

**CURSO**

# **Análisis de datos sociales con R**

**Módulo I: Introducción al análisis y visualización de datos  
con R**

**Módulo II: Visualización y comunicación con R**

---

**Septiembre 2022**

## Presentación

En la actualidad, cada vez son más las áreas en las que el análisis de datos contribuye al abordaje de preguntas de investigación y al estudio de problemáticas, tanto desde una perspectiva exploratoria, como explicativa.

En este sentido, existen distintas técnicas y herramientas posibles de ser empleadas en el análisis de datos. El presente curso se centra en el uso del lenguaje R a través del entorno de desarrollo Rstudio (recientemente renombrado Posit).

## Destinatarios

El curso se orienta a diversos perfiles que busquen incorporar conocimientos en las técnicas, herramientas y conceptos del análisis de datos mediante el lenguaje de programación R. Dado que se trabajará con datos vinculados a ámbitos de las ciencias sociales, es posible que resulte más atractivo para aquellas personas que se desempeñen académica o profesionalmente en dichas áreas; aunque la propuesta se encuentra abierta a profesionales y/o perfiles de todos los ámbitos.

No requiere de conocimientos previos en programación ni estadística.

## Objetivos

Al finalizar el curso se espera que los participantes sean capaces de:

- Trabajar de manera fluida con R y Rstudio.
- Utilizar herramientas de desarrollo colaborativo en la nube para proyectos de análisis de datos.
- Identificar fuentes de datos, extraer, transformar y desarrollar data para el abordaje de preguntas o problemas de índole social .
- Desarrollar un análisis exploratorio de los datos mediante técnicas estadísticas y herramientas introductorias de visualización.
- Emplear paquetes específicos de R para la visualización de datos cualitativos, cuantitativos y georeferenciados.

- Utilizar técnicas de minería de texto.
- Utilizar herramientas de R para la creación de documentos, de presentaciones dinámicas, de informes y/o de aplicaciones.
- Incorporar nociones básicas para el armado de una Shiny App.

## Organización del curso

**Fecha de inicio:** Sábado 24 de Septiembre

**Fecha de finalización:** sábado 12 de noviembre

**Cantidad de clases:** 8 clases de 2hs (16 horas totales)

**Horario y modalidad de cursada:** sincrónica virtual - sábados de 11 a 13hs

**Precio: \$6000**

Fecha límite de pago: 23 de septiembre - vía transferencia bancaria al Ateneo

CUIL: 30715663763

Cuenta: 000000510000031426:

CBU 0290051300000000314264

Alias: TITULO.VERSO.FIGURA

## Evaluación y certificación

El curso contará con dos instancias de evaluación, en función de los dos módulos de contenidos propuestos por el equipo docente. Quienes aprueben la primera evaluación tendrán una certificación por la realización del módulo introductorio y quienes entreguen ambos trabajos de evaluación recibirán el certificado de aprobación del curso.

## Acompañamiento y continuidad

Las personas que participen del curso serán guiadas y acompañadas durante el tiempo que dure el mismo al igual que durante el período destinado a la realización de las instancias evaluatorias. Además, quienes finalicen el curso y así lo deseen, podrán integrarse a los equipos de investigación del Laboratorio de Datos y Políticas Públicas de El Ateneo.

## Contenidos

### MODULO 1

#### **Clase 1: Introducción a la Ciencia de Datos - Programación en R**

Introducción al análisis de datos mediante R. Explicación de software libre, lenguaje de programación e IDLE. Instalación de R y RStudio. Descripción de un entorno de trabajo. Instalación e introducción a Github con el objetivo de poder crear un repositorio del trabajo en clase. Primeros pasos en la herramienta. Instalación de paquetes y operadores lógicos.

#### **Clase 2: Procesos ETL: Extract, Transform and Load**

Operaciones básicas en R. Identificación de elementos: vector, lista, dataframe, matriz, funciones. Importación y exportación de bases de datos (.csv, .xlsx, .sav, entre otros). Desarrollo de procesos de extracción, limpieza y transformación de datos con Tidyverse.

#### **Clase 3: Análisis exploratorio**

Profundización de técnicas de limpieza y transformación de datos con Tidyverse. Exploración de datos mediante técnicas estadísticas para conocer la estructura del dataset, casos atípicos, y casos perdidos o nulos. Introducción a la visualización para el análisis exploratorio (visualizaciones estáticas).

#### **Clase 4: Data viz: técnicas de visualización en R**

Desarrollo de recursos y técnicas de visualización de datos. Visualizaciones en capas y customización del gráfico. Desarrollo de gráficos interactivos en HTML. Paquetes específicos para distintos tipos de datos y casos de uso: plotly, gganimate, highcharter.

### Evaluación Módulo 1

Para la aprobación del **Módulo I** se exigirá un 75% de asistencia (3 clases) y la entrega de un trabajo final en el que se apliquen las técnicas desarrolladas durante el curso. Los participantes que realicen únicamente el Módulo I, contarán con 2 semanas para realizar el trabajo, una vez finalizada la cursada.

El trabajo deberá partir de una pregunta de investigación o situarse en un área de interés a ser indagada mediante el análisis de datos. Se espera que lxs

participantes exploren, manipulen y transformen los datos, desarrollen visualizaciones y elaboren conclusiones sobre la información procesada.

En el caso de aquellxs participantes que formen parte del Laboratorio de Datos del Ateneo, se espera que puedan abordar -junto a compañerxs del mismo equipo- temáticas vinculadas a su área de participación en el Lab.

**Fecha de presentación del trabajo final del Módulo I:** 29 de octubre

## MODULO 2

### **Clase 1: Data viz II: visualización de mapas**

Introducción a datos espaciales y uso de coordenadas geográficas. Desarrollo de mapas estáticos y dinámicos en HTML. Paquetes específicos para la creación de mapas: sf, leaflet, ggplot2.

### **Clase 2: Datos no estructurados: Análisis de texto**

Introducción al uso de datos no estructurados. Técnicas de extracción de datos textuales y representaciones gráficas. Uso de paquetes específicos para el análisis de texto (stringi, tidytext y quanteda) y desarrollo técnicas de text mining (sentiment analysis).

### **Clase 3: Datos no estructurados II: Redes y APIs**

Importación de datos desde APIs. Introducción al Social Network Analysis. Análisis de datos de Twitter: relaciones, nodos, links. Uso de paquetes específicos: ggnet2, igraph, Gephi.

### **Clase 4: Comunicar con datos**

Comunicación del análisis de datos. Elaboración de informes, customización y publicación con Rmarkdown. Introducción a Flexdashboard para el desarrollo y la publicación de tableros; y Shiny para el desarrollo y la publicación de aplicaciones web.

## Evaluación Módulo II

Para la aprobación del **Módulo II** se exigirá un 75% de asistencia (3 clases) y la entrega de un trabajo final en el que se apliquen las técnicas desarrolladas durante el curso. Los participantes que realicen el Módulo I y II deberán entregar y publicar un único trabajo final, integrando las técnicas y contenidos de ambos

módulos. Los participantes que realicen únicamente el Módulo II deberán entregar un trabajo final, pudiendo optar por desarrollar solamente los contenidos vistos en el Módulo II o incorporar los del Módulo I también. En ambos casos, contarán con 2 semanas para realizar el trabajo, una vez finalizada la cursada del segundo Módulo.

El trabajo deberá partir de una pregunta de investigación o situarse en un área de interés a ser indagada mediante el análisis de datos. Se espera que lxs participantes exploren, manipulen y transformen los datos, desarrollen visualizaciones y puedan publicar su producción. Además podrán exponer su trabajo en un encuentro extra, exclusivamente dedicado a tal fin (en el que no se abordarán contenidos, sino que se presentarán y comentarán los trabajos realizados).

En el caso de aquellxs participantes que formen parte del Laboratorio de Datos del Ateneo, se espera que puedan abordar -junto a compañerxs del mismo equipo- temáticas vinculadas a su área de participación en el Lab.

**Fecha de presentación del trabajo final del Módulo II:** 26 de noviembre.

## Acerca de El Ateneo Asociación Civil

Somos una asociación civil que tiene como objetivo general velar por el ejercicio de los derechos de los ciudadanos. Tenemos una profunda vocación social y buscamos el bien común, con especial interés por la relación entre tecnología y sociedad. Desde el año 2017 cuando nace El Ateneo, su objetivo principal ha sido construir inclusión social a través de la difusión y promoción de la tecnología, realizando talleres de robótica, actividades de alfabetización digital, capacitaciones en programación y ciencia de datos, investigación, desarrollo de soluciones digitales y apoyo a organizaciones para la transformación digital.

## Acerca del Laboratorio de Datos y Políticas Públicas

El Laboratorio de Datos y Políticas Públicas es un espacio que nuclea a distintos profesionales y estudiantes provenientes de las ciencias sociales y las ciencias de la computación, sistemas e informática, al igual que cualquier persona interesada en la ciencia de datos, la estadística y/o el diseño de políticas públicas. De modo colaborativo se trabaja en la producción de conocimiento y en la formación

continua, en equipos de investigación temáticos, con acompañamiento y mentoría.

## **Acerca del Programa de Educación, Capacitación y Formación Profesional**

Busca generar oportunidades educativas, de formación profesional y de capacitación promoviendo la inclusión digital, social, laboral y económica. Además se propone ampliar la oferta formativa del Ateneo en términos de diversidad, alcance y calidad. Para ello se propone articular con distintas organizaciones sociales, gremiales, políticas, gubernamentales y del sector privado. La oferta formativa para el 2022 será: Robótica Territorial-Taller; Robótica Territorial-Jornadas; Workshop de Robótica; Economía del Conocimiento; Ciencia de Datos y otros; Habilidades Blandas y metodologías ágiles; Capacitaciones y charlas; Políticas Públicas y formulación de proyectos; Workshop de creación de aplicaciones con App Inventor; Instalaciones inmersivas.

