

## **BIO 4022. Análisis y manipulación de datos en R**

Inicio clases: martes 3 octubre 2017

Término clases: jueves 26 octubre 2017

Horario de clases: martes y jueves 10:00 – 13:00 horas

### **Instructor**

Dr. Derek Corcoran

### **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 instructor del curso

### **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://englianhui.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>)

### **Formato:**

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.

### **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.

							<b>Investigación</b>				<b>Disciplinar</b>	<b>Comunicación</b>				<b>Ética</b>
<b>Clase</b>	<b>Tema</b>	<b>Lectura</b>	<b>Paquetes requeridos</b>	<b>Actividad</b>	<b>Meta</b>	Diseña experimentos que generen la evidencia necesaria para contrastar las predicciones derivadas de las hipótesis biológicas	Registra conforme a protocolos establecidos los procedimientos usados y los datos obtenidos en la investigación propuesta, de manera que los resultados puedan ser verificados y reproducidos posteriormente	Aplica los procedimientos estadísticos pertinentes al diseño experimental propuesto para poder realizar inferencias válidas respecto a las predicciones derivadas de las hipótesis biológicas	Interpreta los resultados obtenidos dentro del marco conceptual para evaluar las predicciones establecidas	Identifica posibles fuentes de error durante el proceso de investigación para diseñar protocolos experimentales rigurosos	Discrimina en función de un problema particular los datos necesarios para su resolución	Diseña una estrategia pertinente para la solución de un problema específico aplicando principios de investigación reproducible	Constuye artículos o informes científicos susceptibles de ser publicados, para difundir las problemática, procedimientos, resultados e implicancias de sus investigaciones.	Debate conforme a formatos preestablecido los argumentos técnicos y disciplinares, los procedimientos, resultados y conclusiones de investigaciones para contrastar perspectivas o explicaciones diferentes a fenómenos científicos.	Adecua su lenguaje y formalidades a distintas audiencias/lectores y contextos de presentación para favorecer la comprensión de lo que se desea transmitir.	Expone en forma dinámica, clara, lógica y precisa los aspectos principales de investigaciones para explicarlas y difundirlas.
1	Investigación reproducible	Peng, Roger. "Reproducible research in computational science." Science 334.6060 (2011): 1226-1227.	knitr, represtools, 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	x										
2	Tidy Data	Tidy data by Hadley Wickham, Journal of Statistical Software August 2014, Volume 59, Issue 10.	dplyr, reshape2, lubridate	Hacer que tus datos sean limpios		x				x	x					

3	El Hadleyverso y el tidyverso		dplyr, tidyr, broom, lubridate	El flujo del tidyverso. Limpiar datos complejos						x	x	x				
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	x			x							
5	Creación de funciones propias y loops			Generar funciones propias en R y loops		x		x								
6	Escribir manuscritos en R	<a href="https://github.com/rstudio/rticles">https://github.com/rstudio/rticles</a>	Rticles	Transformar documento Rmd en un manuscrito	Agregar bibliografía al documento								x			
7	Presentaciones en R y documentos interactivos		Shiny	Transformar tus datos en una presentación o en una Shiny app	Generar una presentación o aplicación en R									x	x	x
8	Presentaciones				Presentación del trabajo realizado									x	x	x