

Taller Procesando datos censales en R

Taller práctico para aprender a explorar y procesar indicadores demográficos con los paquetes **ARcenso** y **tidyverse**.

Andrea Gomez Vargas

Tabla de contenidos

1	Manos al código: calculando indicador demográfico	2
1.1	Instalación de paquetes de trabajo	2
1.1.1	ARcenso	2
1.1.2	tidyverse	2
1.2	Activación de paquetes con library	2
1.3	Importo datos con la función get_census()	2
1.4	Reviso la estructura de mis datos	3
1.5	Elijo la tabla a trabajar como DataFrame	3
1.6	Calculo indicador con dplyr	4
1.7	Grafico el indicador con ggplot2	4
1.8	Gráfico estático con ggplot2	5
1.9	Tabla de resultado en gt	5
1.10	Tabla final	6

1 Manos al código: calculando indicador demográfico

Vamos a calcular la estructura de población de la provincia de Buenos Aires del año 1970.

1.1 Instalación de paquetes de trabajo

1.1.1 ARcenso

```
install.packages("remotes")  
remotes::install_github("SoyAndrea/arcenso")
```

1.1.2 tidyverse

```
install.packages("tidyverse")  
install.packages("gt")
```

1.2 Activación de paquetes con library

esto realizo cada vez que use R

Acá empezamos

```
library(arcenso) # datos censales  
library(tidyverse) # procesar la información  
library(gt) # hacer tablas
```

1.3 Importo datos con la función get_census()

```
estructura_pba_1970 <- get_census(year = 1970,  
                                  topic = "ESTRUCTURA DE POBLACION",  
                                  geolvl = "Provincia de Buenos Aires")  
  
estructura_pba_1970
```

```
$c70_buenosaires_poblacion_c1
# A tibble: 162 x 4
  sexo grupo_de_edad lugar_nacimiento poblacion
  <chr> <chr>         <chr>         <chr>
1 Total 0-4          Total          833150
2 Total 0-4          Nacidos en el país      827450
3 Total 0-4          Nacidos en el extranjero 5700
4 Total 5-9          Total          791450
5 Total 5-9          Nacidos en el país      782700
6 Total 5-9          Nacidos en el extranjero 8750
7 Total 10-14        Total          759500
8 Total 10-14        Nacidos en el país      746150
9 Total 10-14        Nacidos en el extranjero 13350
10 Total 15-19       Total          743300
# i 152 more rows
```

1.4 Reviso la estructura de mis datos

```
str(estructura_pba_1970)
```

List of 1

```
$ c70_buenosaires_poblacion_c1: tibble [162 x 4] (S3: tbl_df/tbl/data.frame)
..$ sexo          : chr [1:162] "Total" "Total" "Total" "Total" ...
..$ grupo_de_edad : chr [1:162] "0-4" "0-4" "0-4" "5-9" ...
..$ lugar_nacimiento: chr [1:162] "Total" "Nacidos en el país" "Nacidos en el extranjero" ...
..$ poblacion      : chr [1:162] "833150" "827450" "5700" "791450" ...
```

1.5 Elijo la tabla a trabajar como DataFrame

```
estructura_pba_1970 <- estructura_pba_1970[[1]]
```

```
estructura_pba_1970
```

```
# A tibble: 162 x 4
  sexo grupo_de_edad lugar_nacimiento poblacion
  <chr> <chr>         <chr>         <chr>
1 Total 0-4          Total          833150
2 Total 0-4          Nacidos en el país      827450
```

```

3 Total 0-4          Nacidos en el extranjero 5700
4 Total 5-9          Total 791450
5 Total 5-9          Nacidos en el país 782700
6 Total 5-9          Nacidos en el extranjero 8750
7 Total 10-14        Total 759500
8 Total 10-14        Nacidos en el país 746150
9 Total 10-14        Nacidos en el extranjero 13350
10 Total 15-19       Total 743300
# i 152 more rows

```

1.6 Calculo indicador con dplyr

Estructura de población por grupos quinquenales de edad según sexo

```

pba_1970 <- estructura_pba_1970 %>%
  filter(sexo != "Total") %>%
  mutate(
    grupo_de_edad = case_when(
      grupo_de_edad == "0-4" ~ "00-04",
      grupo_de_edad == "5-9" ~ "05-09",
      TRUE ~ grupo_de_edad),
    poblacion = as.numeric(poblacion)) %>%
  group_by(sexo) %>%
  mutate(poblacion_rel = if_else(sexo == "Varones", -poblacion / sum(poblacion),
                                poblacion / sum(poblacion))) %>%
  ungroup()

```

1.7 Grafico el indicador con ggplot2

```

g1 <- pba_1970 %>%
  ggplot(aes(x = poblacion_rel,
             y = grupo_de_edad,
             fill = sexo)) +
  geom_col() +
  scale_fill_manual(values = c("#003049", "#6a994e")) +
  scale_x_continuous(
    limits = c(-0.12, 0.12),
    breaks = seq(-0.12, 0.12, by = 0.03),
    labels = paste0(abs(seq(-12, 12, by = 3)), "%") ) +
  labs(

```

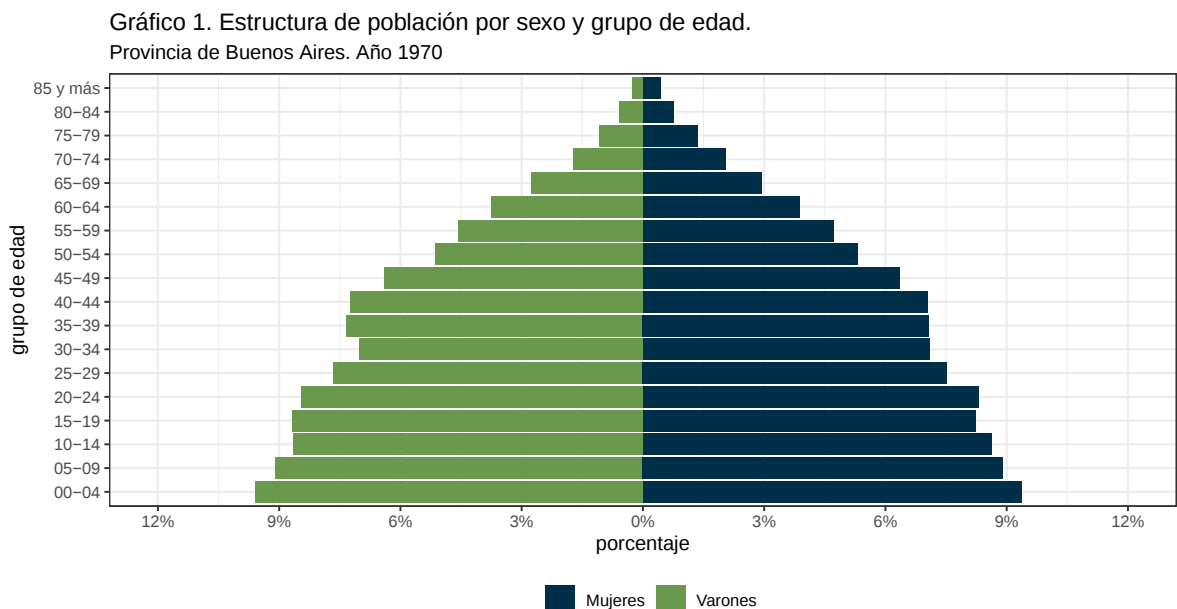
```

title = "Gráfico 1. Estructura de población por sexo y grupo de edad.",
subtitle = "Provincia de Buenos Aires. Año 1970",
x = "porcentaje",
y = "grupo de edad",
fill = " ",
caption = "Fuente: INDEC, Censo Nacional de Población, Familias y Viviendas 1970." ) +
theme_bw() +
theme(legend.position = "bottom")

```

1.8 Gráfico estático con ggplot2

```
g1 # observo el resultado
```



Fuente: INDEC, Censo Nacional de Población, Familias y Viviendas 1970.

1.9 Tabla de resultado en gt

```

tabla1 <- estructura_pba_1970 %>%
  pivot_wider(names_from = "sexo", values_from = "poblacion") %>%
  filter(lugar_nacimiento == "Total") %>%
  select(-2) %>%

```

```

mutate(Total = as.numeric(Total),
       Mujeres = as.numeric(Mujeres),
       Varones = as.numeric(Varones))

tabla_pba <- tabla1 %>%
  gt() %>%
  tab_header(
    title = "Tabla 1. Estructura de población por sexo y grupo de edad. Provincia de Buenos A
  cols_label(
    grupo_de_edad = "Grupo quinquenal de edad"
  ) %>%
  grand_summary_rows(
    columns = c(Total, Varones, Mujeres),
    fns = list("Total población" = ~sum(.))) %>%
  tab_options(
    heading.align = "left",
    heading.background.color = "#2B5597",
    footnotes.background.color = "#cedcf1",
    column_labels.font.weight = "bold"
  ) %>%
  tab_source_note('Fuente: INDEC, Censo Nacional de Población, Familias y Viviendas 1970.')

```

1.10 Tabla final

```
tabla_pba
```

Tabla 1. Estructura de población por sexo y grupo de edad. Provincia de Buenos Aires.
Año 1970

	Grupo quinquenal de edad	Total	Varones	Mujeres
	0-4	833150	422550	410600
	5-9	791450	400900	390550
	10-14	759500	381250	378250
	15-19	743300	382150	361150
	20-24	736150	372250	363900
	25-29	666950	337150	329800
	30-34	620600	309200	311400
	35-39	634350	323900	310450
	40-44	627400	318500	308900
	45-49	559700	281650	278050
	50-54	458800	225900	232900
	55-59	408350	201450	206900
	60-64	335400	165000	170400
	65-69	250250	121600	128650
	70-74	165950	75950	90000
	75-79	106450	47300	59150
	80-84	59250	25450	33800
	85 y más	31250	12050	19200
Total población	—	8788250	4404200	4384050

Fuente: INDEC, Censo Nacional de Población, Familias y Viviendas 1970.