VMS data processing pipeline

Jameal Samhouri

10/11/2016

## How to deal with VMS data from OLE

1. jameal begs for data
2. OLE sends data in wrong format (html), but eventually in right format (csv or txt)
3. blake compiles csv files from OLE, hand scrubs according to his "Blake's Methods Log" google doc and "VMS csv processing summary"
4. run 01\_process R code (bare\_bones() function) on .csv files for individual years to remove junk

* barebones() will:

1. Remove all garbage headers from each .txt file (this part of bare\_bones() is currently off)
2. Ensure each column is formatted properly (date-time, etc) c)Remove NAs
3. Spit out clean RDS and csv files to be used instead of raw .txt files
4. run 02\_process R code to remove duplicates, heisenbergs, complete cases, and subset raw VMS data to on-water points only
5. overlay NGDC coastal bathymtery grid on VMS data points

* for on-water points step:
* bathymetry data: NOAA NGDC West Coast Coastal Relief, 3-arcsecond (~90 m) resolution. Elevation in decimeters re. sea level.

1. use extract() function as extract(all\_vms, bathymetry.ascii) to assign elevation to each row of vms data.
2. subset new all\_vms to elevations <-1m and not embayments (following Blake's thematic map and atlas).
3. retain data points that are deeper than 2000m (NAs in this data set <=-115 W) so as not miss any fun tuna action.
4. slice up trips for each vessel by assigning all VMS points to 0 or 1 for at-sea or on-land, and using sequences of 2 consecutive on-land points to define the end of a trip. this part of the code will also remove abnormally high speeds and make compilation file for each vessel of all trips in the database
5. Emma will run 00\_make.R lines 1-20 to process the new fish ticket data, and then send the output to tigress to do the metier analysis (including 2014 metiers).
6. Jameal will run 00\_make.R lines 21-54 to sewTrips (including 2014) to create a new tickets.RDS
7. Jameal will run 03\_overlapMetier.R to drop all VMS tracks not fishtickets.RDS.
8. Jameal will run 04\_link\_mets\_vms.R to link VMS data to metiers.

* will comment out observer data crap
* will beware of any() statements that use numerical indices (eg any(duplicated(c2\_bydate[,3:8]))).

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.