



MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PROPUESTA DE CURSO DE POSGRADO

1- DATOS GENERALES DE LA ACTIVIDAD CURRICULAR	
1.1 Título del Curso	Análisis de datos, Mapas web y Tableros con Lenguaje R
1.2 Área temática ¹	Ingeniería de Software; Algoritmos y Lenguajes; Bases de Datos y Sistemas de Información.

1- COMPOSICION DEL EQUIPO DOCENTE	
2.1 Responsable a cargo de la actividad curricular	Dr. Luis Reynoso
2.2 Docentes	Dr. Luis Reynoso

2- CARGA HORARIA	
Carga horaria teórica	20
Carga horaria práctica	40
Carga horaria total	60
Distribución horaria semanal	Lu Ma Mie Jue Vie 4 hs. 4 hs
Fecha de inicio sugerida	1 ^{er} ó 2 ^{do} cuatrimestre

3- BREVE RESUMEN DE CONTENIDOS (hasta 400 palabras)

¹ Corresponde a uno de los siguientes tópicos: Algoritmos y Lenguajes; Teoría de la Computación; Ingeniería de Software, Bases de Datos y Sistemas de Información; Arquitecturas, Sistemas Operativos y Redes.



La Ciencia de Datos es una disciplina emergente que permite analizar datos con el objetivo de extraer conocimiento de los mismos como así también producir un mejor entendimiento de los datos.

El objetivo de este curso en conjunto con otros impartidos en la maestría es comprender los conceptos y herramientas más importantes en permitan abordar un proyecto típico en Ciencia de Datos.

En el presente siglo se han introducido diseños curriculares de diferentes carreras destinadas a formar competencias de un científico de datos, que permitan desarrollar nuevas capacidades para describir, explotar y visualizar información existente como así también producir información agregada, combinarla o incluso predecir nueva información. El auge de la ciencia de datos se ve potenciada con el crecimiento exponencial de la información, su tratamiento, y los paradigmas de datos abiertos e infraestructura de datos espaciales.

La visualización de datos es una habilidad fundamental para cualquier persona que utilice habitualmente datos cuantitativos en su trabajo, constituyendo una herramienta que casi todos los profesionales necesitan en la actualidad. Una de las herramientas críticas para la visualización y análisis de datos en la actualidad es el lenguaje de programación R. Este lenguaje permite instalar numeroso paquetes de software como **ggplot**, **tidyverse**, **flexdashboard**, **shiny**, **leaflet** etc., convirtiéndose en una plataforma extremadamente poderosa y flexible para hacer figuras, tablas, mapas, informes y tableros de control.

El presente curso proporciona herramientas para construir diferentes tipos de gráficos. Ilustrará las funciones avanzadas de **ggplot2**. Introduce a la problemática de análisis de datos permitiendo profundizar sus conocimientos en visualización de datos, tanto estáticos como dinámicos. La visualización de datos geográficos a través de mapas web, es otro de los contenidos del curso, a través del paquete **leaflet** de R. El uso de datos geográficos es inevitable ya que se conoce que (aproximadamente) más del 80 por ciento de la información tiene una vinculación directa o indirecta con información geográfica.

Para el caso de informes, tanto el paquete **knitr** como **R Markdown**, proporcionan un marco flexible de creación de informes que permiten integrar conceptos relacionados de ciencia de datos, admitiendo docenas de formatos de salida estáticos y dinámicos, incrustando visualizaciones y tablas, y facilitando la integración de código **Latex** y de múltiples lenguajes de programación.

El curso también aborda el aprendizaje en la construcción de informes en tableros de control interactivos. En este último caso, los participantes aprenderán sobre paneles y gráficos interactivos (paquete **shiny**).

4- CONOCIMIENTOS PREVIOS REQUERIDOS

Título de grado en licenciatura en ciencias de la computación para los alumnos de la maestría en Ciencias de la Computación. También son admitidos al curso, alumnos de grado o investigadores de todas aquellas disciplinas que necesiten analizar datos y producir mapas web.

5- OBJETIVOS

El curso tiene diversos objetivos para los alumnos:

- Analizar datos existentes (ejemplo datasets de datos abiertos u otros provenientes de la actividad de investigación) empleando técnicas y paquetes que favorezcan el tratamiento, combinación, visualización y estructuración de los mismos.
- Aprender a producir informes y tableros de control a partir de los datos analizados.
- Conocer herramientas el proceso del desarrollo de un proyecto de ciencia de datos.
- Involucrar a alumnos en el uso de literatura actualizada y relevante a la temática.
- Conocer y explorar las principales herramientas del ecosistema de R y R Studio.
- Colaborar con los alumnos de la maestría en la producción de sus tesis, como así también en su formato, sistematización, documentación y redacción.



6- CONTENIDOS (organizados en unidades, ejes, módulos, otros)

MÓDULO 1: Introducción

Propedéuticos de Ciencia de Datos y Data Wrangling. Introducción y Objetos. Trabajando con Datos. Data Management con R. Entorno de Trabajo en R. Instalación de Paquetes, Funciones básicas de Cálculo. Scripts.

MÓDULO 2: Tratamiento de Datos

Operaciones sobre los datos. Calidad de Datos. Formatos. Exportarlos. Tratamiento del dato: Crear Subsets de Datos, Agregar campos calculados, combinar, mezclar y ordenar datos. Data wrangling. Paquetes. Vectores y Dataframe. Manipulando datos (filtrar, reordenar y agrupar), juntando datos (distintos joins), estructurando datos.

MÓDULO 3: Visualización con Gráficos

Visualización de Datos. Abordaje general sobre Estadística descriptiva. El paquete **ggplot2**. Creando plots con **ggplot**: Gráfico de dispersión, Gráfico de línea, Histograma, Gráfico de barras, Gráfico de caja, Gráfico de superficie. Personalizando Plots: estética, leyenda, ejes, facetas y temas.

MÓDULO 4: Informes

Producción de informes con formatos de salida estáticos y dinámicos. Aplicación con R Markdown. Fragmentos de código. El paquete **knitr**. Tablas. Tipos de salida.

MÓDULO 5: Mapas en la web

Información Geográfica y Mapas Datos Georreferenciados. Formatos de archivos. Exploración de Datos Geográficos Visualización de Datos Geográficos. Combinación de fuentes de datos. Mapas con la librería **leaflet**. Ejemplo del uso de la librería **rtweet** para obtener información de twitter.

MÓDULO 6: Tableros de Control (Dashboard)

Introducción a tableros de control (dashboard). Estructura general: encabezado, Sidebar, Cuerpo. Apariencia. Comportamiento. El paquete **shiny** para crear aplicaciones web interactivas. La estructura de una aplicación **shiny**: el lado del servidor y el lado del usuario. Comprendiendo el concepto de reactividad en **shiny**. Overview de widgets en html para tableros con el paquete **shiny**.



7- PROPUESTA DIDÁCTICA (metodología de trabajo de clases teóricas y prácticas)

Las actividades teóricas serán principalmente expositivas. En ellas también participarán los alumnos del curso presentando aplicaciones o reportes sobre distintas temáticas (en lo posible de aplicación en su tema de tesis). Se fomentará el aprendizaje colaborativo y el trabajo en grupo.

ACTIVIDADES PRÁCTICAS:

El curso contiene mucha actividad práctica en lenguaje R que el alumno deberá realizar durante las clases y en forma autónoma fuera del horario de clase. Las actividades requerirán generar plots y mapas web, reportes y tableros utilizando el lenguaje R.

MATERIALES:

El curso de doctorado dispondrá en PEDCO de apuntes sobre las unidades principales, actividades y enlaces a recursos en la web.

8- MODALIDAD DE EVALUACIÓN Y CONDICIONES DE ACREDITACIÓN²

Para la acreditación de la materia se deberá asistir al 80% de las clases. Se tomará asistencia a las mismas.

Para aprobar el cursado de la materia se deberá aprobar un examen práctico que consistirá en el desarrollo de ejercicios en lenguaje R.

La acreditación de la aprobación de la materia el alumno deberán realizar un trabajo final en el cual aplique los conocimientos obtenidos a un tema de investigación, extensión o docencia. El mismo será publicado en un repositorio de trabajos en el sitio de opendata.fi.uncoma.edu.ar

² Son condiciones mínimas para la aprobación de todos los cursos: cumplir con un mínimo del 80% de asistencia a las clases, realizar las tareas y aprobar las evaluaciones que se hayan propuesto en el programa, con una calificación no menor a 7 (puntos). El trabajo final de evaluación pautado deberá realizarse en un período no mayor a un año, o dentro de los límites que estipule la Secretaría de Investigación de la Fal.



9- BIBLIOGRAFÍA DE LECTURA OBLIGATORIA CORRESPONDIENTE A CADA UNIDAD Y GENERAL

Chambers, J. M. (2020). "S, R, and Data Science". The R Journal. 12 (1): 462-476. ISSN 2073-4859.

Chang, W. (2021) R Graphics Cookbook, 2nd edition, Disponible en: <https://r-graphics.org/>

Fox, J. ,Andersen, R. (2005). "Using the R Statistical Computing Environment to Teach Social Statistics Courses" (PDF). Department of Sociology, McMaster University. Retrieved 6 August 2018.

Li, J. (2021) [The Exploration of the Approach to Data Preparation for Chinese Text Analysis Based on R Language. Open Access Library Journal Vol.8 No.9](#), September 3, 2021

Long, J. D., Teetor, P. (2021) R Cookbook, 2nd Edition, disponible en: <https://rc2e.com/>

Paradis, E..(2002) R para Principiantes. Institut des Science de l'Evolution. Universit Montpellier II. France. Traducción Jorge A. Ahumada. RCUCH University of Hawaii USGS/National Wildlife Health Center.

R Core Team (2017) R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>

R Development Core Team (2021).disponible en <https://cran.r-project.org/manuals.html>

R Development Core Team (2021). R documentation, disponible en: <https://www.r-project.org/other-docs.html>

Vance, A. (2009). "Data Analysts Captivated by R's Power". New York Times.

Thieme, Nick (2018). "R generation". Significance. 15 (4): 14-19.

Tippmann, S. (2014). "Programming tools: Adventures with R". Nature News. 517 (7532): 109-110.

10- INFRAESTRUCTURA E INSUMOS REQUERIDOS³

El curso no tiene cupos mínimo, ni máximo.

El curso puede ser dictado en forma presencial ó virtual.

En el caso de su dictado en forma presencial se requiere de un aula con pizarrón blanco. No se requiere el uso de proyector debido a que el docente a cargo dispone de uno y puede facilitarlo para el dictado.

11- OTRA INFORMACIÓN RELEVANTE

El responsable a cargo ha obtenido el grado de doctor, y cuenta con conocimiento y experiencia en el tema del curso:

En relación a datos geográficos ha colaborado activamente con el Instituto Geográfico Nacional y es actualmente Coordinador de Academia y Ciencia en IDERA (Infraestructura de Datos Espaciales de la República Argentina).

Ha elaborado numerosos reportes y tableros en lenguaje R utilizando distintos paquetes. Algunos de los cuales están publicados en opendata.fi.uncoma.edu.ar