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

# Manos al código: calculando indicador demográfico

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

## Instalación de paquetes de trabajo

### ARcenso

install.packages("remotes")
remotes::install\_github("SoyAndrea/arcenso")

### tidyverse

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

## Activación de paquetes con library

esto realizo cada vez que use R

Acá emmpezamos

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

## 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 × 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

## Reviso la estructura de mis datos

str(estructura\_pba\_1970)

List of 1
 $ c70\_buenosaires\_poblacion\_c1: tibble [162 × 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" "Total" ...
 ..$ poblacion : chr [1:162] "833150" "827450" "5700" "791450" ...

## Elijo la tabla a trabajar como DataFrame

estructura\_pba\_1970 <- estructura\_pba\_1970[[1]]

estructura\_pba\_1970

# A tibble: 162 × 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

## 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()

## 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")

## Gráfico estático con ggplot2

g1 # observo el resultado



## 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 Aires. Año 1970" ) %>%
 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.')

## Tabla final

tabla\_pba

Table 1: 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. |