

# Rscript\_\_Diagram

Karolina

12/10/2020

## R script

The script developed to assign metiers based on several variables is called `script__metiers__test.R`.

## Prerequisites

The packages required to run the script are,

- `stringr`
- `data.table`
- `openxlsx`
- `purrr`

as well as auxiliary information described in sections XXX and a set of functions developed to facilitate the readability of the script. These functions are described below in detail.

- `loadInputData.R` : reads the Input file provided it's in csv format
- `loadAreaList.R` : reads the RCG area file from the Github repository
- `loadSpeciesList.R` : reads the RCG species file from the Github repository
- `loadMetierList.R` : reads the RCG metier file from the Github repository
- `loadGearList.R` : reads the Gear file from ?? (why not repo)
- `getMeasure.R` : Determines the if the sequence factor is weight or value.
- `getMetier.R` : Assigns the metier level 6.
- `getMetierLvl5FromPattern.R` : Assignes metier level 5

## Data

The data used as an input should be a csv file format as described in detail in section 2 of the report (Input format for transversal data). The first X rows of the example data set are shown below for clarity.

**Remove if not in pdf output or add as annex**

Country	year	vessel_id	vessel_length	trip_id	haul_id	fishing_day	area	ices_rectangle	gear	mesh	selection	registered
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201807011230		02-07-2018	27.3.d.25	37G5	OTB	110		
POL	2018	AAA-1	15	POLAAA1201807011230		02-07-2018	27.3.d.25	37G5	OTB	110		

selection	selection_type	selection_mesh
1_120	1	120
1_120	1	120
1_120	1	120

## Code

First the selection column is split between the selectivity device code, if any, and the selectivity device mesh size.

```
input.data[, ':='(c("selection_type", "selection_mesh"), data.table(str_split_fixed(selection, 2))))]
input.data[, ':='(selection_type, ifelse(selection_type == "",
NA, selection_type))]
input.data[, ':='(selection_mesh, ifelse(selection_mesh == "",
NA, selection_mesh))]
```

The RCG area is assigned based on the area (Subarea, Division, SubDivision) that was provided in the input data.

```
input.data <- merge(input.data, area.list, all.x = T, by = "area")
input.data[is.na(RCG) & substr(area, 1, 2) %in% c("31", "34",
"41", "47", "51", "57", "58", "87"), ':='(RCG, "LDF")]
input.data[is.na(RCG) & substr(area, 1, 2) == "37", ':='(RCG,
"MED")]
```

Next the target assemblage is assigned, depending on the species, and it's indicated if they belong to the Deep Water Species (DWS) group.

```
input.data <- merge(input.data, species.list, all.x = T, by = "FAO_species")
```

A set of variables is necessary to define the level the metier is assigned to and calculate the total weight and value of this group. The default sequence is shown below, however it can be changed to reflect national groupings.

```
sequence.def <- c("Country", "year", "vessel_id", "vessel_length",
"trip_id", "haul_id", "fishing_day", "area", "ices_rectangle",
```

FAO_species	species_group	dws_group
CJM	SPF	
COD	DEF	
FLE	DEF	
FPE	FWS	
HER	SPF	
HXC	LPF	DWS
MAS	SPF	
PLE	DEF	
SHO	DEF	DWS
SPR	SPF	
TUR	DEF	
YYY		

```
input.data[, ':='(seq_measure = getMeasure(measure)), by = sequence.def]
```

Next, the target assemblage is assigned based on the highest weight or value of each sequence.

```
input.data[seq_measure == "weight", ':='(seq_dom_group = species_group[which.max(seq_group_weight)], by = sequence.def]
input.data[seq_measure == "value", ':='(seq_dom_group = species_group[which.max(seq_group_value)], by = sequence.def]
```

The DWS rule is applied; when more than 8% of the species in the sequence are deep water, it is assigned in the DWS group.

```
input.data[dws_group == "DWS", ':='(seq_DWS_kg, sum(KG, na.rm = T)), by = c(sequence.def, "dws_group")]
input.data[, ':='(seq_total_kg, sum(KG, na.rm = T)), by = sequence.def]
input.data[, ':='(seq_DWS_perc, ifelse(is.na(seq_DWS_kg), 0, seq_DWS_kg/seq_total_kg * 100))
input.data[, ':='(seq_DWS_perc, max(seq_DWS_perc)), by = sequence.def]
input.data[seq_DWS_perc > 8, ':='(seq_dom_group, "DWS")]
```

The function getMetier.R first checks if the user has provided a target assemblage and gives it priority over the calculated dominant assemblage. Then it conditionally assigns the **metier level 6** in a stepwise procedure:

- The metier is assigned based on the RCG, year, gear, target assemblage, dominant target assemblage, mesh size, selectivity device and the mesh size of the selectivity device. The metiers with 0 or >0 mesh size are disregarded.
- If the metier is not assigned, the next step is to disregard the selectivity device code and mesh size and try to assign a metier again. The metiers with 0 or >0 mesh size

Country	year	vessel_id	vessel_length	trip_id	haul_id	fishing_day	area	ices_rectangle	gear	mesh	selection	register
POL	2017	GDY-555	40	POLGDY555201701011230		02-01-2017	87.3.3		OTM	50		
POL	2018	BBB-3	12	POLBB3123456		21-06-2018	27.3.d.24	37G4	GNS	220		
POL	2018	AAA-1	15	POLAAA1201807011230		02-07-2018	27.3.d.25	37G5	OTB	110		
POL	2018	ZZZ-2	12	POLZZZ2201801250940		25-01-2018	27.3.d.25	38G5	OTB	105	1_120	
POL	2010	ZZZ-2	12	POLZZZ2201001250940		17-03-2010	27.3.d.25	38G5	OTB	120		
POL	2020	DDD-4	28	POLDD9898		14-04-2020	27.3.d.26	39G9	GNS	110		FWS
POL	2020	DDD-4	28	POLDD9898		15-04-2020	27.3.d.26	39G9	GNS	110		DEF
POL	2020	DDD-4	28	POLDD9898		02-05-2020	27.3.d.26	39G9	OTB	115		
POL	2020	DDD-4	28	POLDD9898		28-04-2020	27.3.d.26	39G9	OTB	115		
POL	2020	DDD-4	28	POLDD9898		29-04-2020	27.3.d.26	39G9	OTB	115		
POL	2020	DDD-4	28	POLDD9898		30-04-2020	27.3.d.26	39G9	OTB	115		
POL	2020	DDD-4	28	POLDD9898		01-05-2020	27.3.d.26	39G9	OTB	115		
POL	2020	DDD-4	28	POLDD9898		14-04-2020	27.3.d.26	39G9	GNS	110		FWS
POL	2020	DDD-4	28	POLDD9898		15-04-2020	27.3.d.26	39G9	GN	1		DEF
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2018	AAA-1	15	POLAAA1201807011230		03-07-2018	27.3.d.25	37G5	OTB	110		
POL	2018	AAA-1	15	POLAAA1201806100325		10-06-2018	27.2.a		OTB	100		
POL	2020	DDD-4	28	POLDD9898		08-05-2020	27.3.d.26	39G9	XYZ	115		

FAO_species	area	measure	RCG	species_group	seq_group_KG	s
CJM	87.3.3	weight	LDF	SPF	18,000	
COD	27.3.d.24	value	BALT	DEF	396	
COD	27.3.d.25	weight	BALT	DEF	800	
COD	27.3.d.25	value	BALT	DEF	1,696	
COD	27.3.d.25	value	BALT	DEF	180	
COD	27.3.d.26	weight	BALT	DEF	150	

FAO_species	area	measure	RCG	species_group	seq_group_KG	seq_group_EUR
CJM	87.3.3	weight	LDF	<b>SPF</b>	18,000	0
COD	27.3.d.24	value	BALT	<b>DEF</b>	396	376
COD	27.3.d.25	weight	BALT	<b>DEF</b>	800	0
COD	27.3.d.25	value	BALT	<b>DEF</b>	1,696	1,611
COD	27.3.d.25	value	BALT	<b>DEF</b>	180	0
COD	27.3.d.26	weight	BALT	<b>DEF</b>	150	300

FAO_species	area	measure	RCG	species_group	seq_group_KG	seq_DWS_perc	seq_DWS_KG
HER	27.2.a	weight	NSEA	SPF	2,300	21	DWS_21
HXC	27.2.a	weight	NSEA	LPF	400	21	DWS_21
SHO	27.2.a	weight	NSEA	DEF	200	21	DWS_21
SPR	27.2.a	weight	NSEA	SPF	2,300	21	DWS_21

RCG	year	gear	seq_dom_group	mesh	selection_type	selection_mesh	metier_le
LDF	2017	OTM	SPF	50			OTM_SPF
BALT	2018	GNS	DEF	220			GNS_DEF
BALT	2018	OTB	DEF	110			OTB_DEF
BALT	2018	OTB	DEF	105	1	120	OTB_DEF
BALT	2010	OTB	DEF	120			OTB_DEF
BALT	2020	GNS	DEF	110			GNS_FW

year	vessel_id	metier_level_5	seq_no_lvl5	seq_perc_lvl5
2017	GDY-555	OTM_SPF	1	100
2018	BBB-3	GNS_DEF	1	100
2018	AAA-1	OTB_DEF	2	67
2018	ZZZ-2	OTB_DEF	1	100
2010	ZZZ-2	OTB_DEF	1	100
2020	DDD-4	GNS_FWS	1	11
2020	DDD-4	GNS_DEF	1	11
2020	DDD-4	OTB_DEF	5	56
2020	DDD-4	GN_DEF	1	11
2018	AAA-1	OTB_DWS	1	33
2020	DDD-4	XYZ_NA	1	11

assemblage. The column `seq_no_lvl5` shows the number of times the metier level 5 was used by the respective vessel in the same year. The percentage (`seq_perc_lvl5` column) is used to assign the dominant metier level 5 for each vessel.

```
input.data[, ':='(metier_level_5, paste(gear, ifelse(is.na(registered_target_assemblage),
  seq_dom_group, registered_target_assemblage), sep = "_"))]
pattern <- unique(input.data[, .SD, .SDcols = c(sequence.def,
  "metier_level_5")])
pattern <- pattern[, .(seq_no_lvl5 = .N), by = .(year, vessel_id,
  metier_level_5)]
pattern[, ':='(seq_perc_lvl5, seq_no_lvl5/sum(seq_no_lvl5, na.rm = T) *
  100), by = .(year, vessel_id)]
pattern <- pattern[!is.na(metier_level_5)]
input.data <- merge(input.data, pattern, all.x = T, by = c("year",
  "vessel_id", "metier_level_5"))
```

If a “rare” metier is assigned based on one different catch composition or a mis-assigned métier code, it will be set to one of the main métiers of the vessel belonging to this pattern. The percentage threshold for defining rare metiers can be set in the script. The default is 13, meaning that if there is less than 13 percent of the sequences that has a level 5 allocated, it is considered rare and will get the suffix “rare\_” in the column “metier\_level\_5”. If possible, the script will suggest another major metier used by the vessel to be used instead. If it is different from the métier code assigned it will have the suffix “pattern\_”.

vessel_id	year	gear	seq_dom_group	metier_level_5
DDD-4	2020	GN	FWS	pattern_OTB_DEF
BBB-3	2018	GNS	DEF	GNS_DEF
BBB-3	2018	GNS	DEF	GNS_DEF
BBB-3	2018	GNS	DEF	GNS_DEF
BBB-3	2018	GNS	DEF	GNS_DEF
DDD-4	2020	GNS	DEF	pattern_OTB_DEF

```

rare.threshold <- 13
input.data[seq_perc_lvl5 < rare.threshold, ':='(metier_level_5,
  NA)]
pattern <- pattern[seq_perc_lvl5 >= rare.threshold]
pattern[, ':='(c("gear", "target_assemblage"), data.table(str_split_fixed(metier_level_5,
  "_", 2))))]
pattern <- merge(pattern, gear.list, all.x = T, by.x = "gear",
  by.y = "gear_code")
input.data <- merge(input.data, gear.list, all.x = T, by.x = "gear",
  by.y = "gear_code")
input.data[is.na(metier_level_5), ':='(metier_level_5, as.character(pmap(list(vessel_id,
  year, gear, gear_group, registered_target_assemblage, seq_dom_group),
  function(v, y, g, gg, rt, d) getMetierLvl5FromPattern(v,
    y, g, gg, rt, d)))))]

```

## Diagram (test)

