

Manejo básico de Redatam usando R

Andrés Gutiérrez, Felipe Molina y Stalyn Guerrero

Introducción

El siguiente documento presenta algunas características para el manejo de datos censales en redatam utilizando R.

R y RStudio

Las siguientes librerías nos permiten lectura, uso y manejo de datos Redatam en R.

```
library(Rcpp)
library(RcppProgress)
library(redatam)
library(dplyr)
library(tidyverse)
library(magrittr)
```

La librería redatam en R

Para ejemplificar el uso de estas librerías, se utilizará el conjunto de datos Redatam correspondiente al censo de PAN 2010. Dicho conjunto de datos contiene la información censal en diferentes archivos de entidades. Para leer el conjunto de datos mediante la función `redatam.open` se utiliza el archivo `.dicx` que se encuentra en la carpeta correspondiente.

```
PAN <- redatam.open("Data/CP2010PAN/BaseOriginal/PAN2010.dicx")
```

La función `redatam.entities()` permite verificar las entidades contenidas en la información censal, dichas entidades poseen la información censal correspondiente. Las preguntas censales de cada entidad pueden ser verificadas mediante `redatam.variables`.

```
redatam.entities(PAN)
```

```
## [1] "PAN2010" "PROVIN" "DISTRITO" "CORREG" "LUGPOB" "BARRIADA"
## [7] "SEGMENTO" "VIVIENDA" "HOGAR" "PERSONA"
```

```
redatam.variables(PAN, "PROVIN")
```

```
## [1] "PROV" "PROVNOMB"
```

```
redatam.variables(PAN, "PERSONA")
```

```
## [1] "NPERSONA" "P01RELA" "P01RELA0" "P011NUCL" "P02SEX0" "P03EDAD"
## [7] "P04SEGSOC" "P05ESTC" "P06A0IR" "P06BVER" "P06CCAMI" "P06DUSARB"
## [13] "P06EHABL" "P06FAPREN" "P07DISCA" "P07DISCO" "P08INDIG" "P08INDIO"
## [19] "P09AFROD" "P09AFROO" "P10NACI" "P10APERI" "P11VIVE" "P12VIVIA"
## [25] "P12APERI" "P13ESCU" "P14GRADO" "P15SLYE" "P16TITUC" "P17TRAB"
## [31] "P18TRAUS" "P19ALGT" "P20BSEM" "P21BMES" "P22MOTI" "P22MOTIO"
```

```
## [37] "P230CUPC" "P25ACTIC" "P26TRAB" "P27EMPL" "P28ASUEL" "P28BXIIM"
## [43] "P28CINDE" "P28DJUBI" "P28EPENS" "P28F1PENS" "P28F2DINE" "P28F3OTRA"
## [49] "P28F3OTRD" "P28GALQU" "P28HBECA" "P28IAGRO" "P28JOTRIN" "P28AINGR"
## [55] "P29HIJOS" "P30VIVOS" "P31HU12M" "P32VU12M" "INGRPER" "RP14GRADO"
## [61] "RP17TRAB" "RP230CUP" "RP25ACTI" "RP26TRAB" "RPROVDIST" "RP10NACI"
## [67] "RP11VIVE" "RP12VIVIA" "AAPROB" "RINGRPER" "EDADQUIN"
```

`redata.query` permite realizar un tabulado de la información censal disponible en cada entidad, para ello, utilizando la programación básica de `redatam` se llama a cada variable anteponiendo el nombre de la entidad. Dicha función genera de forma automática una tabla llamada `Tabla1`, también puede ser guardada como un objeto en R.

En el siguiente ejemplo se agrega el número de personas según sexo por provincia.

```
Tabla1 <- redatam.query(PAN,
  "
  freq PROVIN.PROVNOMB
  by PERSONA.P02SEXO
  "
)
```

```
head(Tabla1)
```

```
##   PROVNOB1_value PROVNOB1_label P02SEX02_value P02SEX02_label value
## 1 BOCAS DEL TORO BOCAS DEL TORO             1      Hombre 65043
## 2 BOCAS DEL TORO BOCAS DEL TORO             2      Mujer 60418
## 3      CHIRIQUÍ      CHIRIQUÍ             1      Hombre 211618
## 4      CHIRIQUÍ      CHIRIQUÍ             2      Mujer 205255
## 5          COCLÉ          COCLÉ             1      Hombre 119417
## 6          COCLÉ          COCLÉ             2      Mujer 114291
```

Notar que `_value` corresponde al valor que toma la variable en la base censal, mientras que `_label` corresponde al nombre de la etiqueta de dicho valor.

Con la sentencia `by` podemos generar tabla con mayor cantidad de desagregaciones. Por ejemplo, la siguiente tabla muestra el número de individuos por provincia según edad quinquenal en la categoría de trabajo.

```
Tabla2 <- redatam.query(PAN,
  "
  freq PROVIN.PROVNOMB
  by PERSONA.EDADQUIN
  by RP17TRAB
  "
)
```

```
head(Tabla2)
```

```
##   PROVNOB1_value PROVNOB1_label EDADQUIN2_value EDADQUIN2_label
## 1 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
## 2 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
## 3 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
## 4 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
## 5 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
## 6 BOCAS DEL TORO BOCAS DEL TORO             2      10-14
##   RP17TRAB3_value RP17TRAB3_label value
## 1          1110      Ocupado, trabajÃ³ la semana de referencia 198
## 2          1211      Desocupado, trabajÃ³ antes, buscÃ³ trabajo    7
## 3          1212      Desocupado, trabajÃ³ antes, espera noticias    4
```

```
## 4      1213 Desocupado, trabajÃ³ antes, se cansÃ³ de buscar trabajo      1
## 5      1214      Desocupado, trabajÃ³ antes, ya consiguiÃ³ trabajo      1
## 6      1221      Desocupado, nunca trabajÃ³, buscÃ³ trabajo      1
```

```
names(Tabla2)
```

```
## [1] "PROVNOMB1_value" "PROVNOMB1_label" "EDADQUIN2_value" "EDADQUIN2_label"
## [5] "RP17TRAB3_value" "RP17TRAB3_label" "value"
```

```
unique(Tabla2$EDADQUIN2_label)
```

```
## [1] "10-14"      "15-19"      "20-24"      "25-29"      "30-34"      "35-39"
## [7] "40-44"      "45-49"      "50-54"      "55-59"      "60-64"      "65-69"
## [13] "70-74"      "75-79"      "80-84"      "85-89"      "90-94"      "95-98"
## [19] "99 y mÃ¡s"
```

Como se puede apreciar en la tabla y en la verificaci3n que se realiza, no hay conteos para edad quinquenal entre 0 y 9 aÃ±os, esto se debe a que para aquel rango de edad la pregunta de trabajo se omite por lo que el conteo se realiza solo sobre los casos vÃ¡lidos. Para hacer un conteo sobre casos no vÃ¡lidos (NA) se debe aÃ±adir la opci3n `tot.omit = FALSE` como se muestra en la tabla a continuaci3n.

Como se aprecia en la tabla, dicha opci3n calcula ademÃ¡s los totales para cada una de las desagregaciones realizadas. En el caso de las etiquetas `_label` los totales y casos no vÃ¡lidos son etiquetados como `__tot__` y `__na__` respectivamente mientras que la etiqueta `_value` contendrÃ¡ NA. Finalmente, la tabla aÃ±ade una etiqueta `_mask` a cada una de las desagregaciones la cual se categoriza como 1 cuando se realiza un conteo sobre el total de dicha variable, 2 cuando se realiza un conteo sobre casos no vÃ¡lidos de la variable y 0 en otro caso. Para descartar los totales y observar los conteos sobre las desagregaciones se puede filtrar mediante `_mask` o `_label` cada una de las variables escogidas como se muestra en la siguiente sintaxis. note que a `_label` en cada variable se antepone un nÃºmero, dicho nÃºmero corresponde al orden en el cual fueron ingresadas las variables.

```
Tabla3 <- redatam.query(PAN,
  "
  freq PROVIN.PROVNOMB
  by PERSONA.EDADQUIN
  by RP17TRAB
  ",
  tot.omit = FALSE) %>%
  filter(PROVNOMB1_label != "__tot__",
    EDADQUIN2_label != "__tot__",
    RP17TRAB3_label != "__tot__")
```

```
names(Tabla3)
```

```
## [1] "PROVNOMB1_value" "PROVNOMB1_label" "PROVNOMB1_mask" "EDADQUIN2_value"
## [5] "EDADQUIN2_label" "EDADQUIN2_mask" "RP17TRAB3_value" "RP17TRAB3_label"
## [9] "RP17TRAB3_mask" "value"
```

```
head(Tabla3)
```

```
## PROVINOMB1_value PROVINOMB1_label PROVINOMB1_mask EDADQUIN2_value
## 1 BOCAS DEL TORO BOCAS DEL TORO 0 0
## 2 BOCAS DEL TORO BOCAS DEL TORO 0 1
## 3 BOCAS DEL TORO BOCAS DEL TORO 0 2
## 4 BOCAS DEL TORO BOCAS DEL TORO 0 2
## 5 BOCAS DEL TORO BOCAS DEL TORO 0 2
## 6 BOCAS DEL TORO BOCAS DEL TORO 0 2
## EDADQUIN2_label EDADQUIN2_mask RP17TRAB3_value
```

```
## 1      0-4      0      NA
## 2      5-9      0      NA
## 3     10-14     0      NA
## 4     10-14     0     1110
## 5     10-14     0     1211
## 6     10-14     0     1212
##              RP17TRAB3_label RP17TRAB3_mask value
## 1              __na__                2 17884
## 2              __na__                2 17673
## 3              __na__                2   10
## 4  Ocupado, trabajÃ³ la semana de referencia      0   198
## 5  Desocupado, trabajÃ³ antes, buscÃ³ trabajo      0    7
## 6  Desocupado, trabajÃ³ antes, espera noticias      0    4
```

Ejemplo: Cálculo de la tasa de ocupación por provincia

Utilizando la variable de trabajo RP17TRAB contenida en la entidad de PERSONA se calcula la tabla correspondiente.

```
Tabla4 <- redatam.query(PAN,
  "
  freq PROVIN.PROVNOMB
  by PERSONA.RP17TRAB
  "
)
```

```
##
## Redatam process: 1% Redatam process: 2% Redatam process: 3% Redatam process: 3% Redatam process: 4%
## tot outputs: 1
## cols: 7
## tot : 226
```

```
head(Tabla4)
```

```
##  PROVINOMB1_value PROVINOMB1_label RP17TRAB2_value
## 1  BOCAS DEL TORO  BOCAS DEL TORO      1110
## 2  BOCAS DEL TORO  BOCAS DEL TORO      1120
## 3  BOCAS DEL TORO  BOCAS DEL TORO      1211
## 4  BOCAS DEL TORO  BOCAS DEL TORO      1212
## 5  BOCAS DEL TORO  BOCAS DEL TORO      1213
## 6  BOCAS DEL TORO  BOCAS DEL TORO      1214
##              RP17TRAB2_label value
## 1      Ocupado, trabajÃ³ la semana de referencia 32838
## 2      Ocupado, trabajador ocasional      220
## 3      Desocupado, trabajÃ³ antes, buscÃ³ trabajo 1237
## 4      Desocupado, trabajÃ³ antes, espera noticias 513
## 5  Desocupado, trabajÃ³ antes, se cansÃ³ de buscar trabajo 501
## 6      Desocupado, trabajÃ³ antes, ya consiguiÃ³ trabajo 425
```

```
cbind(
  unique(Tabla4$RP17TRAB2_value),
  unique(Tabla4$RP17TRAB2_label))
```

```
##      [,1] [,2]
## [1,] "1110" "Ocupado, trabajÃ³ la semana de referencia"
## [2,] "1120" "Ocupado, trabajador ocasional"
```

```
## [3,] "1211" "Desocupado, trabajÃ³ antes, buscÃ³ trabajo"
## [4,] "1212" "Desocupado, trabajÃ³ antes, espera noticias"
## [5,] "1213" "Desocupado, trabajÃ³ antes, se cansÃ³ de buscar trabajo"
## [6,] "1214" "Desocupado, trabajÃ³ antes, ya consiguiÃ³ trabajo"
## [7,] "1221" "Desocupado, nunca trabajÃ³, buscÃ³ trabajo"
## [8,] "1222" "Desocupado, nunca trabajÃ³, espera noticias"
## [9,] "1223" "Desocupado, nunca trabajÃ³, se cansÃ³ de buscar trabajo"
## [10,] "1224" "Desocupado, nunca trabajÃ³, ya consiguiÃ³ trabajo"
## [11,] "2100" "Jubilado(a) o pensionado(a) por vejez"
## [12,] "2200" "Pensionado(a) (por accidente o enfermedad)"
## [13,] "2300" "Estudiante solamente"
## [14,] "2400" "Ama de casa solamente o trabajador del hogar"
## [15,] "2500" "Incapacitado(a) permanentemente para trabajar"
## [16,] "2600" "Edad avanzada (70 aÃ±os y mÃ¡s)"
## [17,] "2700" "Otros(as) inactivos(as)"
```

De dicha tabla es posible identificar mediante `_label` aquellas etiquetas `_value` que corresponden a ocupados y desocupados. Con dicha informacion se generan nuevas columnas para identificar mediante variables dummies a la poblacion ocupada y desocupada, se renombra ademas la etiqueta de provincia.

```
Tabla5 <- Tabla4 %>%
  transmute(
    provin = PROVNOMB1_value,
    ocupados = ifelse(
      RP17TRAB2_value %in% c(1110,1120),
      1,
      0),
    desocupados = ifelse(
      RP17TRAB2_value %in% c(1211:1224),
      1,
      0),
    value
  )
```

```
head(Tabla5)
```

```
##      provin ocupados desocupados value
## 1 BOCAS DEL TORO      1           0 32838
## 2 BOCAS DEL TORO      1           0   220
## 3 BOCAS DEL TORO      0           1  1237
## 4 BOCAS DEL TORO      0           1   513
## 5 BOCAS DEL TORO      0           1   501
## 6 BOCAS DEL TORO      0           1   425
```

Con dichas nuevas columnas podemos agrupar la data para sumar la columna `value` que corresponde al conteo de los individuos en cada categoria.

```
Tabla6 <- Tabla5 %>%
  group_by(provin, ocupados, desocupados) %>%
  summarise(value = sum(value)) %>%
  as.data.frame()
```

```
head(Tabla6)
```

```
##      provin ocupados desocupados value
## 1 BOCAS DEL TORO      0           0 52040
## 2 BOCAS DEL TORO      0           1  3989
```

```
## 3 BOCAS DEL TORO      1      0 33058
## 4      CHIRIQUÍ      0      0 176459
## 5      CHIRIQUÍ      0      1 12988
## 6      CHIRIQUÍ      1      0 147147
```

Note que al ser categorías disjuntas, la variable ocupado y desocupados no deben tomar el valor 1 de manera simultánea. Esto es posible comprobarlo en la siguiente sintaxis.

```
Tabla6 %>%
  filter(ocupados == 1, desocupados == 1)
```

```
## [1] provin      ocupados      desocupados value
## <0 rows> (or 0-length row.names)
```

Utilizando la función `pivot_wider` de `tidyverse` pasamos las filas correspondientes a columnas de acuerdo a la llave `value` para realizar los conteos correspondientes.

```
Tabla7 <- pivot_wider(Tabla6,
  names_from = c("ocupados", "desocupados"),
  values_from = value,
  names_prefix = c("ocupados")) %>%
  as.data.frame()
head(Tabla7)
```

```
##      provin ocupados0_0 ocupados0_1 ocupados1_0
## 1  BOCAS DEL TORO      52040      3989      33058
## 2    CHIRIQUÍ      176459      12988      147147
## 3      COCLÉ      101052      6080      80603
## 4      COLÓN      90865      9648      90769
## 5  COMARCA EMBERÁ      3872      80      3042
## 6  COMARCA KUNA YALA      14436      444      8831
```

Finalmente podemos calcular la tasa de ocupación por provincia ponderando las respectivas columnas de la tabla calculada anteriormente donde el primer dígito corresponde a 1 si está ocupado y el segundo dígito a 1 si está desocupado.

```
Consulta = Tabla7 %>%
  transmute(provin,
    tasa_desocupacion =
      ocupados1_0/sum(ocupados0_1 + ocupados1_0 ))
head(Consulta)
```

```
##      provin tasa_desocupacion
## 1  BOCAS DEL TORO      0.023404772
## 2    CHIRIQUÍ      0.104178776
## 3      COCLÉ      0.057066212
## 4      COLÓN      0.064263650
## 5  COMARCA EMBERÁ      0.002153709
## 6  COMARCA KUNA YALA      0.006252270
```