Rmd files for FDI data calls

Second Training on quality checks
16/04/2024

Walter Zupa

COISPA

RDBqc: data quality checks for FDI format data call

Completely new functions have been designed on the FDI (Fishery Dependent Information) tables (in collaboration with HCMR) , relevant for Mediterranean and Black Sea (specifically G, H, I and J tables) aimed at verifying :

- the consistency of transversal variables (effort and landing);
- the spatial coverage;
- temporal coverage.

Cross-checks among the tables have been also implemented, following the checks described in the **STECF EWG 21-12 report**.

Table	Check functions	Table	Check functions
Table A	FDI_Check_empty_field_A	Table I	FDI_Check_empty_field_I
Table A	FDI_Check_record_duplicated_A	Table I	FDI_Check_record_duplicated_I
Table A	FDI_coverage	Table I	FDI_coverage
Table A	FDI_cov_tableA	Table I	FDI_checks_spatial_HI
Table A	FDI_disc_coverage	Table I	FDI_check_coord
Table A	FDI_prices_not_null	Table J	FDI_Check_empty_field_J
Table A	FDI_price_cov	Table J	FDI_Check_record_duplicated_J
Table G	FDI_Check_empty_field_G	Table J	FDI_coverage
Table G	FDI_Check_record_duplicated_G	Table J	FDI_cov_tableJ
Table G	FDI_coverage	Table J	FDI_vessel_lenth
Table G	FDI_cov_tableG	Cross checks	FDI_fishdays_cov
Table H	FDI_Check_empty_field_H	Cross checks	FDI_landweight_cov
Table H	FDI_Check_record_duplicated_H	Cross checks	FDI_vessel_numbers
Table H	FDI_coverage	Cross checks	FDI_Cross_checks_AG
Table H	FDI_checks_spatial_HI	Cross checks	FDI_Cross_Checks_AH
Table H	FDI_check_coord	Cross checks	FDI_Cross_Checks_IG

Table A: catch data

	country	year	quarter	vessel_length	fishing_tech	gear_type	target_assemblage
1	PSP	2014	1	NK	DTS	OTB	DEF
2	PSP	2014	2	NK	DTS	OTB	DEF
3	PSP	2014	3	NK	DTS	OTB	DWS
4	PSP	2014	1	VL0006	DRB	DRB	MOL
5	PSP	2014	4	VL0612	PMP	DRB	MOL
6	PSP	2014	1	VL1218	PMP	DRB	MOL

	mesh_size_range	metier	domain_discards	domain_landings
1	NK	OTB_DEF_>=40_0_0	1_GSA1_OTB_DEF_>=40_0_0	NK
2	NK	OTB_DEF_>=40_0_0	1_GSA6_OTB_DEF_>=40_0_0	NK
3	NK	OTB_DWS_>=40_0_0	1_GSA6_OTB_DWS_>=40_0_0	NK
4	<NA>	DRB_MOL_0_0_0	NK	NK
5	<NA>	DRB_MOL_0_0_0	NK	NK
6	<NA>	DRB_MOL_0_0_0	NK	NK

	supra_region	sub_region	eez_indicator	geo_indicator	nep_sub_region	specon_tech
1	MBS	GSA99	<NA>	NGI	NK	<NA>
2	MBS	GSA99	<NA>	NGI	NK	<NA>
3	MBS	GSA99	<NA>	NGI	NK	<NA>
4	MBS	GSA99	<NA>	NGI	<NA>	NK
5	MBS	GSA99	<NA>	NGI	<NA>	NK
6	MBS	GSA99	<NA>	NGI	<NA>	NK

	deep	species	totwghtlandg	totvallandg	discards	confidential
1	NA	RJC	0.000	0	0.241	N
2	NA	SQR	0.000	0	0.65	N
3	NA	LDB	0.000	0	0.06	N
4	NA	DXL	0.359	5086.562645	NK	A
5	NA	DXL	0.910	12893.51534	NK	N
6	NA	KLK	0.261	779.8609061	NK	A

Main data in Table A

- totwghtlandg
- totvallandg
- discards

Table G: effort data

```
> head(fdi_g_effort)
  country year quarter vessel_length fishing_tech gear_type target_assemblage
1     PSP 2014       1            NK          DTS       OTB                DEF
2     PSP 2014       2            NK          DTS       OTB                DEF
3     PSP 2014       3            NK          DTS       OTB                DWS
4     PSP 2014       1        VL0006          DRB       DRB                MOL
5     PSP 2014       4        VL0612          PMP       DRB                MOL
6     PSP 2014       1        VL1218          PMP       DRB                MOL
  mesh_size_range           metier supra_region sub_region eez_indicator
1              NK OTB_DEF_>=40_0_0          MBS      GSA99          <NA>
2              NK OTB_DEF_>=40_0_0          MBS      GSA99          <NA>
3              NK OTB_DWS_>=40_0_0          MBS      GSA99          <NA>
4            <NA>    DRB_MOL_0_0_0          MBS      GSA99          <NA>
5            <NA>    DRB_MOL_0_0_0          MBS      GSA99          <NA>
6            <NA>    DRB_MOL_0_0_0          MBS      GSA99          <NA>
  geo_indicator specon_tech deep totseadays totkwdaysatsea totgtdaysatsea
1           NGI          NK   NA         19              19        139.650
2           NGI          NK   NA        134             134       1590.408
3           NGI          NK   NA        244             244       2395.365
4           NGI          NK   NA         34              34        440.265
5           NGI          NK   NA         83              83        720.300
6           NGI          NK   NA         17              17        285.915
  totfishdays totkwfishdays totgtfishdays hrsea kwhrsea gthrsea totves
1       22.04        139.65         22.04    NK      NK      NK       1
2      186.78      1590.408        186.78    NK      NK      NK       7
3      241.49      2395.365        241.49    NK      NK      NK      11
4       34.12       440.265         34.12    NK      NK      NK       8
5       74.27         720.3         74.27    NK      NK      NK      13
6       13.69       285.915         13.69    NK      NK      NK       3
  confidential
1            Y
2            N
3            N
4            N
5            N
6            Y
```

Main data in Table G

- totseadays
- totkwdaysatsea
- Totgtdaysatsea
- totfishdays
- totkwfishdays
- totgtfishdays
- hrsea
- kwhrsea
- gthrsea
- totves

Table H: Landings by rectangle data

```
> head(fdi_h_spatial_landings)
  country year quarter vessel_length fishing_tech gear_type target_assemblage
1     PSP 2014       1            NK          DTS       OTB                NK
2     PSP 2014       2            NK          DTS       OTB                NK
3     PSP 2014       3            NK          DTS       OTB                NK
4     PSP 2014       1        VL0006          DRB       DRB              <NA>
5     PSP 2014       4        VL0612          PMP       DRB              <NA>
6     PSP 2014       1        VL1218          PMP       DRB              <NA>
  mesh_size_range           metier supra_region sub_region eez_indicator
1             DEF OTB_DEF_>=40_0_0          OFR      GSA99            NK
2             DEF OTB_DEF_>=40_0_0          OFR      GSA99            NK
3             DWS OTB_DWS_>=40_0_0          OFR      GSA99            NK
4             MOL    DRB_MOL_0_0_0          OFR      GSA99            NK
5             MOL    DRB_MOL_0_0_0          MBS      GSA99            NK
6             MOL    DRB_MOL_0_0_0          MBS      GSA99            NK
  geo_indicator specon_tech deep rectangle_type rectangle_lat rectangle_lon
1            NK        <NA> <NA>          05*05           NA           NA
2            NK        <NA> <NA>          05*05           NA           NA
3            NK        <NA> <NA>          05*05           NA           NA
4            NK        <NA> <NA>          05*05           NA           NA
5            NK          NK <NA>          05*05           NA           NA
6            NK          NK <NA>          05*05           NA           NA
    c_square species totwghtlandg totvallandg confidential
1 1000:100:1     LKO       0.08100           0           NA
2 1000:100:2     MUT       0.05565           0           NA
3 1000:100:2     HOM       0.09000           0           NA
4 1000:100:3     HMM       3.03670           0           NA
5 1000:100:1     MUT       0.00800           0           NA
6 1000:100:1     HOM       0.11700           0           NA
```

Main data in Table H

- rectangle_type
- rectangle_lat
- rectangle_lon
- c_square
- totwghtlandg
- totvallandg

Table I: Effort by rectangle data

```
> head(fdi_i_spatial_effort)
  country year quarter vessel_length fishing_tech gear_type target_assemblage
1     PSP 2014       1            NK          DTS       OTB                NK
2     PSP 2014       2            NK          DTS       OTB                NK
3     PSP 2014       3            NK          DTS       OTB                NK
4     PSP 2014       1        VL0006          DRB       DRB              <NA>
5     PSP 2014       4        VL0612          PMP       DRB              <NA>
6     PSP 2014       1        VL1218          PMP       DRB              <NA>
  mesh_size_range          metier supra_region sub_region eez_indicator
1             DEF OTB_DEF_>=40_0_0          OFR      GSA99            NK
2             DEF OTB_DEF_>=40_0_0          OFR      GSA99            NK
3             DWS OTB_DWS_>=40_0_0          OFR      GSA99            NK
4             MOL    DRB_MOL_0_0_0          OFR      GSA99            NK
5             MOL    DRB_MOL_0_0_0          MBS      GSA99            NK
6             MOL    DRB_MOL_0_0_0          MBS      GSA99            NK
  geo_indicator specon_tech deep rectangle_type rectangle_lat rectangle_lon
1            NK        <NA> <NA>          05*05           NA           NA
2            NK        <NA> <NA>          05*05           NA           NA
3            NK        <NA> <NA>          05*05           NA           NA
4            NK        <NA> <NA>          05*05           NA           NA
5            NK          NK <NA>          05*05           NA           NA
6            NK          NK <NA>          05*05           NA           NA
    c_square totfishdays confidential
1 1000:100:1       1.007            N
2 1000:100:2       1.008            N
3 1000:100:2       1.006            N
4 1000:100:3       1.008            N
5 1000:100:1       1.000            N
6 1000:100:1       1.000            N
```

Main data in Table I

- rectangle_type
- rectangle_lat
- rectangle_lon
- c_square
- totfishdays

Table J: Capacity and fleet segment effort data

```
> head(fdi_j_capacity)
  country year vessel_length fishing_tech supra_region geo_indicator
1     PSP 2014            NK          DTS          MBS           NGI
2     PSP 2014            NK          DTS          MBS           NGI
3     PSP 2014            NK          DTS          MBS           NGI
4     PSP 2014        VL0006          DRB          MBS           NGI
5     PSP 2014        VL0612          PMP          MBS           NGI
6     PSP 2014        VL1218          PMP          MBS           NGI
  principal_sub_region tottrips totkw totgt totves avgage    avgloa maxseadays
1               GSA99       11   221    21      1      7  11.78000        110
2               GSA99       56   471    18      1    444  12.68000        103
3               GSA99       99   441    45     44     15 444.00000        173
4               GSA99      577   444   108      3   4444  16.93333        444
5               GSA99       17   598    44     44     14  21.53000         27
6               GSA99      101   219    68      1     27  20.04000       55.5
```

Main data in Table J

- tottrips
- totkw
- totgt
- totves
- avgage
- avgloa
- maxseadays

```
<!-- USER DEFINED SOURCE FILES -->

```{r user_setup, include=FALSE}

## reading files
## UNCOMMENT THE FOLLOWING LINES FOR STAND-ALONE USE
## SELECT THE APPROPRIATE FILE PATHS ON THE LOCAL FOLDERS

tableA <- read.table("./fdi_a_catch.csv", sep = ";", header = TRUE)
tableG <- read.table("./fdi_g_effort.csv", sep = ";", header = TRUE)
tableH <- read.table("./fdi_h_spatial_landings.csv", sep = ";", header = TRUE)
tableI <- read.table("./fdi_i_spatial_effort.csv", sep = ";", header = TRUE)
tableJ <- read.table("./fdi_j_capacity.csv", sep = ";", header = TRUE)

MS <- "ITA"
SPs <- c("DPS")
GSAs <- c("GSA99")
# vessel_len = "COMBINED"
# fishtech = "COMBINED"
# metier <- "COMBINED"
```
```

The results are reported for all the combionations of vessel length, fishing technique and metier

```
<!-- USER DEFINED SOURCE FILES -->

```{r user_setup, include=FALSE}

## reading files
## UNCOMMENT THE FOLLOWING LINES FOR STAND-ALONE USE
## SELECT THE APPROPRIATE FILE PATHS ON THE LOCAL FOLDERS

tableA <- read.table("./fdi_a_catch.csv", sep = ";", header = TRUE)
tableG <- read.table("./fdi_g_effort.csv", sep = ";", header = TRUE)
tableH <- read.table("./fdi_h_spatial_landings.csv", sep = ";", header = TRUE)
tableI <- read.table("./fdi_i_spatial_effort.csv", sep = ";", header = TRUE)
tableJ <- read.table("./fdi_j_capacity.csv", sep = ";", header = TRUE)

MS <- "ITA"
SPs <- c("DPS")
GSAs <- c("GSA99")
vessel_len <- "COMBINED"
fishtech <- "COMBINED"
metier <- "COMBINED"
```
```

The results are reported aggregated at vessel length, fishing technique and metier level