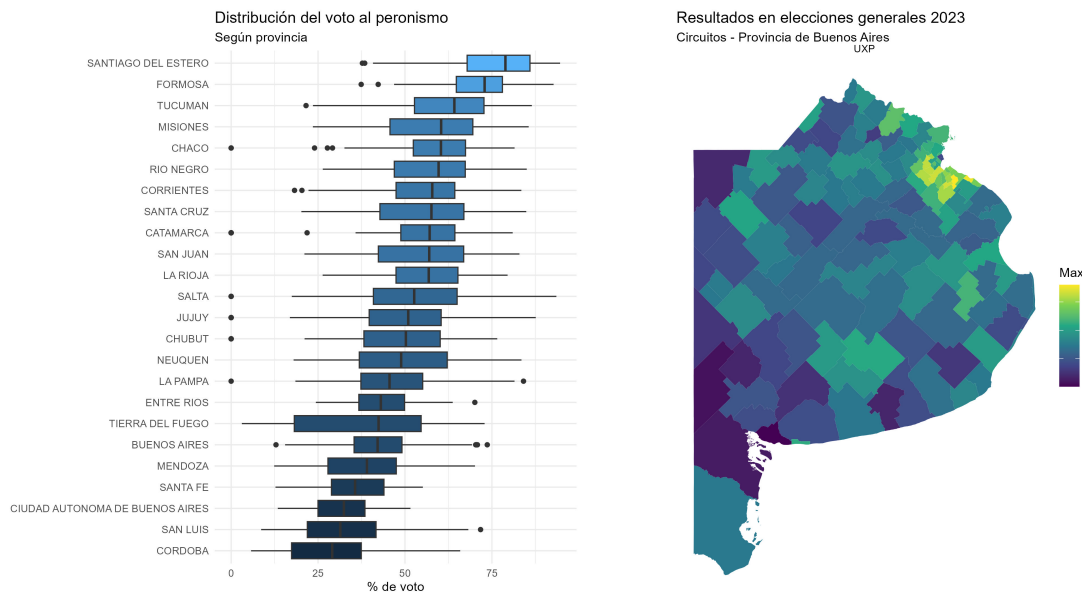


# Presentación

El objetivo de los siguientes párrafos es dar una cálida y armoniosa bienvenida al mundo de datos para personas de la carrera de Sociología de la UBA, pero extensible a distintas carreras de Ciencias Sociales a lo largo y ancho de la República Argentina.



## ¿Por qué “datos”?

Se popularizó en los últimos años el brumoso término de *Big Data*, un espacio *sexy* y confuso en donde se reunieron distintas técnicas, disciplinas y profesiones que lograban incorporar algún elemento novedoso al viejo arte trabajar con datos.

¿Qué es lo realmente interesante? Las tres famosas “V” bocetan un nuevo contexto de circulación de los datos: en grandes **volúmenes**, en tipos y formatos **variados** y con una mayor **velocidad** de intercambio y capacidad de cómputo. Ahora bien, como toda persona que quiera dedicarse a las Ciencias Sociales sabe (¡o debe saber!), poder recolectar, comprender, administrar y analizar datos para construir información es una parte fundamental del proceso de investigación. Al final, el enfoque de este espacio será cómo navegar las nuevas tempestades para ejercer el oficio que nos interpeló en un primer momento.

## ¿Por qué R?

R es un lenguaje de programación enfocado en análisis y visualización de datos. Nos agrada particularmente por ser un producto de software libre y código abierto: puede ser usado, modificado y citado sin pagar licencias de ningún tipo. Además presenta alrededor del globo -y en Argentina en particular- una gran comunidad de usuarios que crean formas específicas de aplicación de este lenguaje que nos facilitará su uso en Ciencias Sociales. Ya sea por facilitar el acceso a ciertas fuentes de información, por sintetizar procesos complejos en funciones simples de ejecutar o tan sólo por tener espacios donde realizar consultas y recibir apoyo.

Un detalle no menor: R, como herramienta, es sólo una de las tantas maneras de acceder *al mundo de datos*. La idea es que funcione como puerta de entrada al entendimiento de las lógicas de la programación y el manejo de los distintos objetos que vamos a utilizar en estos espacios.

¿Por qué utilizar lenguajes de programación y no otras alternativas? Hay otros programas y aplicaciones de renombre que cubren varias de las utilidades que le vamos a dar a R en este espacio. Sin embargo, hay dos ventajas importantes de los lenguajes de programación que es interesante destacar: primero, permiten una **integración vertical** de distintos procesos de un mismo flujo de trabajo (procesamiento, análisis, visualización, incluso la confección de un informe). Casi ninguna herramienta contempla la totalidad de las funciones antedichas y, si está entre sus bondades, probablemente sea engorroso o esté mal documentado. A lo segundo se le podría decir **integración horizontal**: los lenguajes de programación permiten trabajar sobre distintas estructuras de datos. Un mismo código puede albergar trabajo con tablas, información geográfica, texto, imágenes, audio, otros códigos (!) e incluso devolver todo lo anterior en un formato legible directamente por algún programa de renombre. Sin excusas. La principal desventaja es que se tarda más en aprender. Derribar esa última barrera es la tarea que nos encuentra aquí mismo.

## Objetivo

El objetivo general del curso es introducir en el uso del lenguaje R a través de la interfaz **Rstudio**, priorizando los paquetes incluidos en la colección **Tidyverse** y con especial foco en el proceso de extracción, limpieza, análisis y visualización de datos.

Quienes cursen podrán:

1. Comprender las características generales de los distintos tipos de datos que se pueden administrar con herramientas informáticas.
2. Conocer procesos típicos de transformación de bases de datos.
3. Representar visualmente la información recolectada, adquiriendo criterios mínimos para una correcta distribución de los elementos disponibles.

4. Aproximarse a métodos típicos de análisis de datos en Ciencias Sociales, utilizando fuentes como resultados electorales o información censal.

## Organización

El curso se estructura en cuatro partes donde se irá construyendo de manera progresiva el procesamiento, análisis y visualización de la información disponible. Los contenidos de cada una se detallan a continuación.

### 1. Introducción a R en Ciencias Sociales.

- Contextualización del uso de herramientas informáticas en Ciencias Sociales.
- Introducción de la plataforma `Rstudio`, al lenguaje R y a las formas de aprendizaje en el ámbito de la programación.
- Revisión general del proceso de análisis de datos.

### 2. Presentación de la colección Tidyverse para extracción y limpieza de datos.

- Exhibición de funciones típicas de la colección mencionada.
- Aplicación práctica de dichas funciones en bases típicas en el ámbito de las Ciencias Sociales.
- Desarrollo de tipos comunes de bases de datos: formatos *wide* y *long*.

### 3. Uso de la colección Tidyverse para análisis de datos.

- Introducción a herramientas típicas de análisis estadístico para trabajar con información cuantitativa.
- Incorporación de transformaciones complejas sobre bases de datos: agrupaciones, uniones, medidas resumen.
- Exposición de formas simples de visualización de datos.

### 4. Uso de la colección Tidyverse para visualización de datos.

- Repaso por distintos tipos de información y sus formas adecuadas de visualización.
- Presentación de criterios básicos de construcción de un gráfico.
- Aplicación práctica sobre bases típicas en el ámbito de las Ciencias Sociales.

## Agradecimientos

Gran parte de lo volcado acá se lo debo a Antonio Vázquez Brust (especialmente por su libro [Ciencia de datos para gente socaible](#)) y [Germán Rosati](#) por lo realizado desde la Universidad Nacional de San Martín. Por supuesto ninguno de ellos conoce ni tiene responsabilidad sobre el contenido de este trabajo.