

RVSurveyData

A package facilitating sharing of the DFO Maritimes Groundfish Survey Data with external partners.

Mike McMahon

July 15, 2022

This repo provides an authoritative version of the Maritimes RV Survey data in a single format closely resembling the internally-used data structure. This data is contributed to several different institutions in different formats, and it is anticipated that data transformations can be made more repeatable when run against a standard set of data, rather than against a series of potentially different SQL extractions. No functions/scripts are expected to be added to this package. Instead, these will be contributed via the sister package RVTransmogripher. RVTransmogripher is dedicated to the transformation of this RV survey data into formats required by various partners.

Disclaimer

The Information is licensed “as is”, and the Information Provider excludes all representations, warranties, obligations, and liabilities, whether express or implied, to the maximum extent permitted by law.

The Information Provider is not liable for any errors or omissions in the Information, and will not under any circumstances be liable for any direct, indirect, special, incidental, consequential, or other loss, injury or damage caused by its use or otherwise arising in connection with this licence or the Information, even if specifically advised of the possibility of such loss, injury or damage.

Appropriate caution is necessary in the interpretation of results derived from this data. Users must recognize that the analysis and interpretation of survey data requires background knowledge and expertise. Users should be aware of possible errors, including in the use of species names, geo-referencing, data handling, and mapping. They should crosscheck their results for possible errors, and qualify their interpretation of any results accordingly.

Installation

To install this package, run `devtools::install_github('PopulationEcologyDivision/RVSurveyData')`

Contents

Scripts in this package are minimal - it is primarily a data package. `updateCheck()` runs when the package is loaded, and informs the user if updated data is available.

Data

The most recent data currently available in this package is shown below (by Season):

```
##   YEAR SEASON
## 1 2021 SUMMER
## 2 2022 SPRING
## 3 2005  FALL
```

Following is a list of all of the data objects included in this package. Each is documented within the package, accessible via a command like `?GSCAT`.

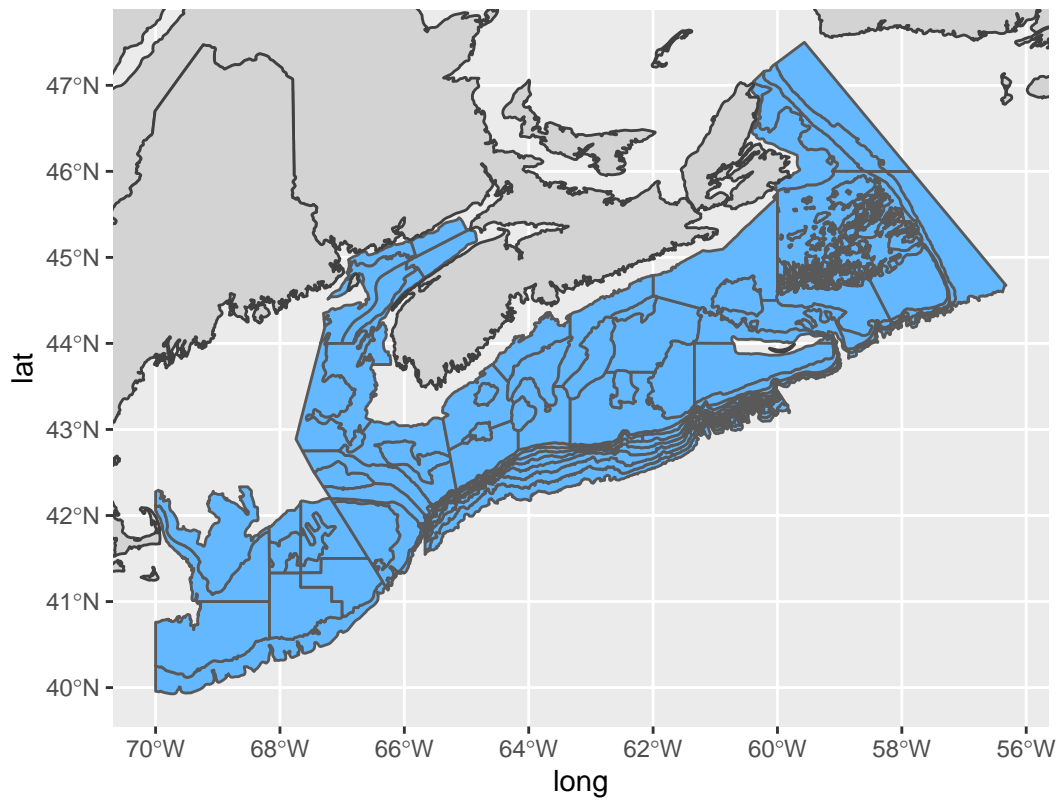
```
## GSAUX
## GSCAT
## GSCURNT
## GSDET
## GSFORCE
## GSGEAR
## GSHOWOBT
## GSINF
## GSMATURITY
## GSMISSIONS
## GSSEX
## GSSPEC
## GSSPECIES
## GSSPECIES_20220624
## GSSPECIES_TAX
## GSSTRATUM
## GSWARPOUT
## GSXTYPE
## nafo_sf
## strataMar_sf
```

Note that `GSSPECIES_20220624` and `GSSPECIES_TAX` are new (and draft), and include APHIAID information for each species.

Geospatial Data In addition to the survey data, two `sf` objects are included for convenience - the Maritimes strata (i.e. `strataMar_sf`) and the NAFO subunits (i.e. `nafo_sf`). Both are in WGS84 (i.e. CRS:4326).

Plots of each are shown below:

strataMar_sf



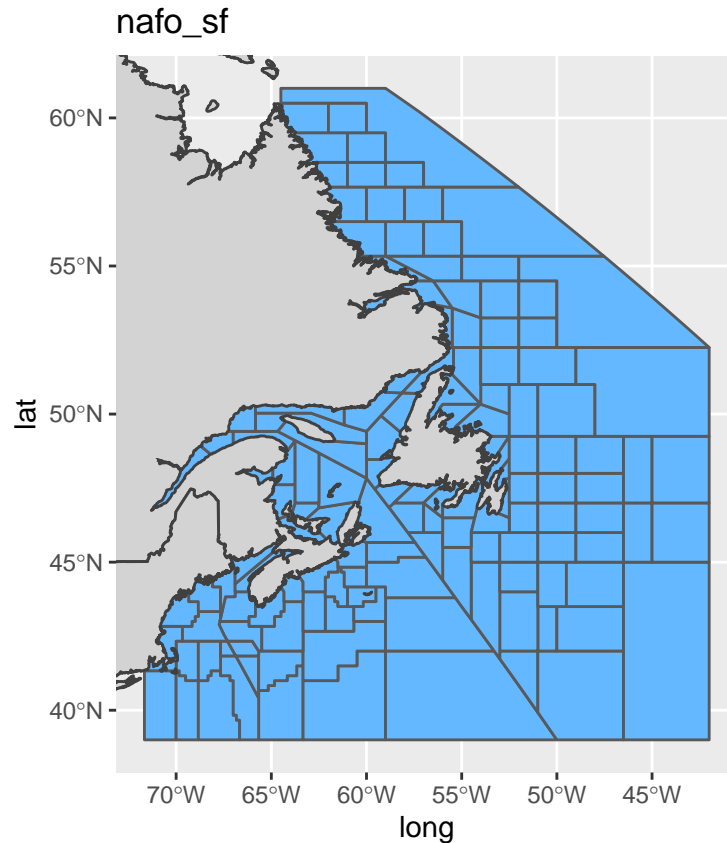


Table-Specific Notes

GSINF

- COORDINATES: The source database stores coordinates in DDMM.MM. These are retained, but have also been duplicated as *_DD, where the values have been converted to decimal degrees.
- DEPTH (values): Where DEPTH existed in the database, it is retained. If it is missing but values were present for both DMIN and DMAX, DEPTH is populated with the average of these 2 values.
- DEPTH (units): The source database stores depth values in fathoms. All depth fields have been duplicated as *_M, where the values have been converted to meters.

The “time” component of *GSINF\$SDATE* has been removed, as has *GSINF\$ETIME*. For the start time of the tow, one should use the “TIME” field. To determine the end time of the tow, add the tow duration (i.e. “DUR”) to the “TIME”.

When joining *GSINF* to *GSCAT*, it must be joined using **both** “MISSION” **and** “SETNO”.

GSCAT

- SAMPWGT - The SAMPWGT is the combined weight of the individuals that were looked at individually, and should be close to the sum of the individual values of FWT in GSDET.
- TOTWGT/TOTNO - When a trawl is hauled, all of the individuals for a given species are weighed and counted. These values become TOTWGT and TOTNO in GSCAT. For huge catches, the TOTNO is sometimes calculated from the weight.

- Cases exist where the `GSCAT$TOTWGT` is reported a “0” despite the `GSCAT$TOTNUM` indicating that at least one individual was caught. Set weights were reported to the nearest kg, and this is due to the measured weight not registering on the old spring balances (which had a precision of 1kg).
- `SIZE_CLASS` - Within the source database, this table includes a field called “`SIZE_CLASS`” which is not present here. This field is used to facilitate at-sea sampling by break a species into manageable chunks based on different observed sizes of animals within a set. For this table, all observations of a single species within a single set have had the weights and numbers of all size classes summed together.

*When joining to `GSCAT` to `GSDDET`, it must be joined using **all three** of the fields “`MISSION`”, “`SETNO`” and “`SPEC`”.*

GSDDET

- `FLEN` is generally in cm, with the caveats below:
 - For Scallops, lobster and crabs it is mm.
 - For herring, it switched from fork length in cm to total length in mm in SUMMER 2016 (missions “`TEL2016002`”, “`TEL2016003`” were still measured in cm).
 - For several grenadier species, we have started switching from ‘fork length’ to pre-anal fin length. This is only for very deep tows, where these species are regularly observed (It is not likely any of these species will be in the standard strata).

*When joining to `GSDDET` to `GSCAT`, it must be joined using **all three** of the fields “`MISSION`”, “`SETNO`” and “`SPEC`”.*

Notes on Species Identification

Not all species have always been confidently identified during the survey. The standard ID guide used is Scott and Scott, Atlantic fishes of Canada, which was published in 1988. The guide available prior to that was less detailed, and had a number of issues, such as it did not distinguish between Red Hake and White Hake. Only ~65 species have been reliably identified since the start of the survey (i.e. 1970). These are identifiable as those species with values of “1970-01-01” in `GSSPECIES_20220624$IDD_CONFID_SINCE`.

In the mid-90’s, an effort began to better differentiate sculpins and pouts more carefully, and as a result, the number of identified species increased.

The proliferation of invertebrate records ~2005 is not a biological phenomenon, but an artifact of an expansion in species identification and commensurate data collection. Prior to 2005, there are some invertebrate records, but the focus was primarily on commercial species (e.g. scallop, lobster).

Below ~40cm, differentiation between Winter and Little skates is questionable.

We are currently working to determine the dates at which other species began to be confidently identified and will continue to populate `GSSPECIES_20220624$IDD_CONFID_SINCE` as we are able.

Usage

`RVTransmogriker` is a companion package for this data, and includes tools for filtering this data to particular surveys, and for generating simple plots. It is recommended that it is used for data exploration. To install that package, run `devtools::install_github('PopulationEcologyDivision/RVTransmogriker')`

That being said, to view a particular data table, just load the package, and then issue a command:

```
library(RVSurveyData)
head(GSCAT)
```

##	MISSION	SETNO	SPEC	LENGTH_TYPE	LENGTH_UNITS	WEIGHT_TYPE	WEIGHT_UNITS	TOTNO	TOTWGT	SAMPWGT
## 1	ATC1970175	1	10	Fork Length	<NA>	Total Weight	Grams	12	59	59
## 2	ATC1970175	1	11	Fork Length	<NA>	Total Weight	Grams	4	10	10
## 3	ATC1970175	1	12	Fork Length	<NA>	Total Weight	Grams	19	19	19
## 4	ATC1970175	1	14	Fork Length	<NA>	Total Weight	Grams	4	0	0
## 5	ATC1970175	1	40	Fork Length	<NA>	Total Weight	Grams	17	6	6
## 6	ATC1970175	1	60	Fork Length	Centimeters	Total Weight	Grams	18	7	7