

IDENTIFICACIÓN

CURSO	:	ANÁLISIS Y MANIPULACIÓN DE DATOS EN R
TRADUCCIÓN	:	DATA ANALYSIS AND PROCESSING USING R
SIGLA	:	BIO4022
CRÉDITOS	:	10
MÓDULOS	:	4
REQUISITOS	:	NO TIENE
RESTRICCIONES	:	NO TIENE
CONECTOR	:	NO APLICA
CARÁCTER	:	OPTATIVO
TIPO	:	CÁTEDRA
CALIFICACIÓN	:	ESTÁNDAR
PALABRAS CLAVE	:	INVESTIGACIÓN REPRODUCIBLE, BASES DE DATOS, ANÁLISIS DE DATOS
NIVEL FORMATIVO	:	PREGRADO

I. DESCRIPCIÓN DEL CURSO

Este curso está enfocado en entregar los principios de investigación reproducible en R, con énfasis en la recopilación y/o lectura de datos de forma reproducible y automatizada. Para esto se trabajará con bases de datos complejos, las cuales deberán ser transformadas y organizadas para optimizar su análisis. Se generarán documentos reproducibles integrando en un documento código, bibliografía, exploración y análisis de datos. Se culminará el curso con la generación de un manuscrito, presentación y/o documento interactivo reproducible.

II. RESULTADOS DE APRENDIZAJE

- Comprender el concepto de Investigación reproducible como un forma y filosofía de investigación que permite que las investigaciones sean más ordenadas y replicables, desde la toma de datos hasta la escritura de resultados.
- Aplicar el concepto de pipeline, el cual permite generar una modularidad desde la toma de datos hasta la escritura de resultados, donde la corrección independiente de un paso tiene un efecto cascada sobre el resultado final.
- Implementar buenas prácticas de recolección y estandarización de bases de datos, con la finalidad de optimizar el análisis de datos y la revisión de estas por pares.
- Realizar análisis críticos de la naturaleza de los datos al realizar análisis exploratorios, que permitan determinar la mejor forma de comprobar hipótesis asociadas a estas bases de datos.

III. CONTENIDOS

1. Investigación reproducible. Generación de documentos Rmarkdown en R, códigos en chunk y en texto.
2. Generación de tablas en Rmd. Tidy Data, optimización de base de datos, limpieza y transformación de base de datos.
3. El tidyverso y el concepto de pipeline. Limpieza de datos complejos.
4. Visualización de datos, visualizar datos vs. visualizar modelos. Insertar gráficos con leyenda en un documento Rmd.
5. Creación de funciones propias y loops.
6. Escritura de manuscritos en R, transformación de documentos Rmd en un manuscrito y manejo de bibliografía en R.
7. Presentaciones en R y generar documentos interactivos. Transformación de datos en una presentación o en una Shiny app. Realizar una presentación o aplicación en R.

IV. ESTRATEGIAS METODOLÓGICAS

Todas las clases serán prácticas y estarán divididas en dos partes: I. Clases de principios y herramientas, donde se presentarán los principios de investigación reproducible y tidy data, junto con las herramientas actuales más utilizadas, y II. Clases aplicadas con datos proporcionados en clase o que forman parte de la librería de R.

Además, se deberán generar informes y presentaciones siguiendo los principios de investigación reproducible. A los estudiantes les será proporcionado un set de datos para trabajar o podrán hacerlo con datos propios, previa discusión de objetivos y complejidad necesaria. La evaluación final será una presentación.

V. ESTRATEGIAS EVALUATIVAS

1. Evaluación 1: Informe exploratorio de base de datos (20 puntos)
2. Evaluación 2: Presentación (20 puntos)
3. Evaluación 3: Informe final (60 puntos)

VI. BIBLIOGRAFÍA

Bibliografía Mínima

- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS computational biology*, 9(10), e1003285.
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1-23.

Bibliografía Complementaria

- Gandrud, C. (2013). *Reproducible research with R and R studio*. CRC Press.
- Stodden, V., Guo, P., & Ma, Z. (2013). *Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals*.
- Stodden, V., Leisch, F., & Peng, R. D. (Eds.). (2014). *Implementing reproducible research*. CRC Press.