



FACULTAD DE CIENCIAS BIOLÓGICAS

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

PROGRAMA DE DOCTORADO EN CIENCIAS BIOLÓGICAS

I. ANTECEDENTES GENERALES

Nombre del Curso	: ANÁLISIS Y MANIPULACIÓN DE DATOS EN R
Sigla	: BIO4022
Carácter	: Optativo
Créditos	: 5
Profesor Encargado	: Dr. Derek Corcoran
Fechas inicio y término	: 05 de Octubre al 30 de Octubre
Días	: Martes - Jueves
Horario	: 10:00 a 13:00 hrs
Semestre que se dicta	: Segundo
Sala	: Sala Z06 – Subterráneo Edificio N°210

II. REQUERIMIENTOS DE SOFTWARE

La última versión de RStudio y R, también se requiere de la instalación de miktex (windows) o mactex (mac), por último, se necesita tener instalado git bash o equivalente, para windows revisar (<https://git-for-windows.github.io/>). Si necesitan ayuda para la instalación contactarse con el profesor del curso.

III. OBJETIVOS

Aprender los principios de investigación reproducible y tidy data a través del aprendizaje de programación y uso de R. Los principios de este curso están explicados en los siguientes libros gratuitos.

- Gandrud, Christopher. Reproducible Research with R and R Studio. CRC Press, 2013. Available for free in the following URL (<https://englianhu.files.wordpress.com/2016/01/reproducible-research-with-r-and-studio-2nd-edition.pdf>)
- Stodden, Victoria, Friedrich Leisch, and Roger D. Peng, eds. Implementing reproducible research. CRC Press, 2014. Available for free in the following URL (<http://web.stanford.edu/~vcs/papers/ijclp-STODDEN-2009.pdf>)

IV. METODOLOGÍA

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 donde se trabajará con datos propios para desarrollar un documento reproducible. A los estudiantes que no cuenten con datos propios, se les será proporcionado un set de datos o se simularán dependiendo del caso.

V. EVALUACIÓN (100 puntos totales)

El curso contará con 6 tareas u objetivos cada uno con un valor de 12 puntos (para un total de 12 puntos). Con un documento o presentación final con un valor de 28 puntos.

VI. PROGRAMA DEL CURSO

Clase	Tema	Lectura	Paquetes requeridos	Actividad	Meta
1	Investigación reproducible	Peng, Roger. "Reproducible research in computational science." <i>Science</i> 334.6060 (2011): 1226-1227.	knitr, repretools, stargazer	Generar un documento Rmarkdown en R y primer repositorio de github	Al menos un código en un chunk y uno en texto, y primera tabla en Rmd y primer commit
2	Tidy Data	Tidy data by Hadley Wickham, <i>Journal of Statistical Software</i> August 2014, Volume 59, Issue 10.	dplyr, reshape2, lubridate	Hacer que tus datos sean limpios	
3	El Hadleyverso y el tidyverso		dplyr, tidyr, broom, lubridate	El flujo del tidyverso. Limpiar datos complejos	
4	Visualización de Datos	Wickham, H. (2010). A layered grammar of graphics. <i>Journal of Computational and Graphical Statistics</i> , 19(1), 3-28.	ggplot2	Visualizar tus datos vs visualizar tu modelo	Al menos un gráfico con leyenda en el documento Rmd
5	Creación de funciones propias y loops			Generar funciones propias en R y loops	

6	Escribir manuscritos en R	https://github.com/rstudio/rarticles	Rarticles	Transformar documento Rmd en un manuscrito	Agregar bibliografía al documento
7	Presentaciones en R y documentos interactivos		Shiny	Transforma tus datos en una presentación o en una Shiny app	Generar una presentación o aplicación en R
8	Presentaciones				Presentación del trabajo realizado