

Observaciones sobre la Metodología de la Base de Metadatos

Juliano Palacios Abrantes

5/11/2017

Contents

MMID[601:751]	1
Autor: CONABIO	1
MMID[612]	2
Autor: CONABIO	2
MMID[759:1246]	2
Autor: Secretaria de Economia	2
MMID[972:994]	4
Autor: Secretaria de Economia	4
MMID[1523:1527]	4
Autor: Monitoreo NorOeste	4
MMID[1962:2075]	5
Autor: NOAA	5
MMID[34468:35071]	5
Autor: Damien Olivier Kunz	5
MMID[35647:35763]	5
Autor: Bertha Lavaniegos et al	5

##

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

##

filter, lag

The following objects are masked from 'package:base':

##

intersect, setdiff, setequal, union

MMID[601:751]

Autor: CONABIO

Publication__Year

There are (often) three different dates on CONABIO's portal:

- 1. Publication year (*Fecha de publicación*)

- 2. Metadata Publication year (*Fecha de publicación del metadato*)
- 3. Sometimes Author comes with a year

Prioritization was given in the following order: 3,1,2.

MMID[612]

Autor: CONABIO

Data_Time_Points

Because we are not sure what social information is relevant for researchers we aggregated all the options on the category *Población* with the exception of “*Pobalción Indígena*”. This adds up to 134 different databases witch are used as “Data_Points”. The category MMID[613] is Indigena and has Data_Points = 43, following the same methodology.

Población:

- Aspectos generales [89]
- Riesgo [1]
- Indicadores Sociales [44]

MMID[759:1246]

Autor: Secretaria de Economia

Data_Time_Points

The records provided by the *Secretaría de Economía* are not standard. They are highly variable at all levels (*Marinos, filetes, por estado and por punto de cotizacion*). For example, for Abadejo there are **10** different *Puntos de cotiización* for *Precios en Origen*, **18** for *Precios de Menudeo*, and **6** for *Precios en Destino*. In addition to these variations, all of these categories have subsets that are not the same (see picture below).

Hence, for each of the main categories I randomly selected options until get to a final number. As exemplified here:

```
#### FIRST CATEGORY (MARINO) ####
```

```
N_Marinos <- 67 #Total options within MARINO category
```

```
# Select the [n] elements to gather data from
```

```
Sample_Marino <- sort(round(runif(
  30, #Total Number of options to select
  min=1,
  max=N_Marinos)
)
```

```
##### SECOND CATEGORY (PRODUCTO) ####
```

```
# Lets assume Number 1 (Abadejo) got selected. For each of the Three categories (Precios en Origen, Pre
```

Producto: Abadejo

Registros del Día 11/May/1998 al 11/May/2017

PRECIOS en ORIGEN					
Punto de Cotización	Min	Max	Prom	Origen Frec	Distrib Orig
Alvarado, Veracruz	0.00	115.00	68.40	VER	VER(735)
Campeche, Campeche	7.00	70.00	19.95	CAM	CAM(215)
Coahuayana, Michoacán	7.00	25.00	17.40	MCH	CAM(2)MCH(3)
La Paz, Baja California Sur	7.00	85.00	46.75	BCS	BCS(4)
Lazaro Cardenas, Michoacán	30.00	40.00	40.00	MCH	MCH(1)
Merida, Yucatán	72.00	200.00	136.00	YUC	YUC(2)
San Blas, Nayarit	80.00	90.00	90.00	NAY	NAY(1)
Tamaulipas, Tamaulipas	2.00	110.00	47.43	TAM	TAM(37)
Tuxpan, Veracruz	13.00	90.00	52.00	VER	VER(13)
Veracruz, Veracruz	22.00	85.00	49.27	VER	VER(298)
PRECIOS de MENUDEO					
Punto de Cotización	Min	Max	Prom	Origen Frec	Distrib Orig
Campeche, Campeche	10.00	110.00	35.09	CAM	CAM(17)AGS(4)VER(1)
Colima, Colima	85.00	180.00	116.00	COL	COL(5)
Culiacán, Sinaloa	50.00	55.00	52.50	SIN	SIN(2)
Ecatepec, México	35.00	64.00	45.00	DF	DF(3)
Irapuato, Guanajuato	37.00	70.00	53.50	YUC	YUC(2)
Jalapa, Veracruz	9.90	110.00	53.61	VER	VER(227)AGS(3)
Merida, Yucatán	80.00	80.00	80.00	YUC	YUC(1)
Monterrey, N. León	62.90	87.90	78.68	SON	SON(2)SIN(1)IMP(1)
Morelia, Michoacán	36.00	70.00	45.56	MCH	MCH(9)
Oaxaca, Oaxaca	30.00	110.00	65.71	OAX	OAX(7)
Pescaderías en Cd. Obregón	28.00	75.00	52.88	SON	SON(13)
Pescaderías en Hermosillo	28.00	80.00	42.00	SON	SON(5)
Querétaro, Querétaro	39.00	130.00	85.00	DF	DF(1)QUE(1)
San Luis Potosí, San Luis Potosí	30.00	80.00	49.00	TAM	JAL(1)TAM(5)
Tamaulipas, Tamaulipas	6.00	75.00	42.45	TAM	TAM(94)
Torreón, Coahuila	25.00	28.90	26.90	TAM	TAM(1)
Tuxtla Gutiérrez, Chiapas	25.00	120.00	81.50	CHIS	CHIS(10)
Veracruz, Veracruz	36.00	8,080.00	75.07	VER	VER(351)
PRECIOS en DESTINO					
Punto de Cotización	Min	Max	Prom	Origen Frec	Distrib Orig
León, Guanajuato	35.00	102.00	68.50	TAM	YUC(3)TAM(5)MCH(4)
Monterrey, N. León	25.00	85.00	56.00	SON	SON(3)TAM(1)SIN(1)
Oaxaca, Oaxaca	12.00	110.00	58.40	OAX	OAX(5)
Puebla, Puebla	13.00	100.00	34.52	CAM	CAM(7)TAM(3)NOESP(6)VER(4)SIN(1)
Tuxtla Gutiérrez, Chiapas	40.00	90.00	48.57	CHIS	CHIS(7)
Zapopan, Jalisco	10.00	45.00	23.82	SIN	MCH(2)NAY(3)BC(6)SIN(7)SON(1)CAM(1)JAL(2)

Figure 1: Fig.1 Example of how complicated the SE-SNIIM data is

```

#Abadejo
#Precio en el Orgien
N_Pe0 <- round(10 * 0.3,0) #<- 30% of the total options since they fluctuate
# Select the locations to review
Pe0_Total <- sort(round(runif(
  N_Pe0,
  min=1,
  max=10),
  digits = 0))

#Pe0_Total <- [1] 5 6 7
# 5 <- Lazaro Cardenas, Michoacan
# 6 <- Merida, Yucatan
# 10 <- Veracruz, Veracruz

# Now, for each one of them, we see how many data points there are
# Lazaro Cardenas, Michoacan (n =1 )
# Merida, Yucatan (n = 2)
# Veracruz, Veracruz (n =298)

```

This process would be repeated for each of the initial selections under each of the main categories (Marinos, Crustaceos, Moluscos y Otros, Filetes y Otros, Por Estado and Por Punto de Cotización).

After having a set of numbers for each of the categories I randomly allocated the numbers to each record of the metadata. Hence, each product has three categories:

- Precio de Venta de *product* en el Origen (1998-2016)
- Precio de Menudeo de *product* (1998-2016)
- Precio de *product* en Destino de Venta (1998-2016)

With random Data Time Points but start year of 1998 and end year of 2017.

MMID[972:994]

Autor: Secretaria de Economia

Subject_Name

There are different shrimp presentations (e.g. *camaron grande*, *vamaron chico*, *cabeza de camaron*, etc). We included all as “*camaron*”. The same for *Langosta*, *Pulpo*, and any other species with different presentaions.

MMID[1523:1527]

Autor: Monitoreo NorOeste

Because Monitoreo’s program is to generate metadata and not data itself (like us) we are not duplicating records. Because of this, there are only five categories for Monitoreo that represent a total of 252 data points.

Observation

Monitoreo gave us their disaggregated dataset for the final analysis.

MMID	Short_Title	Institution	Data_Time_Points
1523	Monitoreo de Especies en el Noroeste	MonitoreoNoroeste	142
1524	Monitoreo de Grupo de Especies en el Noroeste	MonitoreoNoroeste	30
1525	Monitoreo Fisico-quimico en el Noroeste	MonitoreoNoroeste	17
1526	Monitoreo Socioeconomico en el Noroeste	MonitoreoNoroeste	44
1527	Monitoreo de Ecosistemas en el Noroeste	MonitoreoNoroeste	19

MMID[1962:2075]

Autor: NOAA

MMID

Oceanographic data presented by NOAA's is available by season and depth. Hence, each season has its own record for each depth:

MMID	Short_Title	Author
1965	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Annual/10m)	NOAA
1966	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Fall/10m)	NOAA
1967	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Winter/10m)	NOAA
1968	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Spring/10m)	NOAA
1969	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Summer/10m)	NOAA
1970	Gulf of Mexico Dissolved Oxygen - Climatological Mean (Annual/250m)	NOAA

MMID[34468:35071]

Autor: Damien Olivier Kunz

Data_Time_Points

All the records have 9 data time points. Damien used 9 publications (of different years) to gather all data. Hence, I thought that it will be the best to just use that number instead of going publication by publication searching for each year of sampling. Also, he compiled a database of other bases that we already have (e.g. Hector Relles Pangas). Hence, those years are not captured on Damien's database to avoid duplication.

MMID[35647:35763]

Autor: Bertha Lavaniegas et al

Data_Time_Points

Appendix 1, Abundance of each species is shown seasonally (4). However, instead of writing down 4 records per species, I will multiply the Number of data_Time_Points x 4.

Observation

NOAA's data on the Gulf of Mexico is spread by season*