

## PRESENTACIÓN DE PROPUESTAS

### 1. TÍTULO DEL CURSO

Taller para el procesamiento y análisis de datos: Introducción al uso de R y R-studio

### 2. DOCENTE A CARGO Y EQUIPO DOCENTE

APELLIDO Y NOMBRE	
Docente Titular	Leandro Colisko
Docente Adjunto	Flavio Scargiali

### 3. JUSTIFICACIÓN - FUNDAMENTACIÓN

Nuestras vidas se encuentran atravesadas por la generación de datos: votar, tomar un transporte público, responder una encuesta o el relevamiento de información en la aplicación de una política pública. Por este motivo, implementar políticas públicas o abordar una cuestión social es también saber procesar información para tomar decisiones y evaluar resultados. En este contexto, se convierte en una necesidad el desarrollo de perfiles que conozcan y utilicen herramientas para procesar, transformar, analizar y visualizar información.

R es un lenguaje de código abierto que resulta amigable para quienes no tienen

experiencia previa en programación. Nos brinda herramientas integradas que permiten consolidar un flujo de trabajo articulado en un mismo software.

## 4. 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.

- Introducir el uso de software libre (r) y sus ventajas para el análisis de datos.
- Generar aprendizajes en carga, procesamiento, transformación, modelación y visualización de datos a través de la utilización de paquetes y funciones.
- Desarrollar un proceso de trabajo integrado que inicie en la carga de datos y finalice en la elaboración de un informe.

## 5. PROGRAMA A DESARROLLAR

### **UNIDAD I:** Introducción al procesamiento de datos

En esta unidad realizaremos una introducción al problema de los datos y la información en la administración pública. ¿Por qué es tan importante obtener datos y producir información? Destacaremos la oportunidad que se nos presenta al utilizar herramientas gratuitas y basadas en software libre vs las herramientas propietarias. En particular, el uso de R y Rstudio para procesar datos y construir información confiable. Guiaremos a los participantes en la instalación del software, conoceremos la interfaz de estas herramientas y comenzaremos a aplicar los comandos básicos.

#### *Contenidos mínimos:*

- ¿Por qué R? Instalación de R y Rstudio. Conociendo la interfaz de Rstudio

principal entorno de desarrollo integrado (IDE). Consola, scripts de R, ambiente de trabajo. Seteando el directorio de trabajo. Conociendo los R Projects.

- Operadores aritméticos, 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.

Ejercicios de aplicación: Mostraremos ejemplos para cargar diversos conjuntos de datos, relacionados a diferentes problemáticas de interés y acción públicas, para realizar las primeras métricas.

## **UNIDAD II:** Transformación de datos con R

En esta unidad veremos cómo leer los diferentes archivos con los que habitualmente trabajamos en la administración pública (excel, csv, sav, etc). Con estos datos realizaremos diferentes transformaciones que nos permitan construir las métricas que necesitamos. Vamos a filtrar datos por alguna característica de interés, seleccionaremos las variables que nos interesan analizar, vamos a producir agrupaciones y sacar métricas por grupos (cantidad, proporciones, etc).

### *Contenidos mínimos:*

- 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.

### **UNIDAD III:** Visualizando de información mediante gráficos

Una vez que obtenemos la información que necesitamos, un gran desafío es comunicar los resultados obtenidos. Esto se torna especialmente crítico en las diferentes dimensiones de nuestra vida laboral y académica, donde necesitamos comunicar los resultados a nuestros colegas, otras organizaciones, decisores de políticas públicas y en muchas circunstancias al público general.

Por este motivo, veremos la forma de construir gráficos R. Discutiremos cómo elegir el gráfico indicado de acuerdo a la información a presentar. Veremos gráficos de barra, de columna, de líneas, boxplot. Conoceremos la principal librería de R para extender nuestras posibilidades de construcción de gráficos.

#### *Contenidos mínimos:*

- Conociendo ggplot2. Definiendo el conjunto de datos a graficar. Eligiendo la *geometría*: 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.

Ejercicios de aplicación: Construiremos gráficos para presentar la información obtenida para un conjunto de datos, que estos sean significativos y claros.

### **UNIDAD IV:** Introducción al análisis de información geográfica

En muchos casos en diferentes organizaciones públicas contamos con información que puede/debe analizarse desde una dimensión y perspectiva territorial. Las problemáticas pueden variar de provincia en provincia,

departamento a departamento dentro de una provincia y barrio a barrio en un municipio. Por lo tanto en esta unidad presentamos una serie de recursos que R nos provee para leer este tipo de información, transformarlo y visualizarlo cuando sea necesario y oportuno.

*Contenidos mínimos:*

- Introducción a SF: librería para la carga, transformación y visualización de datos geográficos. Datos vectoriales: puntos, líneas y polígonos. Haciendo nuestros primeros mapas. Mapas con *leaflet*

Ejercicios de aplicación: Desarrollo de nuestros primeros mapas en R para analizar una problema público desde una perspectiva territorial.

**UNIDAD V:** Presentación de informes y documentos con R

Comunicar es clave. Hay que comunicar para tomar decisiones. Para eso en general construimos informes, documentos o tableros, en donde insertamos gráficos, imágenes, análisis de los resultados, etc. En R contamos con herramientas que nos permiten realizar estos informes sin recurrir al Word o al Power Point. Dentro de R podemos construir flujos de trabajo que impliquen la lectura de datos, la transformación, obtención de métricas, visualizaciones y la construcción del informe. A esto llamamos flujo integrado de trabajo en R.

Ejercicios de aplicación: Sistematizar las distintas etapas del flujo de trabajo en el análisis de datos en informes. Que permitan al analista de información compartir su trabajo con otros sectores de la administración pública y con diferentes tomadores de decisión.

*Contenidos mínimos:*

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

## 6. BIBLIOGRAFÍA

Antonio Vázquez Brust (2015). "Ciencia de datos para gente sociable: Una introducción a la exploración, análisis y visualización de datos". Disponible en línea, visitado por última vez el 17/04/2022: [https://bitsandbricks.github.io/ciencia de datos gente sociable/](https://bitsandbricks.github.io/ciencia-de-datos-gente-sociable/)

Garrett Golemund (2014). "Hands-On Programming with R". O'reilly. Disponible en línea, visitado por última vez el 17/04/2022: <https://rstudio-education.github.io/hopr/>

Hadley Wickham, Garrett Golemund (2014). "R for data science: import, tidy, transform, visualize and model data". O'reilly. Disponible en línea, visitado por última vez el 17/04/2022: <https://es.r4ds.hadley.nz/>

Pablo Casas (2015). "Libro vivo de Ciencia de Datos". Disponible en línea, visitado por última vez el 17/04/2022: <https://librovivodecienciadedatos.ai/>

## 7. MODALIDAD DE DICTADO

PRESENCIAL	VIRTUAL	SEMI-PRESENCIAL
	X	

## 8. MODALIDAD DE EVALUACIÓN

Se realizará un seguimiento general de la participación de los y las participantes. Como evaluación individual, se propone un trabajo práctico final integrador que se oriente a responder a diferentes dimensiones de los temas vistos en los encuentros.

## 9. DURACIÓN

CARGA HORARIA	
CARGA HORARIA TOTAL	
20 horas	
CANTIDAD DE CLASES TOTALES	
5 clases de 4 horas cada una	
SINCRÓNICA	ASINCRÓNICA
15 horas	

### 9.1. Días y horarios tentativos de dictado

Días	Horarios
Martes	18.00 hs - 21.00 hs
Jueves	18.00 hs - 21.0 hs

## 10. REQUISITOS MÍNIMOS DE INSCRIPCIÓN

Contar con una computadora con conexión a internet (especialmente para la instalación de r y rstudio) para realizar las prácticas planteadas en el curso.



# CV ABREVIADO

Por cada docente, completar esta hoja copiando y pegando de forma continuada.

<b>APELLIDO Y NOMBRE</b>	Flavio Scargiali
<b>DNI</b>	37930891
<b>CORREO ELECTRÓNICO</b>	<a href="mailto:flavio.scargiali@gmail.com">flavio.scargiali@gmail.com</a>
<b>TELÉFONO</b>	02477 - 15675616
<b>ESTUDIOS</b>	Sociólogo de la Universidad de Buenos Aires. Maestrando en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero.
<b>EXPERIENCIA</b>	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).

<b>APELLIDO Y NOMBRE</b>	Leandro Colisko
<b>DNI</b>	
<b>CORREO ELECTRÓNICO</b>	colisko_leandro@live.com.ar
<b>TELÉFONO</b>	
<b>ESTUDIOS</b>	Licenciado en Ciencia Política Universidad Nacional de La Matanza. Maestrando en Generación y Análisis de



	Información Estadística por la Universidad Nacional de Tres de Febrero.
<b>EXPERIENCIA</b>	Cuenta con una amplia experiencia como analista de datos estadísticos. Actualmente se desempeña como ingeniero de datos en MODO. Se ha desempeñado como docente en cursos de introducción al lenguaje R en diferentes universidades (UNTREF, FSOC UBA).

## NOTAS - OBSERVACIONES

--