the sample (see online fish dataset documentation). The value of 7 indicates “that pseudo-shoreline used for shoreline gear” -- these are the observations that should be excluded from the analysis. The fish database files are included in the data1.zip folder (and one may be buried in a second zipped folder).

Where/how exactly the pseudo-shoreline removal is done is up to you, of course, depending on how you intend to set up the workflow. A quick and dirty way is to add a line in the data processing code (descriptive\_stats\_clean\_updated\_Oct2019.R) to generate results quickly. I think I made a comment in that script about doing it myself… In the long term the exclusion may be better handled in a revised data cleaning script, such as a revised cwd.datacleaning.updated.R script.

[11/26/19: excluded pseudo-shoreline observations from the data from pool 4, 8, and 13 in the cwd.datacleaning.updated.Rmd script].

There is also the issue of cleaning data for the analysis of environmental variables, which could be difficult given that we are still discussing what variables are best. For the time being, you may not want to invest much time into editing the environmental variable cleaning bit. We can revisit that after we settle on a list of variables that are important to consider.

1. **Updated Data Analysis**

Results for the current Oct 2019 version of the manuscript were generated using the script

Analysis/R\_scripts/descriptive\_stats\_clean\_updated\_Oct2019.R

This script takes cleaned versions of the fish/large wood datasets and generates results to populate the tables and figures that are in the working manuscript draft and supplementary material (see documents in LW\_Manuscript\_Oct2019/Manuscript folder). It should be fairly straight forward to step through the analyses once the data files had been cleaned to update results.

When it comes time to make figures, consider looking at what these three journals suggest for figure formatting so that you can update the script to export figures efficiently.

* Geomorphology: <https://www.elsevier.com/journals/geomorphology/0169-555x/guide-for-authors>
* Freshwater Science: <https://www.journals.uchicago.edu/journals/fws/instruct>
* Physical Geography: <https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=tphy20>

Note that we aren’t settled on which journal we’re aiming for yet because we haven’t worked up the whole story yet. But all should have a set of similar minimum figure formatting requirements. Once we decide on a journal we can quickly make adjustments to meet any nuanced figure requirements (e.g., labeling panels with a, b, c, vs i, ii, ii, etc).

1. **Head’s up**

USGS is switching up service providers from Google to Microsoft, so our Google drives will be migrated --hopefully without disastrous consequences--to a different state some time this winter. We aren’t sure what will happen. This Google drive is actually my personal gmail account, so there will be some continuity there. But the migration does affect any large wood or REU material any of us USGS-ers have shared. Just something to be aware of!

This also means that when it comes time to sharing manuscript versions back and forth with each other we may need to have a different plan to do that. KathiJo and I can figure that out as we know more about our file sharing capabilities. Yay for change.