



Taller introductorio para el procesamiento de datos en R

Puedes encontrar el contenido en el siguiente [repositorio](#).

Presentación

El procesamiento de datos se ha vuelto uno de los temas más importantes en la actualidad. Cada día hay más profesionales que están optando por especializarse en la gestión de información para generar nuevo conocimiento y/u orientar la toma de decisiones. En ese sentido, los datos han pasado a ser los protagonistas de nuestra época y es imprescindible desarrollar habilidades para comprender, procesar y analizar la información que se encuentra disponible.

Una de las herramientas más útiles para realizar estas tareas es **R**, un lenguaje y entorno de programación libre para el procesamiento y análisis de datos en un sinnúmero de disciplinas. La gran ventaja de esta herramienta, además de ser gratuito, es su flexibilidad y versatilidad, ya que con más de [15000 librerías disponibles](#) se puede realizar una amplia gama de tareas con datos de distinta naturaleza. No obstante, una “desventaja” de este entorno es la gran cantidad de información disponible (tutoriales, blogs, libros, etc.) lo que dificulta el **iniciarse en R**. He aquí el motivo de este taller.

Objetivo

El objetivo de este taller es introducir de manera práctica a las principales herramientas que ofrecen **R** y **RStudio** para la gestión y procesamiento de datos cuantitativos. Se espera entregar a los/as participantes una base sólida para *iniciarse en R* y que una vez finalizado este taller, puedan seguir una formación autónoma.

Contenidos

Dado que este taller no implica evaluaciones se le pedirá a los/as asistentes una activa participación en cada sesión de trabajo y en el desafío final.

Sesión 1: Exploración inicial con R y RStudio

- Instalar R y Rstudio
- Explorar las principales ventanas de trabajo
- Configurar nuestro espacio de trabajo (consultar y fijar directorios)
- Comprender la lógica de *objetos*
- Explorar vectores, listas y matrices
- Explorar funciones y argumentos
- Uso de librerías/paquetes



Sesión 2: Importar y explorar una base de datos

- Estructura de datos y variables
- Importar bases de datos en distintos formatos
- Explorar nuestra primera base de datos en R
- Análisis descriptivo para variables continuas y categóricas
- Validar la información de una base de datos
- Guardar nuestra base de datos en distintos formatos

Sesión 3: Procesamiento de datos con dplyr I

- Encadenar operaciones con %>%
- Filtrar y ordenar nuestros datos
- Seleccionar columnas
- Generar nuevas variables
- Recodificar y etiquetar variables

Sesión 4: Procesamiento de datos con dplyr II

- Unir bases de datos
- Exploración y manejo de casos perdidos
- Generar medidas de tendencia central y dispersión
- Realizar comparaciones entre grupos
- Guardar tablas de resultados en distintos formatos

Sesión 5: Asociación de variables

- Análisis de correlación para variables continuas
- Análisis de asociación para variables categóricas
- Exploración al modelo de regresión lineal simple

Sesión 6: Introducción a la visualización de datos con ggplot

- Gráficos simples con las funciones de base en R
- Primera exploración al paquete ggplot
- Ajustando parámetros en ggplot
- Gráficos descriptivos y de asociación

Sesión 7: Taller de cierre

- Corrección conjunta de desafíos
- Introducción a la reportería de datos con Rmarkdown

Será posible ajustar algunos contenidos a partir del nivel y avance de cada grupo.



Metodología de trabajo

Un buen camino para **iniciarse en este entorno de programación es a través de la práctica**. En ese sentido, el formato **taller on-line** ofrece ser un espacio de aprendizaje conjunto a través del ensayo y error, por lo que se espera trabajar en sesiones de 5 a 7 personas para aprender de forma colectiva en una video-conferencia de 1 hora y 20 minutos de duración.

¿Cómo van a funcionar las sesiones de trabajo?

Cada sesión estará orientada por objetivos, una pequeña introducción de aspectos teóricos y luego se procederá al trabajo práctico en R. Lo fundamental es que cada participante pueda ir trabajando desde su propio computador para ir resolviendo preguntas en base a sus propias necesidades.

¿A quién le podría servir este taller?

Este es un espacio destinado para personas que recién están comenzando en este lenguaje de programación o que tienen experiencias menores. Nuevamente, el objeto de este taller es entregar las herramientas básicas a una persona para que pueda continuar un aprendizaje autónomo.

¿Evaluaciones?

Este taller no cuenta con evaluaciones, sin embargo, tendrá una actividad de cierre. Al inicio del taller se le entregará a los/as participantes la posibilidad de desarrollar uno de tres desafíos a lo largo de todo el taller. En la última sesión se hará una revisión colaborativa con el fin de entregar una retroalimentación a cada participante.

Sobre el guía

José Daniel Conejeros es Licenciado en Ciencias Sociales de la Pontificia Universidad Católica de Chile con un [Major en Sociología](#) y un [Minor en Políticas Públicas](#). Actualmente es estudiante en vía de titulación del programa de [Magíster en Sociología](#) y título profesional de Sociólogo de la misma universidad. Además es estudiante del [Diplomado de Ciencia de Datos de la UC](#). Cuenta con una variada experiencia en el análisis de datos en R y ha trabajado como apoyo de docencia en varios cursos sobre análisis de datos en Ciencias Sociales y Políticas Públicas. Para mayor detalle puede consultar [AQUI](#).



Referencias

El taller utiliza dos literaturas de base y cuatro referencias on-line. Son excelentes herramientas para iniciarse en R y se complementaran con el trabajo realizado sesión a sesión. Al final de este taller se espera que los/as participantes puedan profundizar en estas referencias.

Literatura

- **R para Ciencia de Datos:** Libro base para el aprendizaje/uso de R. La plataforma de Rstudio no es solo un espacio para el análisis estadístico, sino que cuenta con excelentes herramientas para el procesamiento de datos y reportería. Puede encontrar una versión en español puede hacer click [aquí](#) (aún en desarrollo).
- **Fox, J., & Weisberg, S. (2018). An R companion to applied regression. Sage publications.:** Libro para el análisis de datos sociales enfocado al uso de modelos de regresión lineal. Herramienta útil para profundizar temas de análisis en R.

Referencias web

- **The R Manuals:** Manuales de R generados por el propio equipo de desarrolladores.
- **R-bloggers:** Plataforma con varios tutoriales (especial para resolver dudas más específicas).
- **UCLA:** Espacio para aprender estadística y programación.
- **RStudio Cheat Sheets:** Otros recursos visuales en **inglés**.

En cada sesión habrán referencias específicas y además se entregaran algunos recursos disponibles en español



Programación y valores

El taller está organizado en 7 sesiones 1 vez por semana. Cada sesión será a través de videoconferencias vía zoom con una duración de 80 minutos, las sesiones serán grabadas y estarán a disposición de los/as participantes.

INSCRIPCIONES: escribir a Constanza Lemus al correo cplemus@uc.cl

PROGRAMACIÓN DE LOS TALLERES

Cada horario tendrá una capacidad máxima de 8 personas. Cada participante se puede inscribir en **uno** de los siguientes horarios:

- Viernes de 17:00 a 18:20 hrs. **Horario sin cupos disponibles.**
- Sábados de 12:00 a 13:20 hrs.
- Sábados de 16:00 a 17:20 hrs.

LOS TALLERES INICIAN EL VIERNES 12 DE JUNIO.

VALOR

El valor de este taller será de \$70.000.- los que pueden ser pagados vía transferencia electrónica. El pago del taller garantiza la inscripción.

DONACIÓN DE RECURSOS

1/4 de todo lo recaudado en estos talleres será donado a familias de escasos recursos en la comuna de La Pintana. Está donación será realizada a través de la compra de cajas de mercadería con productos básicos y se le enviará el detalle de la donación a cada uno/a de los/as participantes una vez finalizado el taller.

Documento elaborado con las herramientas de [Rmarkdown](#)

