

# ARCenso: a Package

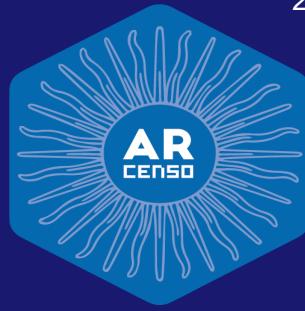


Born From Chaos,  
Powered by  
Community

Andrea Gomez Vargas & Emanuel Ciardullo  
Government Advances in Statistical Programming - GASP 2025

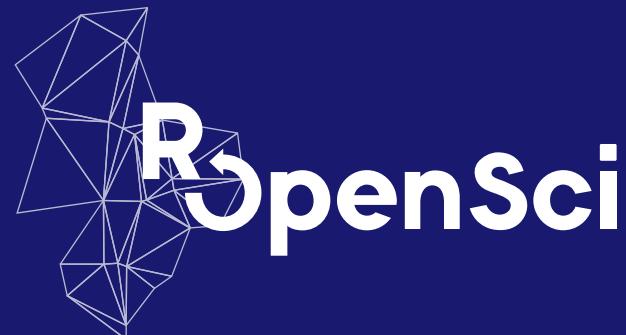
2025-06-25

# ARcenso



Supported by the rOpenSci Champions Program cohort 2023-2024<sup>1</sup>, this project is led by lead developer Andrea Gómez Vargas, co-developer Emanuel Ciardullo, and mentor Luis D. Verde.

It is a citizen-driven initiative that emerged from our professional experience working with census data at the National Institute of Statistics and Census of Argentina (INDEC).

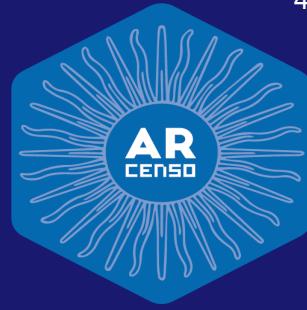


# Project Overview



**Develop an R data package that makes available the official national population census data of Argentina**, produced by the National Institute of Statistics and Censuses (INDEC), covering the period from 1970 to 2022. The data are homogenized, organized, and ready to use. The package provides open access to these datasets, facilitating their use by the public, researchers, and decision-makers.

# Problem Statement



Historical census results for **1970, 1980, 1991, 2001, 2010 and 2022** in Argentina are available in different formats through physical books, PDFs, excel files, and REDATAM outputs, without having a unified system or format that would allow working with the data from these six census periods as a database.

This fragmentation limits data accessibility, interoperability, and reuse, especially for users working within the R environment.

# Proposed Solution



From excel tables to **ordered tables** in R

|    | A   | B          | C                  | D                        | E          | F                  | G                        | H          | I                  | J                        |
|----|---|------------|--------------------|--------------------------|------------|--------------------|--------------------------|------------|--------------------|--------------------------|
| 1  | Cuadro 2. Total del país. Población total, por grupo de edad según sexo y lugar de nacimiento. Año 1970 |            |                    |                          |            |                    |                          |            |                    |                          |
| 2  |   |            |                    |                          |            |                    |                          |            |                    |                          |
| 3  | Total   |            |                    | Varones                  |            |                    | Mujeres                  |            |                    |                          |
| 4  | Grupo de edad   | Total      | Nacidos en el país | Nacidos en el extranjero | Total      | Nacidos en el país | Nacidos en el extranjero | Total      | Nacidos en el país | Nacidos en el extranjero |
| 5  | Total   | 23.390.050 | 21.179.650         | 2.210.400                | 11.617.000 | 10.455.750         | 1.161.250                | 11.773.050 | 10.723.900         | 1.049.150                |
| 6  | 0-4   | 2.355.300  | 2.339.700          | 15.600                   | 1.196.950  | 1.189.250          | 7.700                    | 1.158.350  | 1.150.450          | 7.900                    |
| 7  | 5-9   | 2.297.000  | 2.272.850          | 24.150                   | 1.163.050  | 1.149.900          | 13.150                   | 1.133.950  | 1.122.950          | 11.000                   |
| 8  | 10-14   | 2.201.150  | 2.167.750          | 33.400                   | 1.114.300  | 1.096.850          | 17.450                   | 1.086.850  | 1.070.900          | 15.950                   |
| 9  | 15-19   | 2.098.700  | 2.046.800          | 51.900                   | 1.058.850  | 1.031.150          | 27.700                   | 1.039.850  | 1.015.650          | 24.200                   |
| 10 | 20-24   | 1.950.500  | 1.815.600          | 134.900                  | 969.950    | 900.250            | 69.700                   | 980.550    | 915.350            | 65.200                   |
| 11 | 25-29   | 1.702.700  | 1.571.950          | 130.750                  | 842.550    | 773.300            | 69.250                   | 860.150    | 798.650            | 61.500                   |
| 12 | 30-34   | 1.580.350  | 1.454.500          | 125.850                  | 784.700    | 717.050            | 67.650                   | 795.650    | 737.450            | 58.200                   |
| 13 | 35-39   | 1.546.400  | 1.410.100          | 136.300                  | 779.000    | 705.150            | 73.850                   | 767.400    | 704.950            | 62.450                   |
| 14 | 40-44   | 1.539.100  | 1.376.800          | 162.300                  | 769.500    | 682.100            | 87.400                   | 769.600    | 694.700            | 74.900                   |
| 15 | 45-49   | 1.382.500  | 1.186.350          | 196.150                  | 683.550    | 581.700            | 101.850                  | 698.950    | 604.650            | 94.300                   |
| 16 | 50-54   | 1.147.100  | 1.005.600          | 141.500                  | 562.300    | 489.950            | 72.350                   | 584.800    | 515.650            | 69.150                   |
| 17 | 55-59   | 1.067.050  | 890.400            | 176.650                  | 517.800    | 424.900            | 92.900                   | 549.250    | 465.500            | 83.750                   |
| 18 | 60-64   | 890.800    | 656.700            | 234.100                  | 436.050    | 308.500            | 127.550                  | 454.750    | 348.200            | 106.550                  |
| 19 | 65-69   | 672.550    | 431.050            | 241.500                  | 322.100    | 192.150            | 129.950                  | 350.450    | 238.900            | 111.550                  |
| 20 | 70-74   | 446.450    | 273.050            | 173.400                  | 202.250    | 113.650            | 88.600                   | 244.200    | 159.400            | 84.800                   |
| 21 | 75-79   | 279.700    | 160.400            | 119.300                  | 123.150    | 61.150             | 62.000                   | 156.550    | 99.250             | 57.300                   |
| 22 | 80-84   | 151.850    | 79.250             | 72.600                   | 62.450     | 27.200             | 35.250                   | 89.400     | 52.050             | 37.350                   |
| 23 | 85 y más  | 80.850     | 40.800             | 40.050                   | 28.500     | 11.550             | 16.950                   | 52.350     | 29.250             | 23.100                   |
| 24 | Resultados obtenidos por muestra.   |            |                    |                          |            |                    |                          |            |                    |                          |
| 25 | Fuente: INDEC. Censo Nacional de Población, Familias y Vivienda 1970.                                   |            |                    |                          |            |                    |                          |            |                    |                          |

Original excel download

|    | sex   | grupo_de_edad | lugar_nacimiento         | poblacion |
|----|-------|---------------|--------------------------|-----------|
| 1  | Total | 0-4           | Total                    | 2355300   |
| 2  | Total | 0-4           | Nacidos en el pa?s       | 2339700   |
| 3  | Total | 0-4           | Nacidos en el extranjero | 15600     |
| 4  | Total | 5-9           | Total                    | 2297000   |
| 5  | Total | 5-9           | Nacidos en el pa?s       | 2272850   |
| 6  | Total | 5-9           | Nacidos en el extranjero | 24150     |
| 7  | Total | 10-14         | Total                    | 2201150   |
| 8  | Total | 10-14         | Nacidos en el pa?s       | 2167750   |
| 9  | Total | 10-14         | Nacidos en el extranjero | 33400     |
| 10 | Total | 15-19         | Total                    | 2098700   |
| 11 | Total | 15-19         | Nacidos en el pa?s       | 2046800   |
| 12 | Total | 15-19         | Nacidos en el extranjero | 51900     |
| 13 | Total | 20-24         | Total                    | 1950500   |
| 14 | Total | 20-24         | Nacidos en el pa?s       | 1815600   |
| 15 | Total | 20-24         | Nacidos en el extranjero | 134900    |
| 16 | Total | 25-29         | Total                    | 1702700   |
| 17 | Total | 25-29         | Nacidos en el pa?s       | 1571950   |
| 18 | Total | 25-29         | Nacidos en el extranjero | 130750    |
| 19 | Total | 30-34         | Total                    | 1580350   |
| 20 | Total | 30-34         | Nacidos en el pa?s       | 1454500   |
| 21 | Total | 30-34         | Nacidos en el extranjero | 125850    |
| 22 | Total | 35-39         | Total                    | 1546400   |
| 23 | Total | 35-39         | Nacidos en el pa?s       | 1410100   |
| 24 | Total | 35-39         | Nacidos en el extranjero | 136300    |

Tidy table in R



# Conceptual framework

# FAIR Principles



## **Findable:**

Centralized census data covering six national census periods (1970–2022), openly published in a single R data package with clear versioning.

## **Accessible:**

Publicly available, homogenized census datasets provided in open formats and accompanied by comprehensive documentation and metadata.

## **Interoperable:**

Tidy, well-structured tables that enable easy integration with other demographic, geographic, and socioeconomic datasets within the R ecosystem.

## **Reusable:**

Includes detailed variable descriptions, standardized coding across census years, open licensing, and reproducible data structures that facilitate long-term use and cross-study comparisons.

# Census Data: Themes and Structure (UN Framework)



## Census Topics

- **Core:** Essential variables (e.g., age, sex, population)
- **Derived-core:** Calculated variables (e.g., fertility rates)
- **Additional:** Country-specific topics (e.g., religion)

## Conceptual Units

- **Population:** Individuals
- **Housing:** Physical dwellings
- **Household:** People sharing a dwelling

## Geographic Coverage

- **National level**
- **Jurisdictional level**



# From Chaos to Package



# Problems: rain of excels and non-standardized formats





Cuadro A6. Total del país. Población de 14 años y más económicamente activa. Categoría ocupacional por sexo según edad. Año 1980

| Categoría ocupacional          | Población activa de 14 años y más | Edad           |                  |                  |                  |                  |                |
|--------------------------------|-----------------------------------|----------------|------------------|------------------|------------------|------------------|----------------|
|                                |                                   | 14-19          | 20-24            | 25-34            | 35-44            | 45-44            | 55-64          |
| <b>Total</b>                   | <b>10.033.798</b>                 | <b>996.535</b> | <b>1.415.282</b> | <b>2.674.952</b> | <b>2.106.657</b> | <b>1.728.784</b> | <b>893.151</b> |
| Empleado u obrero              | 7.147.327                         | 812.772        | 1.154.229        | 1.987.265        | 1.449.300        | 1.131.969        | 514.433        |
| Cuenta propia                  | 1.939.884                         | 68.535         | 157.903          | 479.066          | 458.551          | 421.663          | 268.999        |
| Patrón o socio                 | 575.507                           | 8.687          | 29.460           | 133.983          | 153.909          | 138.910          | 86.336         |
| Familiar sin remuneración fija | 326.472                           | 82.938         | 60.887           | 69.152           | 43.457           | 35.518           | 22.991         |
| Nuevo trabajador               | 44.608                            | 23.603         | 12.803           | 5.486            | 1.440            | 724              | 392            |
| <b>Varones</b>                 | <b>7.278.034</b>                  | <b>651.518</b> | <b>940.644</b>   | <b>1.922.139</b> | <b>1.548.901</b> | <b>1.312.658</b> | <b>725.427</b> |
| Empleado u obrero              | 4.876.634                         | 501.802        | 723.212          | 1.350.726        | 1.005.929        | 818.938          | 403.857        |
| Cuenta propia                  | 1.610.043                         | 57.199         | 132.779          | 397.417          | 376.355          | 346.344          | 226.122        |
| Patrón o socio                 | 512.748                           | 7.325          | 25.668           | 118.247          | 136.627          | 123.849          | 78.974         |
| Familiar sin remuneración fija | 250.278                           | 69.714         | 51.100           | 52.427           | 29.262           | 22.989           | 16.214         |
| Nuevo trabajador               | 28.331                            | 15.478         | 7.885            | 3.322            | 728              | 538              | 260            |
| <b>Mujeres</b>                 | <b>2.755.764</b>                  | <b>345.017</b> | <b>474.638</b>   | <b>752.813</b>   | <b>557.756</b>   | <b>416.126</b>   | <b>167.724</b> |
| Empleado u obrero              | 2.270.693                         | 310.970        | 431.017          | 636.539          | 443.371          | 313.031          | 110.576        |
| Cuenta propia                  | 329.841                           | 11.336         | 25.124           | 81.649           | 82.196           | 75.319           | 42.877         |
| Patrón o socio                 | 62.759                            | 1.362          | 3.792            | 15.736           | 17.282           | 15.061           | 7.362          |
| Familiar sin remuneración fija | 76.194                            | 13.224         | 9.787            | 16.725           | 14.195           | 12.529           | 6.777          |
| Nuevo trabajador               | 16.277                            | 8.125          | 4.918            | 2.164            | 712              | 186              | 132            |

Fuente: INDEC. Censo Nacional de Población y Vivienda 1980.



# The Process



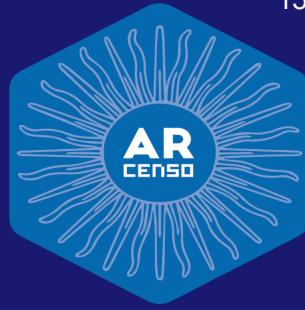
- **Download:** Automated web scraping to collect census tables from official sources.
- **Select:** Listed, classified, and extracted relevant files and metadata (census year, geography, topics).
- **Transform:** Converted Excel tables into tidy, standardized datasets using base R.
- **Function development:** Built R functions to access, manipulate, and visualize the data efficiently.
- **Package creation:** Integrated datasets and functions into the ARcenso package for easy use and reproducibility.
- **Version control:** Used Git and GitHub for tracking changes, collaboration, and release management.

# Data Availability Roadmap



| <b>Stage</b> | <b>Census years</b> | <b>Geographic level</b>       |
|--------------|---------------------|-------------------------------|
| 1            | 1970                | National and 24 jurisdictions |
|              | 1980                | National level                |
| 2            | 1991 and 2001       | National level                |
| 3            | 2010                | National level                |
| 4            | 2022                | National level                |
| 5            | 1980 and 1991       | 24 jurisdictions              |
| 6            | 2001 and 2010       | 24 jurisdictions              |
| 7            | 2022                | 24 jurisdictions              |

# {ARcenso}



## Installation

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

## Package activation

```
1 library(arcenso)
```



# get\_census()

## get tables

```

1 get_census( year = 1970,
2                         topic = "CONDICIONES HABITACIONALES",
3                         geolvl = "Total del país")
4 #> $c70_total_del_pais_poblacion_c18
5 #>             regimen_de_tenencia hogares personas cuartos
6 #> 1                 Propietario 3553250 13778700 11197900
7 #> 2                 Inquilino o arrendatario 1380950 4692800 3305350
8 #> 3 Ocupante en relación de dependencia 353300 1402500 880050
9 #> 4                 Ocupante gratuito 575650 2271150 1196500
10 #> 5                En otro carácter 192950 816350 419800
11 #>
12 #> $c70_total_del_pais_poblacion_c20
13 #>     tama?o_hogar             regimen_tenencia hogares
14 #> 1 De 1 persona                      Total 615900
15 #> 2 De 1 persona                      Propietario 255900
16 #> 3 De 1 persona                      Inquilino o arrendatario 199350
17 #> 4 De 1 persona Ocupante con relación de dependencia 52600
18 #> 5 De 1 persona                      Ocupante gratuito 82100
19 #> 6 De 1 persona                      Other 25050

```



# check\_repository()

## report of available tables

```
1 check_repository( year = 1970,  
2                      topic = "CONDICIONES HABITACIONALES",  
3                      geolvl = "Total del país")  
4 #>                                         Archivo  
5 #> 1 c70_total_del_pais_poblacion_c18  
6 #> 2 c70_total_del_pais_poblacion_c20  
7 #>  
8 #> 1 Cuadro 18. Total del país. Hogares particulares, personas y  
9 #> 2 Cuadro 20. Total del país. Hogares particulares, por tamaño de
```

# ARcenso()



## shinyapp for consulting

1 arcenso()

The screenshot shows a shiny application window titled "~/GitHub/Censuses - Shiny" at "http://127.0.0.1:7830". The header features the ARcenso logo and links for "Censo 80", "CENSO '91", "CENSO 2001", "Censo 2010", and "Censo 2022".

The left sidebar contains four dropdown menus:

- Año censal: 1970
- Alcance geográfico: Total del país
- Temática: ESTRUCTURA DE POBLACION
- Seleccione tabla: Cuadro 1. Total del país. Población total, por grupo y años simples de edad según sexo. Año 1970

The main content area displays a table titled "c70\_total\_del\_pais\_poblacion\_c1".

| grupo_de_edad | sexo    | poblacion |
|---------------|---------|-----------|
| 0-4           | Total   | 2355300   |
| 0-4           | Varones | 1196950   |
| 0-4           | Mujeres | 1158350   |
| 5-9           | Total   | 2297000   |
| 5-9           | Varones | 1163050   |
| 5-9           | Mujeres | 1133950   |
| 10-14         | Total   | 2201150   |
| 10-14         | Varones | 1114300   |
| 10-14         | Mujeres | 1086850   |
| 15-19         | Total   | 2098700   |

At the bottom, it says "1-10 of 54 rows" and includes navigation buttons for "Previous", page numbers 1 through 6, and "Next".

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

ARcenso home page

# Documentation

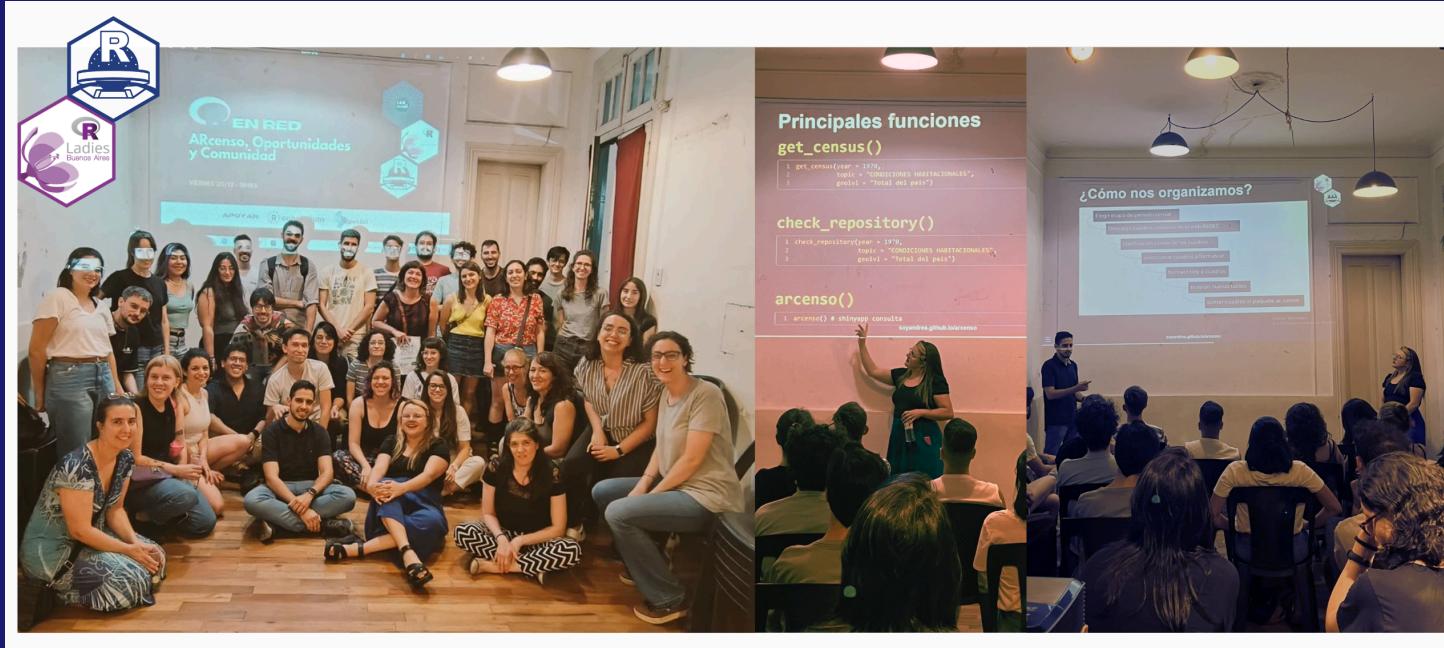




Powered by  
Community



# Thanks to Our Community



Thanks to the feedback, collaboration, and support of R communities such as rOpenSci, LatinR, and R en Buenos Aires, as well as the joint work of fellow demographers, statisticians, and sociologists, ARcenso was built.

# Community involvement provided:



- Technical assistance in solving coding and data challenges.
- Validation: Ensuring data quality and usability.
- Dissemination: Promoting the use and awareness of the package.
- Inspiration: driving continuous improvements and new features.

# Gracias 😊



**Andrea**  
Sociologist  
rOpenSci Champion



**Emanuel**  
Statistician

