TITULO {ARcenso}: A Package Born From Chaos, Powered by Community

**ENFOQUE**

## parte 1: Introducción conceptual

## parte 2: el armado del paquete {arcenso}

\_\_\_\_\_\_\_\_\_\_\_\_

# parte 1:

## Slide 1: Hello

## Welcome to the presentantion {ARcenso}: A Package Born From Chaos, Powered by Community

## Slide 2:

It is a project supported by the rOpenSci Champions program cohort 2023-20241, with Andrea Gómez Vargas as lead developer, Emanuel Ciardullo as co-developer and Luis D. Verde as mentor.

## Slide 3:

The original idea behind the project was to create a package that would make available the official data of the national population censuses in Argentina from INDEC from 1970 to 2022, homogenized, ordered and ready to use.

## Slide 4:

Currently, the historical census results are available in different formats through physical books, PDFs, excel files or REDATAM, without a unified system or format that would allow working with the data from these six census periods as a database.

In addition, the presentation of the data is not homogenized between periods, making it difficult to make historical or serial comparisons of the available information.

## Slide 8:

## It is a key tool for understanding the characteristics and needs of the population.

## They provide essential data for the planning and development of public policies.

## Social and economic planning

## Academic research and social studies

## Market research and much more…

## Slide 9: working progress

## Slide 10: ¿Cómo empezar?

How to work with data elaborated from six different decades? each census year is a different work process and with the passing of the decade, there is an exponential increase in the amount of public information.

## Slide 11: problems

As the decades progressed, excel files began to be counted by the hundreds, in formats that were not standardized, i.e., we could not convert one and repeat the same process for all the others because the organization of the table changed: how it presented the information, rows, columns, etc.

## Slide 12: lets get to work

Well, as we took this whole idea to practice, to something tangible and to the R package, we used two basic tools, python to do web scraping on the INDEC web page, to download the census tables published for each census.

Python not only helped to download the information, but also helped to list the information downloaded in some cases the same excel file had two or three census tables, sometimes with an index on the first sheet, others not, that automatically could read each of the files and at least to record the title of the tables it contained helped a lot to the subsequent organization of the work.

once downloaded the files, with R we changed the original formats that were more presentation of information to a tidy table that can be used for any type of analysis and processing. So that it is well recorded what transformation was used, so that it can be replicated or improved in the future.

github helped us to organize the work, to have a version control and to be able to combine everyone's work without generating more chaos.

exponencial de cuadros se repite de un censo a otro

## Slide 13: Roadmap

We established a roadmap with steps to update the package and make the work achievable. We are currently in phase 1.

## Slide 15: Installation

You can download the package from github. Currently it has three main functions.

## Slide 16:

Get\_cesus brings to the environment all the tables that match the three arguments: year, topic and geographical level.

## Slide 17:

Check\_repository list all the tables available in the package at the time that match the three arguments: year, topic and geographical level.

## Slide 18:

A shinny app for dynamically navigating information.

## Slide 20: Comunidad