

# Práctica 3: Parte 3 (ejemplos útiles)

## Tabla de contenidos

<b>1</b>	<b>Agrupar en intervalos de edad a partir de edades simples <code>case_when()</code></b>	<b>1</b>
1.1	Caso: estudio de “Cantabria en 2016” considerando “ambos sexos” . . . . .	2
1.2	Caso: estudio de “Cantabria en 2016” considerando únicamente a los “niños” .	5
1.3	Caso: estudio de “Cantabria en 2016” considerando únicamente a las “niñas” .	6

```
# source("FuncionesDemografia/funciones_px.R")
# source("FuncionesDemografia/funciones_tablas.R")
# library(tidyverse)
library(DemographyBasic)
```

En este documento, para presentar la información en forma de tablas se utilizará la función R creada: `func_salida_tablas()`.

## 1 Agrupar en intervalos de edad a partir de edades simples `case_when()`

Cargamos los datos almacenados en el fichero “RData”: “`DatosINE_PLC.RData`”, con la función R: `load()`:

```
load("DatosINE_PLC.RData")
```

También se pueden cargar ficheros “.px” del INE (y también del IECA) con el paquete “pxR”. Pero en este caso, se ha creado una función específica `read_px2()` (en el fichero: “`funciones_px.R`”) para que las columnas de tipo “factor” sean convertidas a tipo “character”.

```
Nacimientos_CCAA_EdadMadre_px = DemBas_read_px("02004.px")
head(Nacimientos_CCAA_EdadMadre_px) |>
  DemBas_presentadf(fuentesize = 9)
```

Sexo.del.nacido	Estado.civil.de.la.madre	Edad.de.la.madre	Comunidad.Autónoma	value
Ambos sexos	Todos los nacimientos	Todas las edades	Total	393181
Hombres	Todos los nacimientos	Todas las edades	Total	202478
Mujeres	Todos los nacimientos	Todas las edades	Total	190703
Ambos sexos	De madre casada	Todas las edades	Total	209461
Hombres	De madre casada	Todas las edades	Total	107950
Mujeres	De madre casada	Todas las edades	Total	101511

## 1.1 Caso: estudio de “Cantabria en 2016” considerando “ambos sexos”

Se tiene información de todas las CCAA y también de varios años de calendario. Nuestro objetivo es obtener la información exclusivamente para **Cantabria y 2016**.

**Nota:** este procedimiento se podría generalizar para cualquier comunidad autónoma y año.

```
Gano=2016
GCCAA="Cantabria"
nLxAmbos=TRUE
```

Con la función `unique()` obtenemos los valores diferentes (distintos o únicos) que aparecen en la columna CCAA y en la columna Calendario de `Nacimientos_CCAA_EdadMadre_2010a2016` (en “DatosINE\_PLC.RData”):

```
CCAAdif = unique(Nacimientos_CCAA_EdadMadre_2010a2016$CCAA)
cualCCAA = which(CCAAdif==GCCAA)
(anodif = unique(Nacimientos_CCAA_EdadMadre_2010a2016$Calendario))
```

```
[1] 2016 2015 2014 2013 2012 2011 2010
```

```
cualano = which(anodif==Gano)
```

También se ha obtenido el índice correspondiente a Cantabria y 2016: `cualCCAA` y `cualano`.

Se dispone de datos con edades simples (algunas vienen agrupadas) y además en formato de texto, como puede verse a continuación:

```
unique(Nacimientos_CCAA_EdadMadre_px$Edad.de.la.madre) %>%
  head(20)
```

```
[1] "Todas las edades" "Menos de 15 años" "De 15 años"      "De 16 años"
[5] "De 17 años"      "De 18 años"      "De 19 años"      "De 20 años"
[9] "De 21 años"      "De 22 años"      "De 23 años"      "De 24 años"
[13] "De 25 años"      "De 26 años"      "De 27 años"      "De 28 años"
[17] "De 29 años"      "De 30 años"      "De 31 años"      "De 32 años"
```

Queremos agrupar las edades simples, en grupos de edades. Para ello intentaremos construir una nueva variable o columna: “GEdades”, con ayuda de `mutate()`, y la función del paquete “dplyr”: `case_when()` como veremos en el siguiente ejemplo.

```
Nacimientos_CCAA_EdadMadre_px %>%
  dplyr::filter(Comunidad.Autónoma==CCAAdif[cualCCAA],
    Estado.civil.de.la.madre=="Todos los nacimientos",
    Sexo.del.nacido=="Ambos sexos") %>%
  dplyr::rename(Edades = Edad.de.la.madre) %>%
  dplyr::mutate(
    GEdades = dplyr::case_when(
      ((Edades >= "De 15 años") & (Edades <= "De 19 años")) ~ "15-19",
      ((Edades >= "De 20 años") & (Edades <= "De 24 años")) ~ "20-24",
      ((Edades >= "De 25 años") & (Edades <= "De 29 años")) ~ "25-29",
      ((Edades >= "De 30 años") & (Edades <= "De 34 años")) ~ "30-34",
      ((Edades >= "De 35 años") & (Edades <= "De 39 años")) ~ "35-39",
      ((Edades >= "De 40 años") & (Edades <= "De 44 años")) ~ "40-44",
      ((Edades >= "De 45 años") & (Edades <= "De 49 años")) ~ "45-49"
    )
  ) %>%
  head(20) |>
  DemBas_presentadf(fuentesize = 7)
```

Sexo.del.nacido	Estado.civil.de.la.madre	Edades	Comunidad.Autónoma	value	GEdades
Ambos sexos	Todos los nacimientos	Todas las edades	Cantabria	4118	NA
Ambos sexos	Todos los nacimientos	Menos de 15 años	Cantabria	NA	NA
Ambos sexos	Todos los nacimientos	De 15 años	Cantabria	2	15-19
Ambos sexos	Todos los nacimientos	De 16 años	Cantabria	5	15-19
Ambos sexos	Todos los nacimientos	De 17 años	Cantabria	8	15-19
Ambos sexos	Todos los nacimientos	De 18 años	Cantabria	20	15-19
Ambos sexos	Todos los nacimientos	De 19 años	Cantabria	22	15-19

(continúa)

Sexo.del.nacido	Estado.civil.de.la.madre	Edades	Comunidad.Autónoma	value	GEdades
Ambos sexos	Todos los nacimientos	De 20 años	Cantabria	33	20-24
Ambos sexos	Todos los nacimientos	De 21 años	Cantabria	28	20-24
Ambos sexos	Todos los nacimientos	De 22 años	Cantabria	40	20-24
Ambos sexos	Todos los nacimientos	De 23 años	Cantabria	41	20-24
Ambos sexos	Todos los nacimientos	De 24 años	Cantabria	53	20-24
Ambos sexos	Todos los nacimientos	De 25 años	Cantabria	61	25-29
Ambos sexos	Todos los nacimientos	De 26 años	Cantabria	92	25-29
Ambos sexos	Todos los nacimientos	De 27 años	Cantabria	120	25-29
Ambos sexos	Todos los nacimientos	De 28 años	Cantabria	169	25-29
Ambos sexos	Todos los nacimientos	De 29 años	Cantabria	173	25-29
Ambos sexos	Todos los nacimientos	De 30 años	Cantabria	227	30-34
Ambos sexos	Todos los nacimientos	De 31 años	Cantabria	272	30-34
Ambos sexos	Todos los nacimientos	De 32 años	Cantabria	295	30-34

Hay filas que no nos interesan, ya que aparecen en la nueva variable “GEdades” con el valor NA.

**Nota:** Las miniherramientas del sistema tidyverse se pueden repetir en una misma construcción todas las veces que se necesiten.

Por tanto, además de eliminar esas observaciones, necesitamos resumir esos datos al agrupar para los valores distintos de la nueva variable “GEdades”, de la siguiente forma:

```
tmp2a = Nacimientos_CCAA_EdadMadre_px %>%
  dplyr::filter(Comunidad.Autónoma==CCAAdif[cualCCAA],
    Estado.civil.de.la.madre=="Todos los nacimientos",
    Sexo.del.nacido=="Ambos sexos") %>%
  dplyr::mutate(Edades = Edad.de.la.madre) %>%
  dplyr::mutate(
    GEdades = case_when(
      ((Edades >= "De 15 años") & (Edades <= "De 19 años")) ~ "15-19",
      ((Edades >= "De 20 años") & (Edades <= "De 24 años")) ~ "20-24",
      ((Edades >= "De 25 años") & (Edades <= "De 29 años")) ~ "25-29",
      ((Edades >= "De 30 años") & (Edades <= "De 34 años")) ~ "30-34",
      ((Edades >= "De 35 años") & (Edades <= "De 39 años")) ~ "35-39",
      ((Edades >= "De 40 años") & (Edades <= "De 44 años")) ~ "40-44",
      ((Edades >= "De 45 años") & (Edades <= "De 49 años")) ~ "45-49"
    )
  ) %>%
  dplyr::filter(!is.na(GEdades)) %>%
  dplyr::select(GEdades,value) %>%
  dplyr::group_by(GEdades) %>%
  dplyr::summarise(
    CCAA = CCAAdif[cualCCAA],
```

```

    Ano = anodif[cualano],
    NacimientosAmbos = sum(value, na.rm = T)
  )

```

De esta forma hemos obtenido el número de nacimientos de **ambos sexos**, que hay en Cantabria durante el 2016, por grupos de Edad de la madre:

```
DemBas_presentadf(tmp2a)
```

GEdades	CCAA	Ano	NacimientosAmbos
15-19	Cantabria	2016	57
20-24	Cantabria	2016	195
25-29	Cantabria	2016	615
30-34	Cantabria	2016	1486
35-39	Cantabria	2016	1382
40-44	Cantabria	2016	361
45-49	Cantabria	2016	20

## 1.2 Caso: estudio de “Cantabria en 2016” considerando únicamente a los “niños”

Ahora hacemos operaciones parecidas para obtener el número de nacimientos de **niños**, que hay en Cantabria durante el 2016, por grupos de Edad de la madre:

```

tmp2Hombres = Nacimientos_CCAA_EdadMadre_px %>%
  filter(Comunidad.Autónoma==CCAAdif[cualCCAA],
         Estado.civil.de.la.madre=="Todos los nacimientos",
         Sexo.del.nacido=="Hombres") %>%
  mutate(Edades = Edad.de.la.madre) %>%
  mutate(
    GEdades = case_when(
      ((Edades >= "De 15 años") & (Edades <= "De 19 años")) ~ "15-19",
      ((Edades >= "De 20 años") & (Edades <= "De 24 años")) ~ "20-24",
      ((Edades >= "De 25 años") & (Edades <= "De 29 años")) ~ "25-29",
      ((Edades >= "De 30 años") & (Edades <= "De 34 años")) ~ "30-34",
      ((Edades >= "De 35 años") & (Edades <= "De 39 años")) ~ "35-39",
      ((Edades >= "De 40 años") & (Edades <= "De 44 años")) ~ "40-44",
      ((Edades >= "De 45 años") & (Edades <= "De 49 años")) ~ "45-49"
    )
  )

```

```

) %>%
filter(!is.na(GEdades)) %>%
select(GEdades,value) %>%
group_by(GEdades) %>%
dplyr::summarise(
  CCAA = CCAAdif[cualCCAA],
  Ano = anodif[cualano],
  NacimientosHombres = sum(value,na.rm = T)
)

```

```
DemBas_presentadf(tmp2Hombres)
```

GEdades	CCAA	Ano	NacimientosHombres
15-19	Cantabria	2016	27
20-24	Cantabria	2016	89
25-29	Cantabria	2016	332
30-34	Cantabria	2016	750
35-39	Cantabria	2016	723
40-44	Cantabria	2016	186
45-49	Cantabria	2016	12

### 1.3 Caso: estudio de “Cantabria en 2016” considerando únicamente a las “niñas”

Y repetimos las mismas operaciones parecidas para obtener el número de nacimientos de **niñas**, que hay en Cantabria durante el 2016, por grupos de Edad de la madre:

```

tmp2Mujeres = Nacimientos_CCAA_EdadMadre_px %>%
  filter(Comunidad.Autónoma==CCAAdif[cualCCAA],
         Estado.civil.de.la.madre=="Todos los nacimientos",
         Sexo.del.nacido=="Mujeres") %>%
  mutate(Edades = Edad.de.la.madre) %>%
  mutate(
    GEdades = case_when(
      ((Edades >= "De 15 años") & (Edades <= "De 19 años")) ~ "15-19",
      ((Edades >= "De 20 años") & (Edades <= "De 24 años")) ~ "20-24",
      ((Edades >= "De 25 años") & (Edades <= "De 29 años")) ~ "25-29",
      ((Edades >= "De 30 años") & (Edades <= "De 34 años")) ~ "30-34",
      ((Edades >= "De 35 años") & (Edades <= "De 39 años")) ~ "35-39",
    )
  )

```

```

      ((Edades >= "De 40 años") & (Edades <= "De 44 años")) ~ "40-44",
      ((Edades >= "De 45 años") & (Edades <= "De 49 años")) ~ "45-49"
    )
  ) %>%
  filter(!is.na(GEdades)) %>%
  select(GEdades,value) %>%
  group_by(GEdades) %>%
  dplyr::summarise(
    CCAA = CCAAdif[cualCCAA],
    Ano = anodif[cualano],
    NacimientosMujeres = sum(value,na.rm = T)
  )

```

```
DemBas_presentadf(tmp2Mujeres)
```

GEdades	CCAA	Ano	NacimientosMujeres
15-19	Cantabria	2016	30
20-24	Cantabria	2016	106
25-29	Cantabria	2016	283
30-34	Cantabria	2016	736
35-39	Cantabria	2016	659
40-44	Cantabria	2016	175
45-49	Cantabria	2016	8