

GSL-N: Gulf of St. Lawrence North survey data processing summary

fishglob, Aurore A. Maureaud, Juliano Palacios Abrantes, Zoë Kitchel, Dan Forrest, & Michelle Stuart

December, 2022

Contents

General info	1
Data cleaning in R	1
1. Overview of the survey data table	6
2. Summary of sampling intensity	7
3. Summary of sampling variables from the survey	8
4. Summary of biological variables	9
5. Extreme values	10
6. Summary of variables against swept area	11
7. Abundance or Weight trends of the six most abundant species	12
8. Distribution mapping	13
9. Taxonomic flagging	13
10. Spatio-temporal standardization	13
a. Standardization method 1	13
b. Standardization method 2	17
c. Standardization summary	17

General info

This document presents the cleaning code and summary of the Gulf of St. Lawrence North (Canada) bottom trawl survey provided by Fisheries and Oceans Canada. It contains data from 1980 and up to 2019.

Data cleaning in R

```
#####
##### R code to clean trawl survey for Gulf of St. Lawrence North
##### Public data Ocean Adapt
##### Contacts: Government of Canada; Fisheries and Oceans Canada
#####gddaisss-dmsaisb.XLAU@dfo-mpo.gc.ca
##### Coding: Michelle Stuart, Dan Forrest, Zoë Kitchel November 2021
#####
#NB: there are multiple events at similar locations on the same day because
#there is more than one vessel sampling, keep an eye on vessel name and haul_id
#-----#
##### LOAD LIBRARIES AND FUNCTIONS #####
#-----#
library(tidyverse)
library(lubridate)
library(googledrive)
library(taxize) # for getting correct species names
```

```

library(magrittr) # for names wrangling
library(readr)
library(dplyr)
library(PBSmapping)

source("functions/clean_taxa.R")
source("functions/write_clean_data.R")
fishglob_data_columns <- read_excel("standard_formats/fishglob_data_columns.xlsx")

#Data for the Gulf of St. Lawrence North can be accessed using the public Pinsky
#Lab OceanAdapt Git Hub Repository.
#Contact malin.pinsky@rutgers.edu for questions or help accessing

#Note that there have been gear changes and required calibrations
#for GSL-N
#and described well in here:
#Bourdages, H., Brassard, C., Desgagnés, M., Galbraith, P., Gauthier, J., Lambert, J., Légaré,
#B., Parent, E. and Schwab P. 2015. Preliminary results from the groundfish and shrimp
#multidisciplinary survey in August 2014 in the Estuary and northern Gulf of St. Lawrence.
#DFO Can. Sci. Advis. Sec. Res. Doc. 2014/115. v + 96 p.
#The analysis of 2014 abundance and biomass data were integrated into the combined
#annual summer survey series initiated in 1990. This combined series was developed
#following a comparative study between the two vessel-gear tandems (1990–2005: CCGS
#Alfred Needler – URI 81'/114' trawl; 2004–2012: CCGS Teleost – Campelen 1800 trawl) to
#establish specific correction factors for about twenty species caught (Bourdages et al.
#2007). This resulted in adjustment of Needler catches into Teleost equivalent catches.
#Note that the distinction between the two redfish species, Sebastes fasciatus and S.
#mentella, is based on the analysis of the soft anal fin rays count and the depth of capture
#of individuals (H. Bourdages, DFO Mont-Joli, pers. comm.).
```

-----#
PULL IN AND EDIT RAW DATA FILES #####
-----#

#GSL North Sentinel

GSLnor_sent <- read.csv(
 "https://github.com/pinskylab/OceanAdapt/raw/master/data_raw/GSLnorth_sentinel.csv")

#GSL North Gadus

GSLnor_gad <- read.csv(
 "https://github.com/pinskylab/OceanAdapt/raw/master/data_raw/GSLnorth_gadus.csv")

#GSL North Hammond

GSLnor_ham <- read.csv(
 "https://github.com/pinskylab/OceanAdapt/raw/master/data_raw/GSLnorth_hammond.csv")

#GSL North Needler

GSLnor_need <- read.csv(
 "https://github.com/pinskylab/OceanAdapt/raw/master/data_raw/GSLnorth_needler.csv")

```

#GSL North Teleost

GSLnor_tel <- read.csv(
  "https://github.com/pinskylab/OceanAdapt/raw/master/data_raw/GSLnorth_teleost.csv")

#-----#
#### REFORMAT AND MERGE DATA FILES ####
#-----#


#Bind all datasets

GSLnor <- plyr::rbind.fill(GSLnor_sent, GSLnor_gad, GSLnor_ham, GSLnor_need, GSLnor_tel)
GSLnor$lat <- as.numeric(as.character(GSLnor$Latit_Deb))
GSLnor$lon <- as.numeric(as.character(GSLnor$Longit_Deb))
GSLnor$depth <- as.numeric(as.character(GSLnor$Prof_Max))
GSLnor$Dist_Chalute_Position <- as.numeric(as.character(GSLnor$Dist_Chalute_Position))
GSLnor$Pds_Capture <- as.double(GSLnor$Pds_Capture)
GSLnor$Nb_Ind_Capture <- as.numeric(as.character(GSLnor$Nb_Ind_Capture))
GSLnor$date <- as.Date(GSLnor$date_Deb_Trait)
GSLnor$year <- as.integer(year(GSLnor$date))
GSLnor$verbatim_name <- trimws(as.character(GSLnor$Nom_Scient_Esp), which = "right")


GSLnor <- GSLnor[!is.na(GSLnor$lat),] #only keep rows with latitude
GSLnor <- GSLnor[!is.na(GSLnor$depth),] #only keep rows with depth

GSLnor <- GSLnor %>%
  # Create a unique haul_id
  mutate(
    haul_id = paste(GSLnor$Nom_Navire, GSLnor$No_Releve, GSLnor$Trait,
                    GSLnor$date_Deb_Trait, GSLnor$Hre_Deb, sep="-"),
    #area in km^2 =
    #Dist_Chalute_Position (nautical miles) * 1852 m/1 nautical mile *
    #                                trawl width *(1km^2/1000000m^2)
    area_swept = Dist_Chalute_Position * 1852 * 12.497 *(1/1000000),
    wgt = Pds_Capture, #in kg
    num = Nb_Ind_Capture, #in pieces
    # (via Daniel Ricard) trawl width, 12.497 m. Hurlbut and Clay (1990)
    # catch weight (kg.) per tow /km^2,
    wgt_cpue = (Pds_Capture)/area_swept,
    #weight in kg/time in minutes*60minutes/1hour
    wgt_h = (Pds_Capture)/Duree*60,
    #abundance in number/km^2
    num_cpue = Nb_Ind_Capture/area_swept,
    #abundance in number/hour
    num_h = Nb_Ind_Capture/Duree*60,
  )

GSLnor <- GSLnor %>%
  filter(
    # remove unidentified spp and non-species
    verbatim_name != "" | !is.na(verbatim_name),

```

```

!grepl("EGG", verbatim_name),
!grepl("UNIDENTIFIED", verbatim_name)) %>%
# add survey column
mutate(survey = "GSL-N")

#check that the number of unique haul_ids *
# spp combinations is the same as the number of rows in mar
nrow(GSLnor) == nrow(unique(GSLnor[,c("haul_id", "verbatim_name")]))

#it's not, so let's see why we have extras
#which(duplicated(GSLnor[,c("haul_id", "verbatim_name")]))

GSLnor <- GSLnor %>%
# Adding extra columns and setting proper format
mutate(
  country = "Canada",
  sub_area = NA,
  continent = "n_america",
  stat_rec = NA,
  station = NA,
  stratum = NA,
  season = NA,
  latitude = lat,
  longitude = lon,
  month = month(Date),
  day = day(Date),
  haul_dur = ifelse(Duree > 0, Duree/60, NA),
#get rid of negative duration values and code them as NA
  gear = Engin,
  sbt = NA,
  sst = NA,
  quarter = ifelse(month %in% c(1,2,3),1,
                    ifelse(month %in% c(4,5,6),2,
                          ifelse(month %in% c(7,8,9),3,
                                4
                               )
                            )
  ),
  aphia_id = NA,
  verbatim_aphia_id = NA,
) %>%
select(survey, haul_id, country, sub_area, continent, stat_rec, station, stratum,
       year, month, day, quarter, season, latitude, longitude, haul_dur, area_swept,
       gear, depth, sbt, sst, verbatim_name, num, num_h, num_cpue,
       wgt, wgt_h, wgt_cpue, verbatim_name, verbatim_aphia_id)

#-----#
##### INTEGRATE CLEAN TAXA FROM TAXA ANALYSIS #####
#-----#


# Get WoRM's id for sourcing
wrms <- gnr_datasources() %>%
  filter(title == "World Register of Marine Species") %>%

```

```

pull(id)

### Automatic cleaning
# Set Survey code
GSLnor_survey_code <- "GSL-N"

GSLnor <- GSLnor %>%
  mutate(
    taxa2 = str_squish(verbatim_name),
    taxa2 = str_remove_all(taxa2, " spp.| sp.| spp| sp|NO "),
    taxa2 = str_to_sentence(str_to_lower(taxa2))
  )

# Get clean taxa
clean_auto <- clean_taxa(unique(GSLnor$taxa2),
                           input_survey = GSLnor_survey_code, save = F, output=NA)

#This leaves out the following species, all of which are inverters
#Eualus gaimardii belcheri (invert)

#-----#
#### INTEGRATE CLEAN TAXA in GSL-North survey data ####
#-----#

correct_taxa <- clean_auto %>%
  select(-survey)

clean_GSLnor <- left_join(GSLnor, correct_taxa, by=c("taxa2"="query")) %>%
  filter(!is.na(taxa)) %>% # query does not indicate taxa entry that were
#removed in the cleaning procedure
# so all NA taxa have to be removed from the surveys because: non-existing,
#non marine or non fish
  rename(accepted_name = taxa,
         aphia_id = worms_id) %>%
  mutate(verbatim_aphia_id = NA,
        source = "DFO",
        timestamp = "2021") %>%
  select(fishglob_data_columns$`Column name fishglob`)

# -----#
#### SAVE DATABASE IN GOOGLE DRIVE ####
# -----#

# Just run this routine should be good for all
write_clean_data(data = clean_GSLnor, survey = "GSL-N_v2", overwrite = T)

```

1. Overview of the survey data table

survey	source	timestamp	haul_id		country	sub_area	continent
GSL-N	DFO	2021	Annie-Annick	-3-1-1995-08-07 -05:40:00	Canada	NA	n_america
GSL-N	DFO	2021	Annie-Annick	-3-1-1995-08-07 -05:40:00	Canada	NA	n_america
GSL-N	DFO	2021	Annie-Annick	-3-1-1995-08-07 -05:40:00	Canada	NA	n_america
GSL-N	DFO	2021	Annie-Annick	-3-2-1995-08-07 -07:45:00	Canada	NA	n_america
GSL-N	DFO	2021	Annie-Annick	-3-2-1995-08-07 -07:45:00	Canada	NA	n_america

stat_rec	station	stratum	year	month	day	quarter	season
NA	NA	NA	1995	8	7	3	NA
NA	NA	NA	1995	8	7	3	NA
NA	NA	NA	1995	8	7	3	NA
NA	NA	NA	1995	8	7	3	NA
NA	NA	NA	1995	8	7	3	NA

latitude	longitude	haul_dur	area_swept	gear	depth
49.85500	-62.530	0.5	0.0284677	Chalut de fond	182
49.85500	-62.530	0.5	0.0284677	Chalut de fond	182
49.85500	-62.530	0.5	0.0284677	Chalut de fond	182
49.70167	-62.465	0.5	0.0261532	Chalut de fond	229
49.70167	-62.465	0.5	0.0261532	Chalut de fond	229

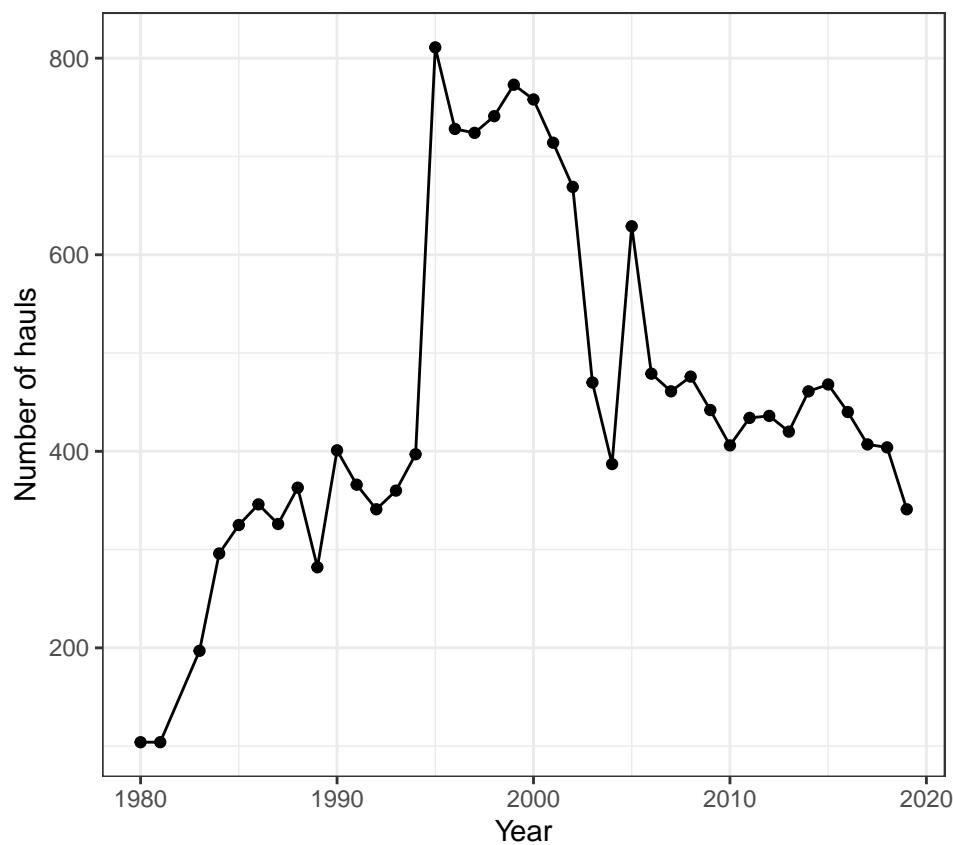
sbt	sst	num	num_h	num_cpue	wgt
NA	NA	16	32	562.04116	7.00
NA	NA	2	4	70.25514	0.06
NA	NA	34	68	1194.33746	3.50
NA	NA	2	4	76.47241	0.11
NA	NA	17	34	650.01552	9.00

wgt_h	wgt_cpue	verbatim_name	verbatim_aphia_id	accepted_name
14.00	245.893006	Gadus morhua	NA	Gadus morhua
0.12	2.107654	Sebastes sp.	NA	Sebastes
7.00	122.946503	Reinhardtius hippoglossoides	NA	Reinhardtius hippoglossoides
0.22	4.205983	Sebastes sp.	NA	Sebastes
18.00	344.125863	Reinhardtius hippoglossoides	NA	Reinhardtius hippoglossoides

aphia_id	SpecCode	kingdom	phylum	class	order	family
126436	69	Animalia	Chordata	Actinopteri	Gadiformes	Gadidae
126175	NA	Animalia	Chordata	Actinopteri	Perciformes	Sebastidae
127144	516	Animalia	Chordata	Actinopteri	Pleuronectiformes	Pleuronectidae
126175	NA	Animalia	Chordata	Actinopteri	Perciformes	Sebastidae
127144	516	Animalia	Chordata	Actinopteri	Pleuronectiformes	Pleuronectidae

2. Summary of sampling intensity

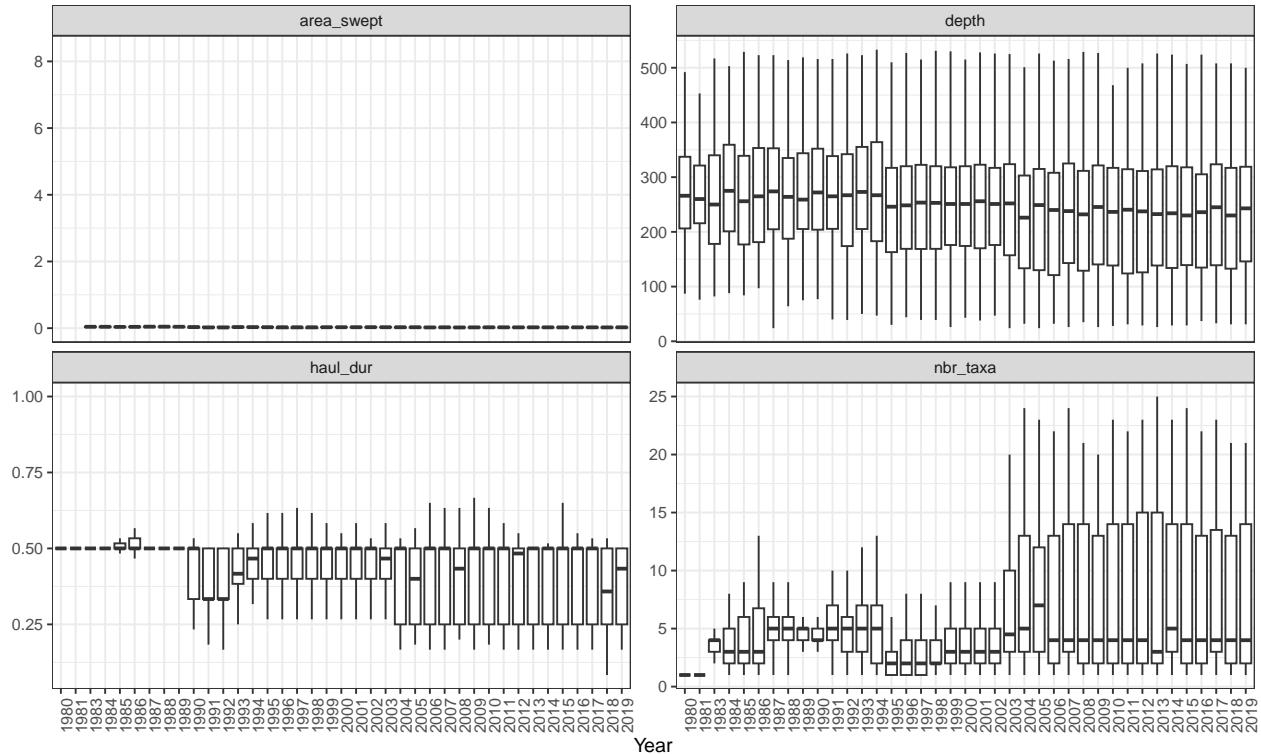
Number of hauls per year performed during the survey after data processing.



3. Summary of sampling variables from the survey

Here we show the yearly total and average of the following variables reported in the survey data:

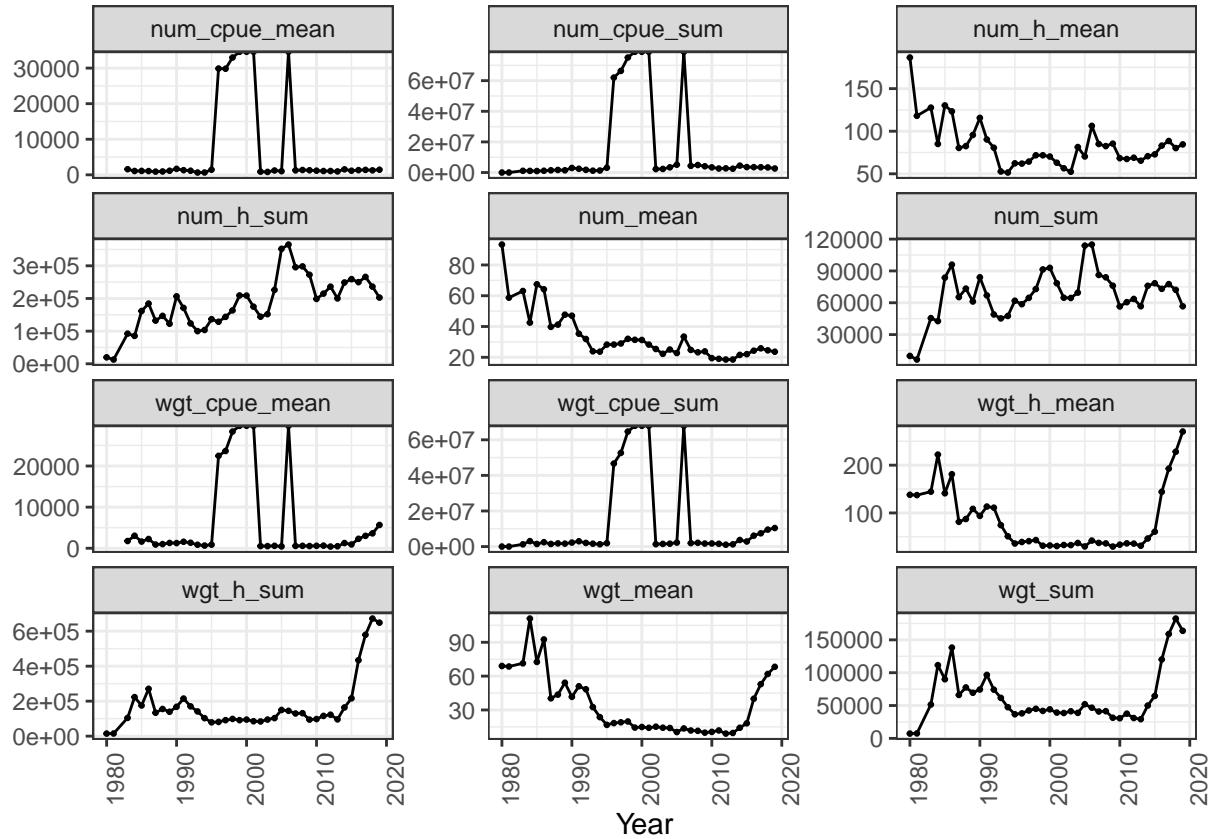
- *area_swept*, swept area by the bottom trawl gear km^2
- *depth*, sampling depth in m
- *haul_dur*, haul sampling duration *hour*
- *number of marine fish taxa*, taxa were cleaned following the last version of taxonomy from the World Register of Marine Species (<https://www.marinespecies.org/>, October 2021)



4. Summary of biological variables

Here we display the yearly total and average across hauls of the following variables recorded in the data:

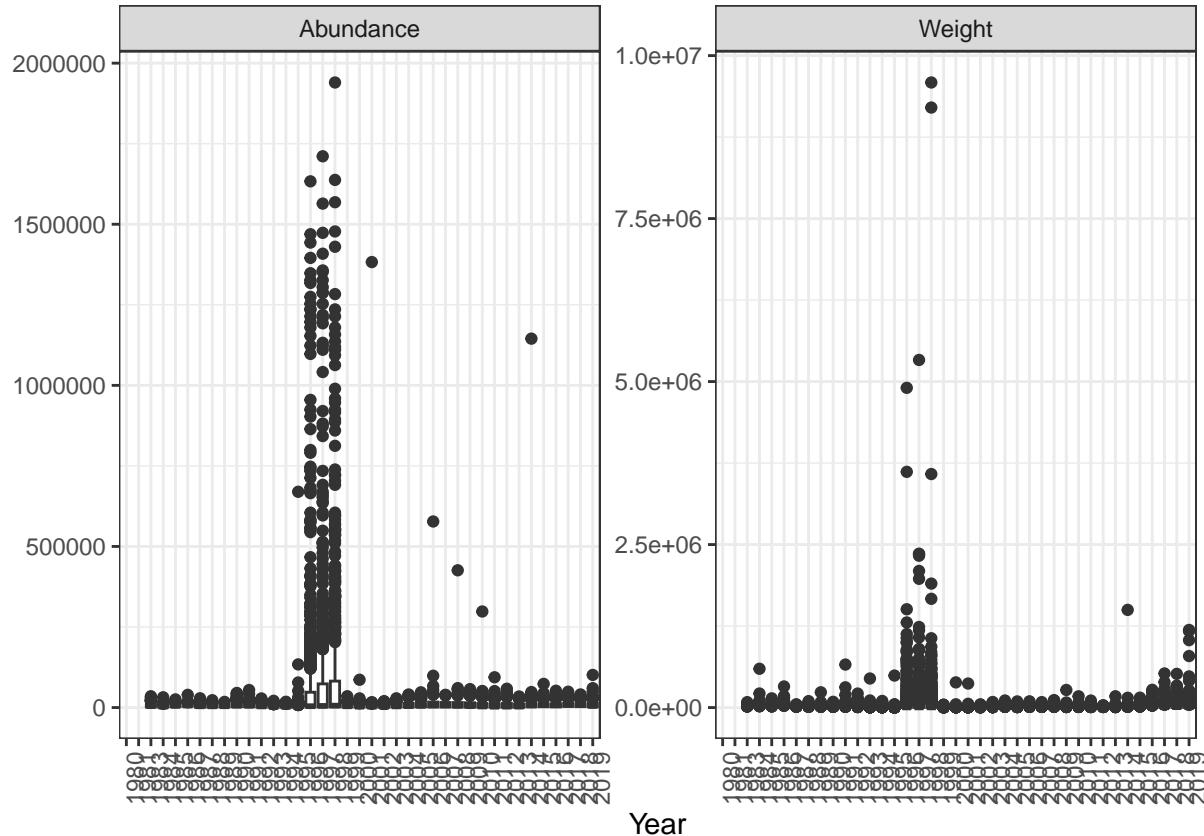
- num_cpue , number of individuals (abundance) in $\frac{individuals}{km^2}$
- num_h , number of individuals (abundance) in $\frac{individuals}{h}$
- num , number of individuals (abundance)
- wgt_cpue , weight in $\frac{kg}{km^2}$
- wgt_h , weight in $\frac{kg}{h}$
- wgt , weight in kg



5. Extreme values

Here we show a yearly total distribution of the biomass data to visualize outliers:

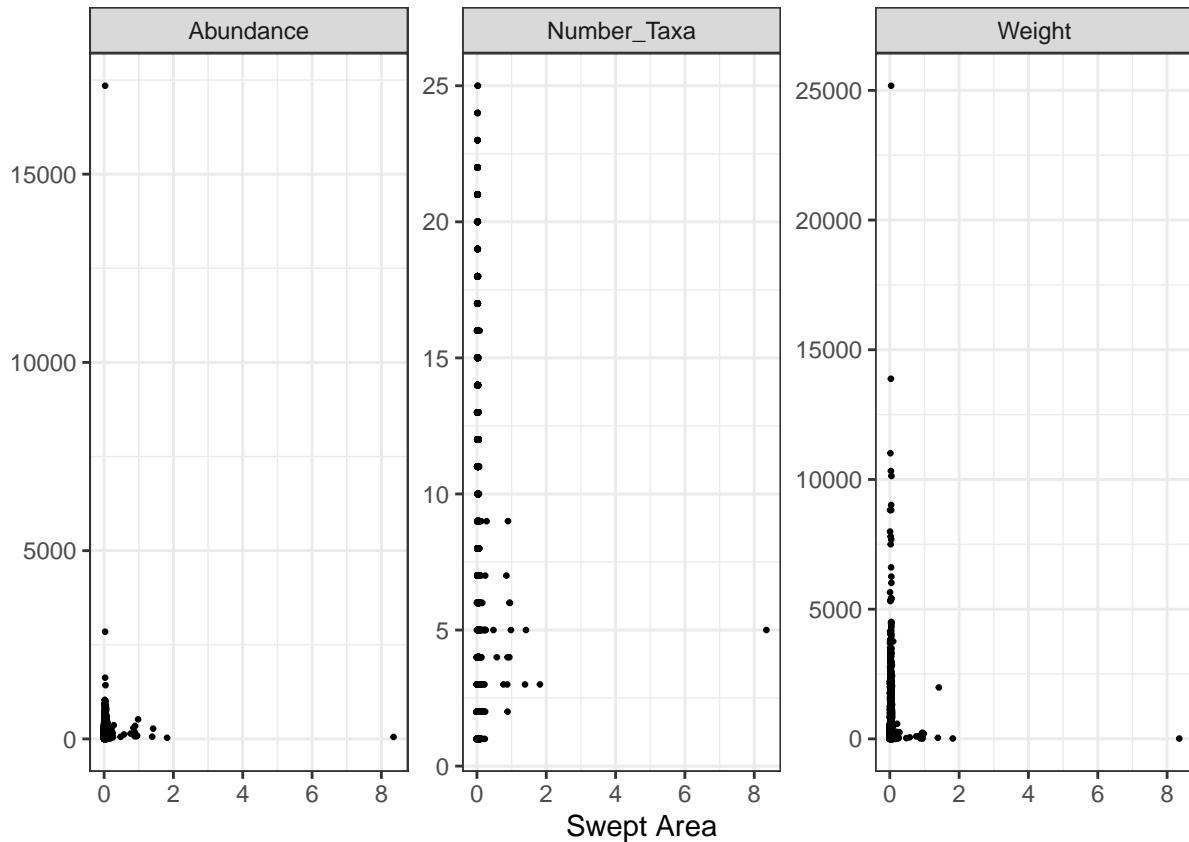
- wgt , total weight in kg per haul and year per haul and year, if available in the survey data
- num , total number of individuals, if available in the survey data



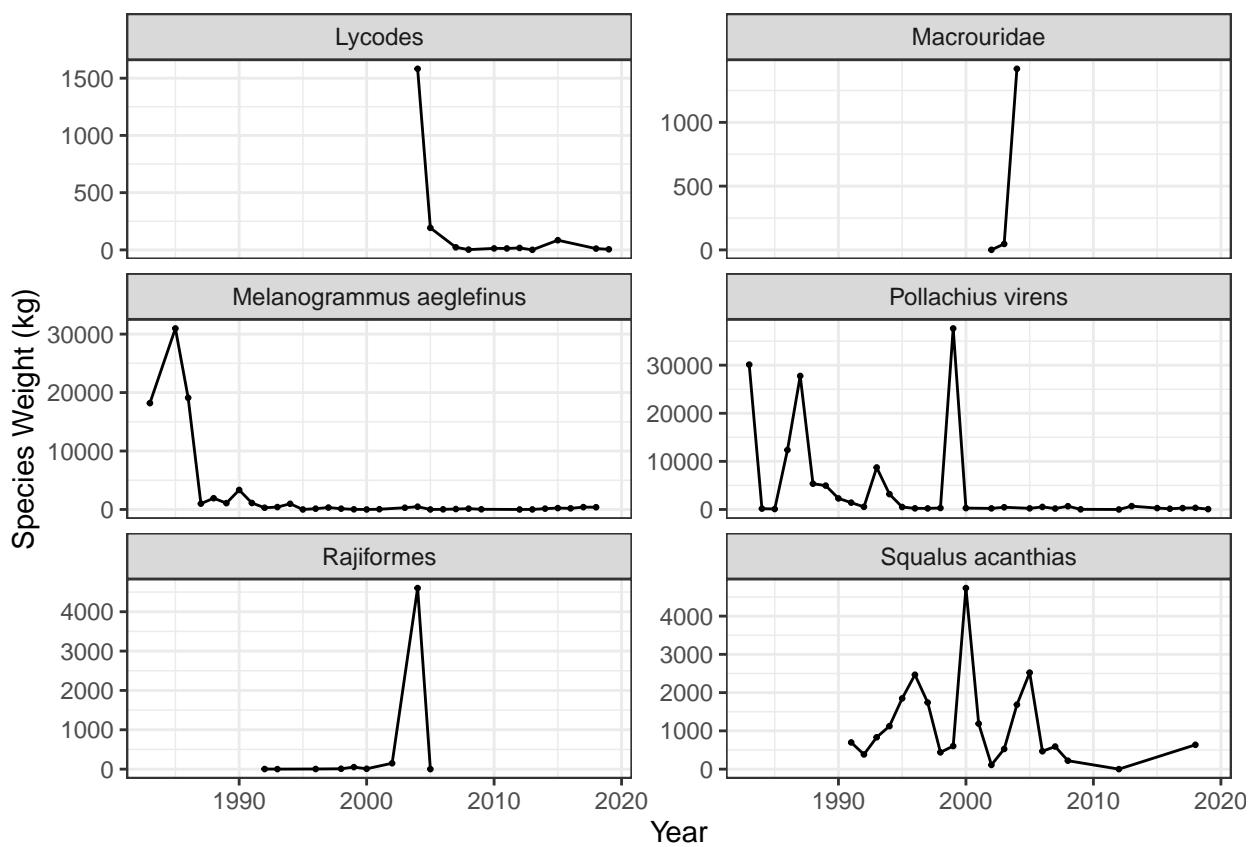
6. Summary of variables against swept area

Here we show the total abundance and number of taxa relationships with the area swept:

- *nbr_taxa*, number of marine fish taxa after taxonomic data cleaning
- *num*, number of individuals, if available in the survey data
- *wgt*, weight in *kg*, if available in the survey data

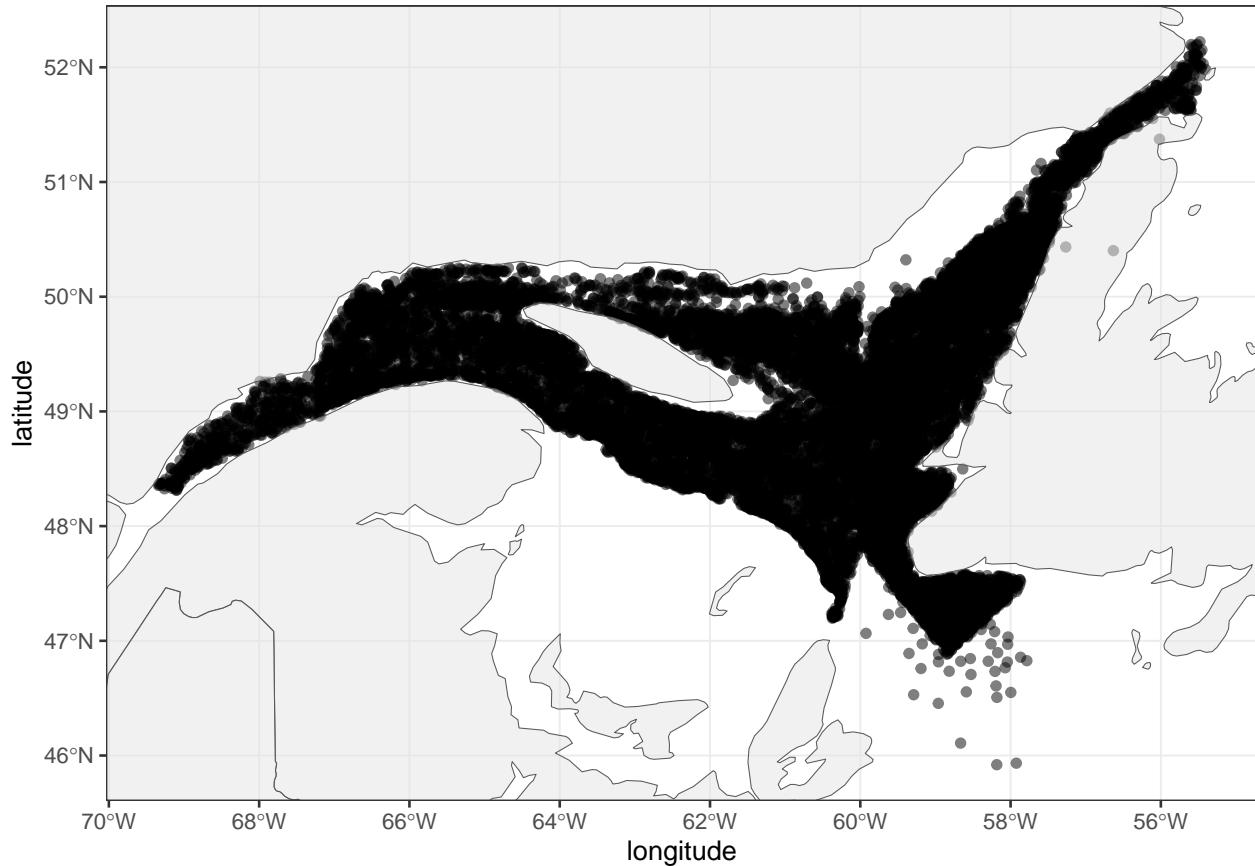


7. Abundance or Weight trends of the six most abundant species



8. Distribution mapping

Map of the sampling distribution in space. Note that we only show one year per coordinate.



9. Taxonomic flagging

This species flagging method was adapted from <https://github.com/pinskylab/OceanAdapt/blob/master/R/add-spp-to-taxonomy.Rmd#L33>

Visualization of flagged taxa

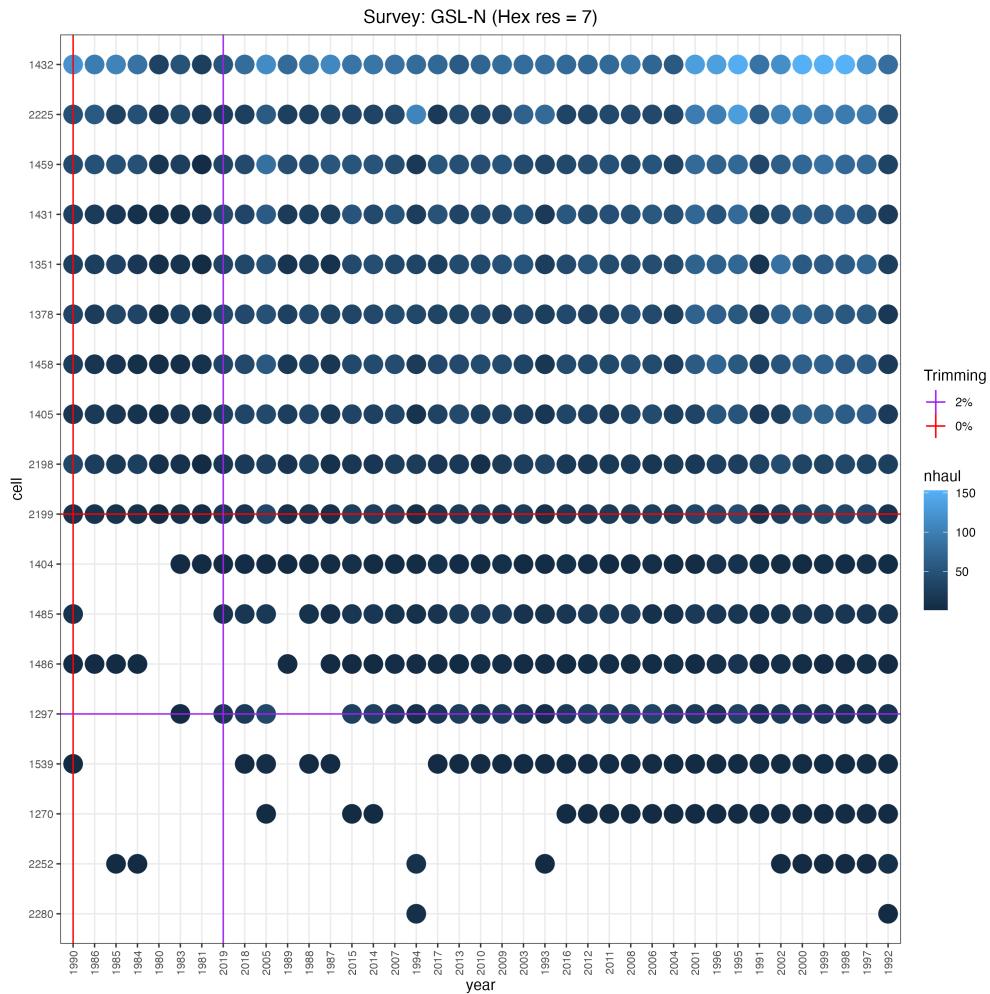
Statistics related to the taxonomic flagging outputs

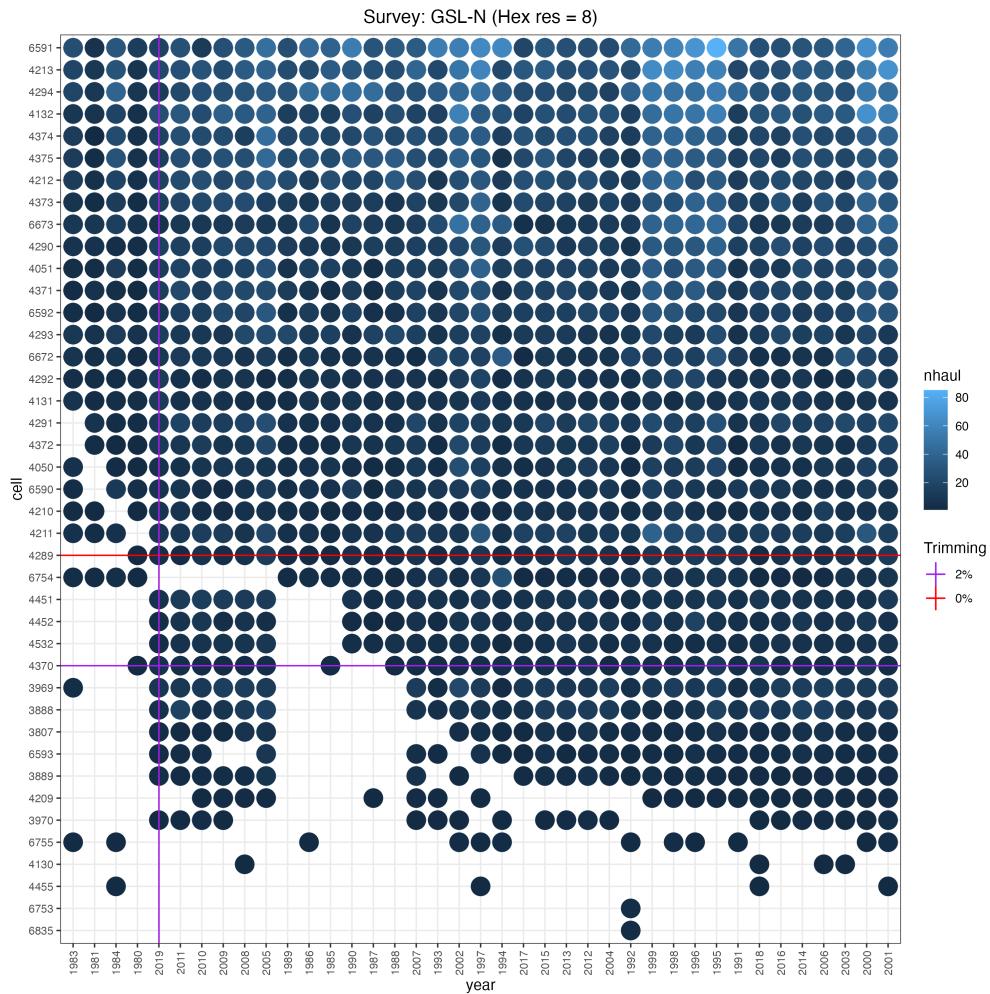
10. Spatio-temporal standardization

a. Standardization method 1

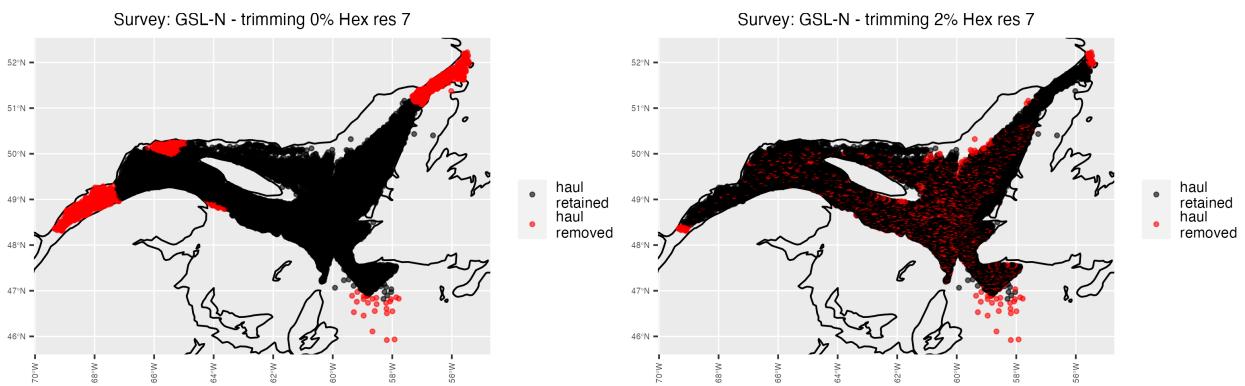
This standardization method was adapted from https://github.com/zoekitchel/trawl_spatial_turnover/blob/master/data_prep_code/species/explore_NorthSea_trimming.Rmd
It was run for hex resolution 7 and 8.

Plot of number of cells x years with overlaid flagging options

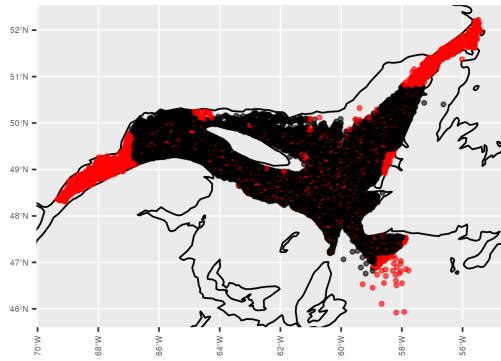




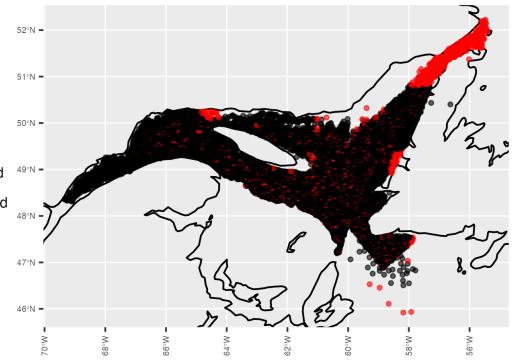
Map of hauls retained and removed per flagging method and threshold



Survey: GSL-N - trimming 0% Hex res 8

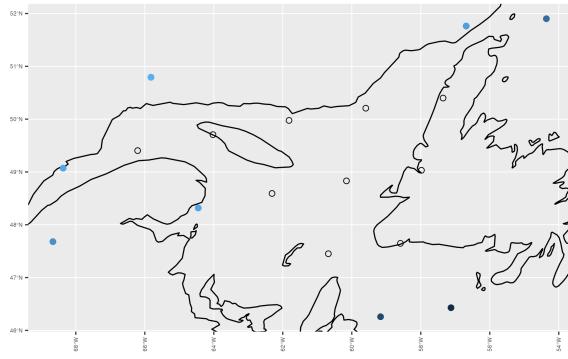


Survey: GSL-N - trimming 2% Hex res 8

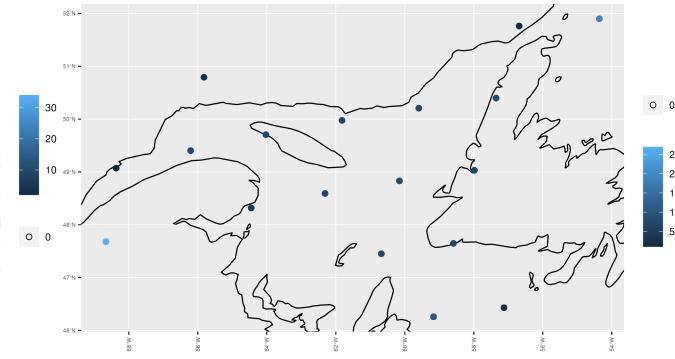


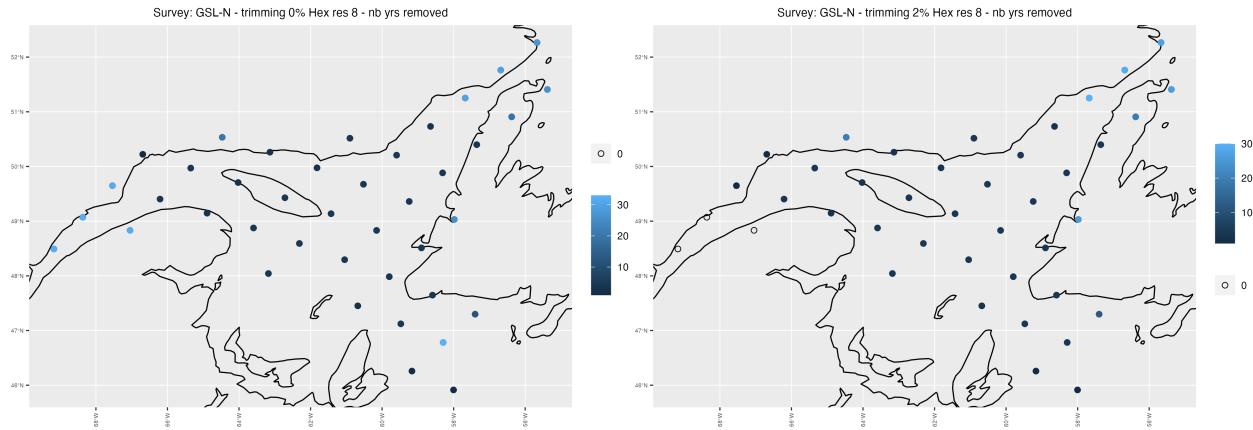
Map of numbers of years removed per grid cell and flagging method/threshold

Survey: GSL-N - trimming 0% Hex res 7 - nb yrs removed



Survey: GSL-N - trimming 2% Hex res 7 - nb yrs removed

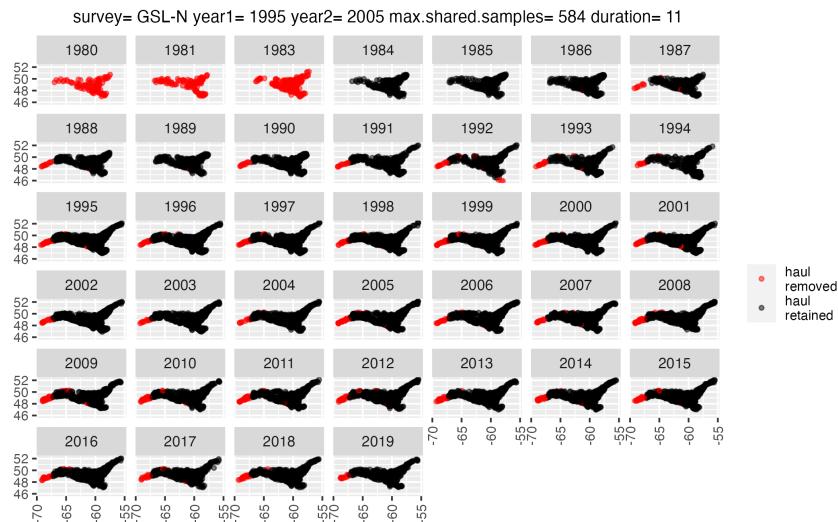




b. Standardization method 2

This standardization method was adapted from BioTIME code from https://github.com/Wubing-Xu/Range_size_winners_losers

Map of hauls retained and removed



c. Standardization summary

Statistics of hauls removed for each standardization method

summary	grid cell 7, 0% threshold	grid cell 7, 2% threshold	grid cell 8, 0% threshold	grid cell 8, 2% threshold	method 2 (biotime)
number of hauls removed	1420	1889.0	2533.0	1643.0	7281.0
percentage of hauls removed	8	10.7	14.3	9.3	7.8