

Taller de iniciación extracurricular:

Procesamiento y análisis de datos: Introducción al uso de R y R-studio

Profesores: Flavio Scargiali y Leandro Colisko

Flavio Scargiali es Sociólogo por la Universidad de Buenos Aires y Maestrando en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero. Se especializa en el análisis de datos globales y *analytics*. Actualmente se desempeña como analista en Equifax S.A. Se ha desempeñado como docente en cursos de introducción al lenguaje R en diferentes universidades (UNTREF, FSOC UBA).

Leandro Colisko es Licenciado en Ciencia Política por la Universidad Nacional de La Matanza y Maestrado en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero. Cuenta con una amplia experiencia como analista de datos estadísticos. Actualmente se desempeña en el Departamento de Estadísticas en la Empresa Argentina de Navegación Aérea. Se ha desempeñado como docente en cursos de introducción al lenguaje R en diferentes universidades (UNTREF, FSOC UBA).

Presentación curso

Nuestras vidas se encuentran atravesadas por la generación de datos: votar, tomar un transporte público, responder una encuesta o el relevamiento de información para la aplicación de una política pública. En las sociedades actuales, se convierte en una necesidad el desarrollo de perfiles técnicos que conozcan herramientas para procesar, transformar, analizar y visualizar información. Organizaciones públicas y privadas están avanzando en la construcción de áreas especializadas en estas tareas. De la misma manera, en el ámbito académico emergen numerosos trabajos que recurren a estos softwares.

R es un lenguaje de código abierto que resulta amigable para quienes no tienen experiencia previa en programación. Este curso invita a los participantes a dar sus primeros pasos en el análisis de información.

Objetivos

Introducir a los participantes en el uso del lenguaje de programación R para fortalecer el perfil profesional en el procesamiento, análisis y visualización de

información.

- Acceder y trabajar con repositorios de datos cuantitativos y espaciales de distintas instituciones públicas y privadas.
- Cargar y ordenar la información a través de aprender las nociones básicas del lenguaje de programación R.
- Transformar y modelar la información y datos cuantitativos y espaciales a través de la utilización de paquetes y funciones.
- Desarrollar visualizaciones que fortalezcan el análisis y la comunicación a través de aprender la estructura básica de una representación gráfica en ggplot2 y de otras técnicas.

Destinatarios

Estudiantes de grado, posgrado, técnicos, profesionales e investigadores que deseen adquirir conocimientos teóricos y prácticos en el uso de R y Rstudio para el procesamiento, manipulación, transformación, análisis, visualización y comunicación de datos. El curso no requiere conocimientos previos en programación estadística.

Temario

UNIDAD I: Introducción a R y Rstudio

¿Por qué R? Instalación de R y Rstudio. Conociendo la interfaz de Rstudio principal entorno del desarrollo integrado (IDE). Consola, scripts de R, ambiente de trabajo. Seteando el directorio de trabajo. Conociendo los R Projects.

Operadores aritméticos, R como calculadora. Operadores lógicos. Objetos en R. Tipos de datos básicos. Clases de datos. Introducción a las funciones. Funciones estadísticas básicas: media, mediana, varianza, desvío estándar.

Aplicación: importando un conjunto de datos. Explorando un data frame. Creando nuestras primeras métricas.

UNIDAD II: Introducción a Tidyverse

Bibliotecas de R, ¿Qué es una biblioteca? Instalando una biblioteca, cargando una librería en nuestro ambiente de trabajo. Tidyverse como un paradigma de programación en R. Presentación de sus principales paquetes. Dplyr: select, filter, mutate, group_by, summarise, arrange. Tydyr: pivot_longer, pivot_wider.

Conociendo otras librerías de tidyverse: stringr, lubridate, forecast, ggplot2.

Aplicación: dos librerías de producción nacional EPH (Encuesta Permanente de Hogares y funModeling. Importando datos de EPH), hogares e individuos. Más exploración de datos con funModeling. Joins de datos con Tidyverse.

UNIDAD III: Visualizando datos estadísticos

Condicionales if, else. Iterando sobre un objeto: for y while. Conociendo ggplot2. Definiendo el conjunto de datos a graficar. Eligiendo el *geom*: columnas, barras, puntos, líneas. Definiendo la estética del gráfico: *aes*. Modificando otros componentes del gráfico, título, subtítulo, escala de colores, escalas de ejes, temas y facetado. Conociendo plotly para la creación de gráficos dinámicos.

Aplicación: graficamos un conjunto de datos (por ejemplo, el hacinamiento en los hogares).

UNIDAD IV: Mapas en R

Conociendo la librería para la carga, transformación y visualización de datos geográficos *sf*. Datos vectoriales: puntos, líneas y polígonos. Haciendo nuestros primeros mapas. Mapas dinámicos con *leaflet*.

Aplicación: graficando los resultados electorales de la provincia de Buenos Aires del 2019.

UNIDAD V: Rmarkdown

Comunicando con Rmarkdown, ¿Qué es y para qué nos puede servir?
Construir reportes en documentos HTML. Curioseando con flexDashboard.

Bibliografía principal

- Antonio Vazquez Brust (2015). *“Ciencia de datos para gente sociable: Una introducción a la exploración, análisis y visualización de datos”*. Disponible en línea:
https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/
- Garrett Golemund (2014). *“Hands-On Programming with R”*. O’reilly. Disponible en línea: <https://rstudioeducation.github.io/hopr/>
- Hadley Wickham, Garrett Golemund (2014). *“R for data science:”*

import, tidy, transform, visualize and model data". O'reilly.

Disponible en línea: <https://es.r4ds.hadley.nz/>

- Pablo Casas (2015). "*Libro vivo de Ciencia de Datos*". Disponible en línea:
- <https://librovivodecienciadedatos.ai/>

Requisitos para la cursada y aprobación

Cumplir con el 80% de asistencia a los encuentros sincrónicos La cursada se organiza en 5 unidades secuenciales a desarrollar durante 6 clases sincrónicas de dos horas de duración junto a material de lectura disponible en el aula virtual y ejercicios prácticos. Entregar un trabajo práctico final integrador, utilizando R para analizar un conjunto de datos de interés del cursante.

Organización del curso

Modalidad: online sincrónica y asincrónica.