Readme WDFW crab logbook data 082820

The zip folder contains a project Corey Niles had going for some other things and to which I just added the 2018-2019 data. It’s organized around an RStudio project. people don’t have to use R or RStudio to get the data. Instead, there’s a CSV file in the “output” folder. And the original data in the “data” folder. The R notebooks just show Corey got from the original data to what’s in the output folder.

* The main directory: contains the RStudio .Rproj file and some Rmarkdown notebooks with the code I used to create the files in the “output” folder.
  + The “Read and prep crab logbook data\_2020-08-28.Rmd” notebook is what I use to stack and join the individual csv files. I used the data.table package to do that. It saves the csv compilation and does a bit of error checking and adding of some convenience fields (e.g. season start dates).
  + The “Create-and-save-GIS-sf\_and\_gdb.Rmd” creates GIS objects with the R sf package and saves those as RDS files and into an ArcGIS geodatabase.
  + The “Check-log-dates-against-WAFT.Rmd” records how I checked some dates against our fish ticket database. It’s there to remind me what I did. You could do something but would need to use PacFIN and the FTID column for our FishTicketNumber field.
* The “data” folder:
  + contains the input csv files, which were exported from the Access databases where we maintain the data.
  + We have a different Access databases for each season. And each database as two main tables, a Header table and a Sets table.
* The output folder:
  + A csv file with the word “compiled” in the file name. That file combines all Set and Header tables from all seasons.
    - I did minimal error checking and corrections and just on the date fields.
  + The .rds files are GIS objects created with R’s sf package and to be loaded into R with readRDS(). Two of those plot start and end coordinates as points and the other combines them into lines.
    - There are two files each for the points and lines, with one being in the WGS84 geographic coordinate system and the other in the UTM10N projected coordinate system.
    - Compared to the csv file, these .rds files drop records that are missing coordinates.
  + An ArcGIS file geodatabase with the same four spatial files, which I loaded from R with the arcgisbinding package.
    - In writing the spatial lines to the geodatabase, I got a “geometry is not valid” argument which prompted me to use *argument 'validate = TRUE' to enforce topological consistency.* It wrote the file despite the error. Yet I rewrote with the validate argument advice. I didn’t research what that meant.