Notes on assembling the dataset for Elise Collett, student at St.FX

Daniel Ricard

Last modified: 2021-10-07 15:14:20

Collaboration with St. FX

Species considered

To get things started, I am generating a dataset similar to that used by Benoit and Swain (2008). Table A1 shows 52 species. Let's try and reproduce that table by using the correct species codes in the RV database.

Table 1: The 52 species of interest, meant to match Table A1 in Benoit and Swain.

		family	LATIN	ENGLISH	aphia_id	species.code
Myxini						
Myxiniformes						
		Myxinidae	Myxine glutinosa	Atlantic hagfish	101170	241
Elasmobranchii						
Rajiformes						
		Rajidae	Amblyraja radiata	Thorny skate	105865	201
			Leucoraja ocellata	Winter skate	158553	204
			Malacoraja senta	Smooth skate	158554	202
Squaliformes						
		Squalidae	Squalus acanthias	Spiny dogfish	105923	220
		Etmopteridae	Centroscyllium fabricii	Black dogfish	105906	221
Actinopteri						
Aulopiformes						
		Paralepididae	Arctozenus risso	White barracudina	126352	712
Clupeiformes						
		Clupeidae	Alosa pseudoharengus	Alewife	158669	62
			Clupea harengus	Atlantic herring	126417	60
Gadiformes						

		Gadidae	Boreogadus saida	Arctic cod	126433	110
			Gadus morhua	Atlantic cod	126436	10
			Gadus macrocephalus	Greenland cod	254538	118
			Melanogrammus aeglefinus	Haddock	126437	11
			Pollachius virens	Pollock	126441	16
		Macrouridae	Nezumia bairdii	Marlin-spike grenadier	183289	410
		Merlucciidae	Merluccius bilinearis	Silver hake	158962	14
		Lotidae	Enchelyopus cimbrius	Fourbeard rockling	126450	114
		Phycidae	Phycis chesteri	Longfin hake	158988	112
			Urophycis tenuis	White hake	126504	12
Perciformes						
		Gasterosteidae	Gasterosteus aculeatus	Three-spined stickleback	126505	361
Osmeriformes						
		Osmeridae	Mallotus villosus	Capelin	126735	64
			Osmerus mordax	Rainbow smelt	126737	63
Perciformes						
		Ammodytidae	Ammodytes dubius	Northern sand lance	151520	610
		Anarhichadidae	Anarhichas lupus	Atlantic wolffish	126758	50
Eupercaria incertae sedis						
		Labridae	Tautogolabrus adspersus	Cunner	159785	122
Perciformes						
		Stichaeidae	Leptoclinus maculatus	Daubed shanny	127072	623
			Lumpenus lampretaeformis	Snakeblenny	154675	622
			Stichaeus punctatus punctatus	Arctic shanny	293745	624
		Zoarcidae	Gymnelus viridis	Fish doctor	127096	616
			Zoarcidae (f.)	Eelpouts unidentified	125575	598
			Melanostigma atlanticum	Atlantic soft pout	127120	646
			Zoarces americanus	Ocean pout	159267	640
Pleuronectiformes						

		Pleuronectidae	Glyptocephalus cynoglossus	Witch flounder	127136	41
			Hippoglossoides platessoides	American plaice	127137	40
			Hippoglossus hippoglossus	Atlantic halibut	127138	30
			Limanda ferruginea	Yellowtail flounder	158879	42
			Pseudopleuronectes americanus	Winter flounder	158885	43
			Reinhardtius hippoglossoides	Greenland halibut	127144	31
		Scophthalmidae	Scophthalmus aquosus	Windowpane flounder	158907	143
Perciformes						
		Agonidae	Aspidophoroides monopterygius, Aspidophoroides olrikii	Alligatorfishes	159458	340, 341
			Leptagonus decagonus	Atlantic poacher	127191	350
		Cottidae	Artediellus sp.	Hookear sculpins	126147	306, 880
			Icelus spatula	Spatulate sculpin	127200	314
			Myoxocephalus octodecemspinosus	Longhorn sculpin	159520	300
			Myoxocephalus scorpioides	Arctic sculpin	127202	316
			Myoxocephalus scorpius	Shorthorn sculpin	127203	301
			Triglops murrayi	Moustache sculpin	127205	304
		Cyclopteridae	Cyclopterus lumpus	Lumpfish	127214	501
		Liparidae	Liparis sp.	Seasnails	126160	503, 504, 505, 506, 508, 512, 513
Scorpaeniformes						
		Hemitripteridae	Hemitripterus americanus	Sea raven	159518	320
Perciformes						
		Psychrolutidae	Cottunculus microps	Polar sculpin	127235	307
		Sebastidae	Sebastes sp.	Redfish unidentified	126175	23

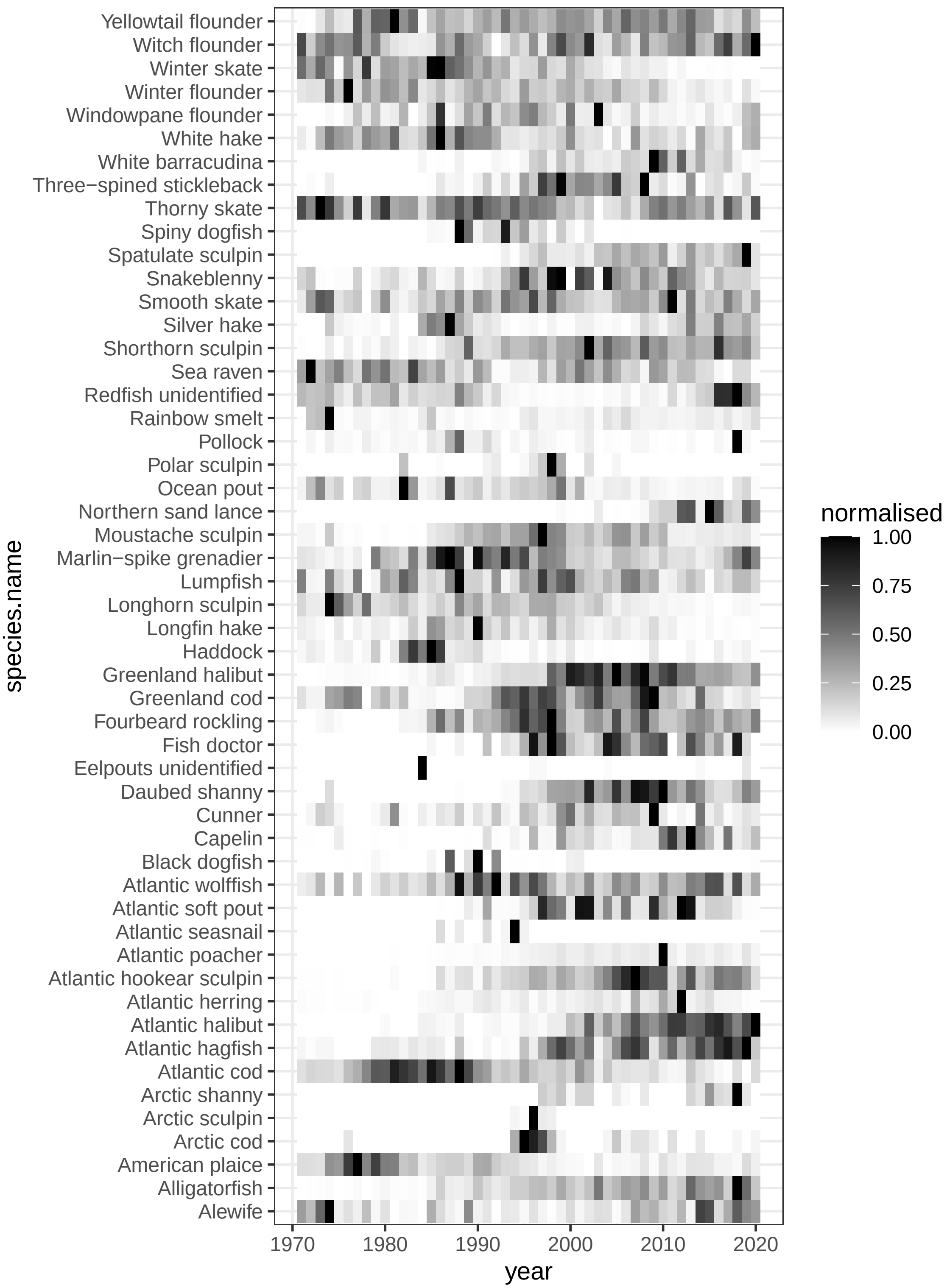
Dataset of abundance per tow

Now extract set cards and catch cards for representative sets in strata 415 to 439, and compute the yearly stratified random estimate of abundance per tow, corrected for gear, vessel and diurnal effects.

```
data(rv)

yrs <- 1971:2020
x <- rv.good.sets(yrs)
```

```
x <- x[x$stratum %in% c(415:439),]
x$unique.id <- paste(x$vessel.code, x$year, x$cruise.number, x$set.number, sep="-")

y <- rv$cat
y$unique.id <- paste(y$vessel.code, y$year, y$cruise.number, y$set.number, sep="-")
y<-adjust(y,x)

## output matrix, called C in Benoit & Swain
C.matrix <- matrix(NA, nr=length(sp.codes), nc=length(yrs), dimnames=list(species=sp.codes, year=yrs))
C.df <- expand.grid(year=yrs, species=sp.codes, normalised=NA)

for(i in 1:length(sp.codes)){
  s <- sp.codes[i]
  if(s==503) { # snailfishes
    ss <- c(503, 504, 505, 506, 508, 512, 513)
    this.y <- y[y$species %in% ss,]
    ## sum for each set
    agg.df <- aggregate(number.caught~unique.id, this.y, sum)
    vars <- c("unique.id", key(x))
    this.agg.y <- merge(x[,vars], agg.df, by="unique.id")
    this.agg.y$species<-503
    z <- merge.catch(x, rvcat(this.agg.y), var = "number.caught") #
  }
    if(s==340) { # alligatorfishes
    ss <- c(340,341)
    this.y <- y[y$species %in% ss,]
    ## sum for each set
    agg.df <- aggregate(number.caught~unique.id, this.y, sum)
    vars <- c("unique.id", key(x))
    this.agg.y <- merge(x[,vars], agg.df, by="unique.id")
    this.agg.y$species<-340
    z <- merge.catch(x, rvcat(this.agg.y), var = "number.caught")
  }
    if(s==880) { # hookear sculpins
    ss <- c(306, 880)
    this.y <- y[y$species %in% ss,]
    ## sum for each set
    agg.df <- aggregate(number.caught~unique.id, this.y, sum)
    vars <- c("unique.id", key(x))
    this.agg.y <- merge(x[,vars], agg.df, by="unique.id")
    this.agg.y$species<-880
    z <- merge.catch(x, rvcat(this.agg.y), var = "number.caught")
  }
  else{
    this.y <- y[y$species==s,]
    z <- merge.catch(x,this.y)
  }

stratified.number.df <- smean(z, "number.caught", by=c("year"))
matrix.row <- (stratified.number.df$mean - min(stratified.number.df$mean)) / (max(stratified.number.df$

C.matrix[i,] <- matrix.row
```

```r
C.df[C.df$species==s,"normalised"] <- matrix.row
}

# visualize C matrix using ggplot
C.df$species.name <- species.str(C.df$species, "english")
g <- ggplot(data=C.df,
       aes(x=year, y=species.name, fill=normalised)) + geom_tile() +
       theme_bw() + scale_fill_gradient(low = "white", high = "black")
g
```



Tow-level dataset

Also provide tow-level data for exploration purposes.

```r
strat.stats <- stratum.info(region = "gulf", survey = "rv", stratum = 415:439)

## write to CSV file
csv.fn2 <- paste0("sGSL-RV-data-for-Elise-", format(Sys.time(), "%Y-%m-%d"),"-strata-stats.csv")
keep.vars <- c("stratum","area","trawlable.units")
write.csv(strat.stats[,keep.vars], file=csv.fn2)

csv.fn3 <- paste0("sGSL-RV-data-for-Elise-", format(Sys.time(), "%Y-%m-%d"),"-sets.csv")
x.keep.vars <- c("unique.id","year","month","day","vessel.code","cruise.number","gear","stratum","set.nu
o.x <- order(x$year, x$set.number)
write.csv(x[o.x,x.keep.vars], file=csv.fn3)


csv.fn4 <- paste0("sGSL-RV-data-for-Elise-", format(Sys.time(), "%Y-%m-%d"),"-catch.csv")
y.keep.vars <- c("unique.id", "species", "number.caught", "weight.caught")
k.spec.keep <- c(strsplit(paste0(taxo.df.out$species.code, collapse="", sep=","), ",")[[1]])
write.csv(y[y$species %in% k.spec.keep,y.keep.vars], file=csv.fn4)

csv.fn5 <- paste0("sGSL-RV-data-for-Elise-", format(Sys.time(), "%Y-%m-%d"),"-C-matrix.csv")

write.csv(C.matrix, file=csv.fn5)

## create a backbone Excel file to copy and paste the CSV files into
library(xlsx)
xl.fn <- "sGSL-RV-data-for-Elise-2021-10-07.xlsx"
t.df <- data.frame(x=1)
write.xlsx(t.df, file=xl.fn, sheetName = "Species list")
write.xlsx(t.df, file=xl.fn, sheetName = "Strata statisitics", append = TRUE, row.names=FALSE)
write.xlsx(t.df, file=xl.fn, sheetName = "Sets", append = TRUE, row.names=FALSE)
write.xlsx(t.df, file=xl.fn, sheetName = "Adjusted catch", append = TRUE, row.names=FALSE)
write.xlsx(t.df, file=xl.fn, sheetName = "C matrix", append = TRUE, row.names=FALSE)
```

The xlsx package in R was giving a memory error when writing an Excel file, so I am outputting CSV files and assembling them in an Excel file manually instead.