**{ARcenso}: primeros pasos desarrollando un paquete en comunidad con rOpenSci**

*Andrea Gomez Vargas & Emanuel Ciardullo*

**Palabras clave:**  R, paquete, censo, Estadísticas de población, Argentina, comunidad.

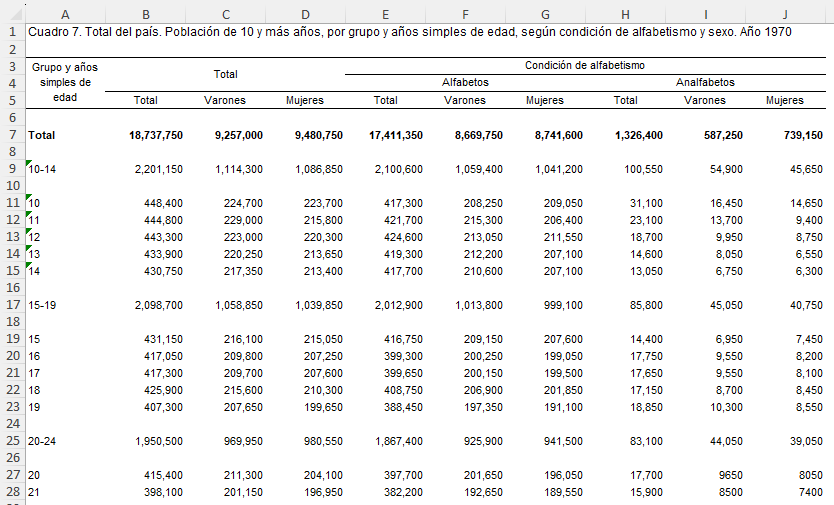
En el marco del programa de Campeones y Campeonas de rOpenSci[[1]](#footnote-0) y con el objetivo de identificar, reconocer y recompensar a los miembros apasionados de la comunidad que ayudan a la comunidad a crecer y mejorar, nace el proyecto de crear el paquete {ARcenso}.



**{ARcenso}**[[2]](#footnote-1) Es un paquete en desarrollo que permitirá acceder a los datos oficiales de los censos nacionales de población en Argentina provenientes del Instituto Nacional de Estadísticas y Censos - INDEC.

Actualmente los resultados históricos censales[[3]](#footnote-2) de 1970, 1980, 1991, 2001, 2010 y 2022 están disponibles en distintos formatos a través de libros físicos, PDFs, archivos en formato excel o en REDATAM[[4]](#footnote-3), sin contar con un sistema o formato unificado que permita trabajar con los datos de estos seis periodos censales como base de datos (Figura 1). Además, la presentación de los datos no está homogeneizada entre periodos dificultando la comparación histórica o en serie de la información disponible.

*Figura 1. Cuadro censal del año 1970 disponible para descargar en la web de INDEC.*

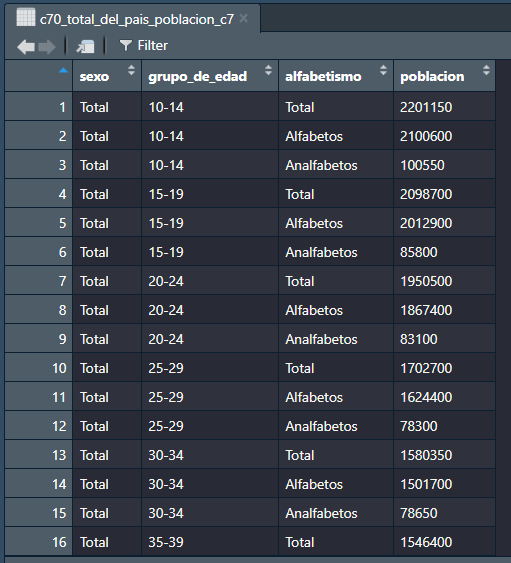


En este sentido contar con un paquete de información censal permitirá al sector público y privado, a los ciudadanos y a otros actores de la sociedad acceder a información actual e histórica sobre la población, los hogares y las viviendas de Argentina de una manera más accesible amplificando el uso de software libre.

El proyecto {ARcenso} consiste en generar un paquete que sirva de repositorio de los datos censales de estos seis periodos, ya homogeneizados en un formato de datos ordenados (tidy data) listos para su uso en R. Disponer de un paquete con estas características requiere de un armado metodológico y de diseño conceptual previo para la clasificación de los cuadros censales, lo que implica un extenso proceso 1:1 de revisión, consolidación y conversión de cuadros en excel a tabulados en R por cada indicador, por cada referencia geográfica (total del país y cada una de las 24 jurisdicciones) y por cada año censal, solo para el año 1970 se trabajaron con 357 cuadros en formato excel.

Para obtener los cuadros en este formato de datos ordenados (tidy data) (Figura 2), el paquete busca además aportar funciones elaboradas en R base para la selección de información censal, simplificar la comparabilidad y utilización conjunta de la información de los seis años censales, y de otra función que ejecute un tablero de datos soportado por los paquetes de *shiny, quarto y gt* para la exploración y consulta amigable de la información disponible.

*Figura 2. Tabulado en formato “tidy data” del cuadro censal presentado en la figura 1*



En esta presentación nos proponemos compartir la experiencia de armar un paquete desde cero acompañados por rOpenSci y toda su comunidad, con las capacitaciones, la mentoría 1:1 y la sesiones de coworking que delimitaron los puntos de partida y estrategias para el armado del proyecto, la toma de decisiones, el armado conceptual, los desafíos y aprendizajes a la hora de construir un paquete, las primeras funciones y la hoja de ruta para incorporar los seis periodos censales.

Paquete en desarrollo en <https://github.com/SoyAndrea/arcenso>

1. [Introducing rOpenSci Champions - Cohort 2023-2024](https://ropensci.org/blog/2024/02/15/champions-program-champions-2024/#andrea-gomez-vargas) [↑](#footnote-ref-0)
2. <https://github.com/SoyAndrea/arcenso> [↑](#footnote-ref-1)
3. [Censo Nacional de Población, Hogares y Viviendas](https://www.indec.gob.ar/indec/web/Nivel3-Tema-2-41) [↑](#footnote-ref-2)
4. REDATAM es un software desarrollado por la CEPAL (Comisión Económica para América Latina Y el Caribe) y utilizado ampliamente en los países de América para la difusión de estadísticas censales. [↑](#footnote-ref-3)