

1. Horario atención a estudiantes, correos electrónicos y nombres de los profesores complementarios

Clase magistral

Salón: ML-108B

Horario de clase: miércoles 17:00 – 18:20

Profesor: Eduard Fernando Martínez González

Horario de atención a estudiantes: martes 16:00 – 17:00

Lugar de atención a estudiantes: W722

2. Introducción y descripción general del curso

El curso Taller de R busca que los estudiantes adquieran una comprensión general del lenguaje R y su utilidad para el análisis de datos, la investigación económica y otras aplicaciones académicas y profesionales. El énfasis será sobre la importación, organización y manejo de bases de datos, la presentación de resultados, el uso de R para el análisis de datos espaciales y otras aplicaciones de R.

El curso no profundiza en conceptos estadísticos o econométricos; en cambio, busca que los estudiantes entiendan la estructura general de programación en R y la complementen con los conocimientos adquiridos en otras clases. Al final se espera que los estudiantes se sientan cómodos utilizando R para resolver problemas relacionados con análisis de datos; asimismo, se espera que identifiquen las diferentes aplicaciones de R y puedan continuar utilizándolo.

3. Objetivos de la materia

- Impulsar el uso de lenguajes de programación *Open Source*.
- Familiarizar a los estudiantes con el entorno de trabajo de R.
- Motivar el uso de R a partir de la presentación de casos prácticos.
- Promover la automatización de procesos usando R.
- Introducir a los estudiantes en el uso de R como un Sistema de Información Geográfica (SIG).
- Proporcionar herramientas para que los estudiantes puedan continuar de manera autónoma su aprendizaje de paquetes y aplicaciones de R.

4. Organización del curso

El curso está dividido en 4 partes. En la primera parte se .. en importar, ordenar y exportar bases de datos

Modulo 1: Introducción a R

- **Clase 1:** 22 de enero
 - ¿Qué es R? ¿Por qué R?
 - Instalación de R y R Studio.
 - Interfaz de R Studio y scripts.
 - Conceptos básicos de la sintaxis en R.
 - Tipos de datos: entero, numérico, lógico, character.
 - Funciones del paquete base: length(), rep(), seq(), as(), is().
 - Estructuras de datos: vectores, matrices, listas, arrays, data frames.

Modulo 2: importar, organizar y exportar bases de datos

- **Clase 2:** 29 de enero

- Trabajar sobre un directorio de trabajo en R
- Instalar, cargar y eliminar paquetes
- Trabajando con dataframes y listas
- Importar y exportar bases de datos desde diferentes formatos
- **Clase 3:** 5 de febrero
 - Limpieza de bases de datos con los paquetes dplyr y tidyr.
 - String functions
- **Clase 4:** 12 de febrero
 - Unir bases de datos por filas y por columnas.
 - Bases de datos *long* y *wide* (función reshape).
 - Estadísticas descriptivas de una base de datos: table(), summary() y summaryBy()
- **Clase 5:** 19 de febrero
 - Loops: while, for, repeat.
 - Controles de flujo (next, break) y condicionales (if, else, ifelse.)
- **Clase 6:** 26 de febrero
 - Socializar taller 1
 - Funciones apply, lapply, sapply y tapply.
- **Clase 7:** 4 de marzo
 - Programar funciones en R (“function”).
 - Depuración de errores en funciones: debug, undebbug, traceback.
 - Funciones de control dentro de una función: warning, stop.

Modulo 3: Visualización de las datos

- **Clase 8:** 11 de marzo

- Histogramas
- Gráficos de dispersión
- Gráficos de series de tiempo
- Gráficos de barras
- Exportar gráficos

Modulo 4: Análisis espacial de información

■ **Clase 9:** 25 de marzo

- Introducción a la información espacial
- Importar y exportar archivos espaciales
- Atributos de los archivos espaciales
- Visualizar información espacial
- Reproyectar información espacial
- Agregar información a mapas
- Visualizar y exportar información espacial

■ **Clase 10:** 1 de abril

- Calcular centroides
- Opciones para cortar entidades espaciales
- Disolver y unir entidades espaciales
- Buffer y áreas de influencia de una entidad espacial
- Calcular distancias entre entidades espaciales
- Calcular el largo de un objeto tipo linea y el área de un polígono

■ **Clase 11:** 15 de abril

- Introducción a datos raster

- Cargar, inspeccionar, extraer información y guardar raster
- Trabajar con rasters desde varios layers
- Operaciones con raster (cortar, medir distancias, calcular valores del raster para un polígono)
- Rasterizar datos vectoriales (puntos o polígonos)
- Aplicaciones con datos raster
- Mostrar algunas fuentes de datos raster para Colombia y el mundo

Modulo 5: Otras aplicaciones de R

- **Clase 12:** 22 de abril
 - Socializar taller 2
 - Regresión lineal
 - Regresión Logit, Probit y Poisson
 - Análisis de series de tiempo
- **Clase 13:** 29 de abril
 - Introducción a RMarkdown.
 - Automatización de resultados con R.
- **Clase 14:** 6 de mayo
 - Introducción a la minería de texto
 - Aplicación con los acuerdos de paz
- **Clase 15:** 13 de mayo
 - Introducción a web-scraping
 - HTML (Hyper Text Markup Language)
 - Selector Gadget, xpath y css
 - Aplicación (atributos del elemento)

- Aplicación (extraer tablas)
- Aplicación (extraer datos de lluvias)
- Aplicacion (web drive)
- **Clase 16:** 20 de mayo
 - Taller final en clase

5. Metodología

El curso se desarrollará en las salas de computadores asignados por la Universidad, sin embargo, los estudiantes podrán hacer uso de sus equipos personales siempre y cuando ello no interfiera con el normal desarrollo de la clase. 15 minutos antes de cada clase, se compartirá en la plataforma de SICUA un link con el material necesario para el desarrollo de la clase. El cual consiste de 3 carpetas, una con los datos a usar en la clase, otra con los resultados resultantes de ejecutar el script y la última con el script de la clase.

El curso tiene cinco (5) módulos, los cuales buscan iniciar al estudiante en el uso de R y presentar algunas aplicaciones, que motiven su continuo aprendizaje y utilización de R durante el curso y después de haberlo terminado.

6. Competencias

Al final del curso, los estudiantes podrán:

- Solucionar de manera autónoma las dificultades básicas que se puedan presentar durante el uso del entorno de programación de R.
- Emplear R para importar, procesar y exportar bases de datos limpias.
- Automatizar el desarrollo de tareas repetitivas.
- Emplear diversas técnicas de visualización de información.
- Usar R para el procesamiento de información espacial.

- Extraer información y construir bases de datos mediante el uso de técnicas como *web scraping* o minería de texto.

7. Criterios de evaluación (Porcentajes de cada evaluación)

Para tomar la nota del curso se llevaran a cabo 3 talleres extra-clase, los cuales serán enviados mínimo 3 días antes de la clase más cercana a la fecha de entrega y 10 días antes de la fecha de entrega. El taller será socializado en clases, pero se asignará únicamente un espacio de 20 minutos de la clase para resolver dudas relacionadas con el mismo. A cada taller se le asignará el siguiente peso dentro de la nota total del curso:

- Taller 1: 30 %
- Taller 2: 36 %
- Taller 3: 34 %

Los criterios de calificación de los talleres se definirán previamente en cada taller. Los script con la solución del taller deberán ser enviados en las fechas establecidas para su entrega al correo ef.martinezg@uniandes.edu.co. No se aceptaran talleres que sean enviados después de las fechas establecidas y no cumplir alguna de las instrucciones será penalizado con el 10 % del valor de la nota del taller.

8. Sistema de aproximación de notas definitiva

Las calificaciones definitivas de las materias serán numéricas de uno punto cinco (1.5) a cinco punto cero (5.0), en unidades y décimas.

Reclamos: ¿Cómo y en qué tiempos debe presentar un estudiante un reclamo sobre su nota en cualquier evaluación del curso? Según los artículos 62 y 63 del Reglamento general de estudiantes de pregrado, el estudiante tendrá cuatro días hábiles después de la entrega de la evaluación calificada para presentar un

reclamo. El profesor responderá al reclamo en los cinco días hábiles siguientes. Si el estudiante considera que la respuesta no concuerda con los criterios de evaluación podrá solicitar un segundo calificador al Consejo de la Facultad en los cuatro días hábiles posteriores a la recepción de la decisión del profesor.

9. Bibliografía

- Gentleman, R., Hornik, K., & Parmigiani, G. (2009). Use R!.
- Mas, Jean-Francois. (2019). Análisis espacial con R: Usa R como un sistema de información geográfica.
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). Automated data collection with R: A practical guide to web scraping and text mining. John Wiley & Sons.
- Quintana, L., & Mendoza, M. (2016). Econometría aplicada utilizando R. Ciudad de México: Universidad Nacional Autónoma de México.

Fechas importantes

Inicio de clases: 20 de enero

Fecha de entrega del 30 % de las notas: 13 de marzo

Semana de receso: 16 al 20 de marzo

Último día para solicitar retiros: 27 de marzo (6pm)

Semana santa: 6 al 10 de abril

Último día de clases: 23 de mayo

Exámenes finales: 26 de mayo al 4 de junio

Último día para subir notas finales en banner: 10 de junio