

Plan de enseñanza

2021

Introducción al análisis de datos biológicos con R

Sobre el curso

¿De qué se trata el curso?

Este es un curso de dos meses en programación R para analizar datos biológicos. Se realiza de forma virtual y contará con una combinación de clases sincrónicas y asincrónicas, con ejercicios que se entregarán semanalmente y un proyecto final. Es un curso intensivo y desafiante que invita al estudiante a pensar y desarrollar algoritmos para el análisis de datos biológicos.

Es un curso de programación en R que prepara a los participantes para poder programar algoritmos y comprender los métodos desarrollados por otros. Específicamente, el curso tiene como objetivo enseñar a los participantes a:

- (a) desarrollar e interpretar scripts R
- (b) importar y manipular datos
- (c) describir y visualizar datos
- (d) crear informes con RMarkdown

Subiremos los videos de las clases al canal de youtube @somaquadrados. Los videos se pueden ver de forma asincrónica. El material escrito, diapositivas, scripts y tareas estarán disponibles en el repositorio de “Soma dos Quadrados”. Las clases presenciales se llevan a cabo semanalmente los viernes, de 2:00 pm a 5:00 pm. Las clases presenciales se utilizarán para resolver dudas y ejercicios.

El curso se ofrece de forma gratuita y la preinscripción estará abierta entre los días 20 y 20 del XX XXXX 2021. El curso comienza el día XX de XXXX hasta finales de XXXX de 2021.

Público

El curso está dirigido a estudiantes de pós-grado y/o investigadores interesados en aprender a utilizar R para el análisis de datos biológicos. Es un curso introductorio que trabajará con la importación, manipulación y análisis descriptivo y visual de datos biológicos.

El cupo máximo es de 20 personas.

Dedicación

Los estudiantes deben ver las lecciones en video y resolver los ejercicios disponibles en el repositorio de Soma dos Quadrados. La duración estimada de las clases es de 3h a 4h. Los ejercicios semanales son obligatorios.

El viernes habrá clases virtuales obligatorias y sincrónicas, donde se hará un repaso de la clase teórica y un lugar para resolver las dudas de los alumnos. Los ejercicios de tareas también se revisarán con los participantes. La duración media de estas clases es de 2h hasta 3h. También habrá un grupo de Telegram (mensajería on-line) para consultas y discusiones entre estudiantes en tiempo real. El curso requiere de seis a ocho horas de dedicación a la semana durante dos meses.

Responsables

Eliana Burgos (burgosef@gmail.com)

Marília Melo Favalesso (mariliabioufpr@gmail.com)

Carga horaria

48 h (8 clases de 6 h)

Enmienda

Clase 1

- **Introducción al curso**
 - Cronograma
 - Los materiales que va a utilizar
 - Cómo descargar e instalar R
 - Cómo descargar e instalar RStudio
 - Referencias bibliográficas del curso
- **Programación**
 - ¿Qué es?
 - ¿Para que sirve?
 - Lenguajes de programación
 - Lenguaje de programación R
 - Ventajas de utilizar la programación para analizar datos
- **IDE**
 - ¿Qué es?

- ¿Para que sirve?
- ¿Qué IDE podemos usar con R?
- **RStudio**
 - Presentación
 - Panel de inicio (área de trabajo, script, consola y salida)
 - “Tools > Global options”: Personalizar mi RStudio
 - “Help”
 - Mi primer código: “¡Hola mundo!”
 - La R como calculadora - operadores matemáticos
 - Hacer comentarios en el script
 - Atajos útiles
 - Cómo y dónde guardar mis archivos
 - Ejercicios

Clase 2

- **Objetos y atribuciones**
 - ¿Qué es un objeto?
 - ¿Para qué sirve un objeto?
 - ¿Cómo crear un objeto?
 - ¿Cómo asignar un objeto?
 - Confirmando el objeto
 - Objeto en la pestaña “environmental”
 - Reglas de asignación
 - Operaciones con objetos (operaciones matemáticas)
 - Atributos de los objetos:
 - * Modos de objeto: “numeric”, “logical” y “character”
 - * Estructura de objeto: Vector, array, matrix, lista, data.frame e factor
 - Ejercicios
- **Estructuras de objetos: qué son y cómo construir en R**
 - Vector
 - Array
 - Matrix

- Lista
- Data.frame (tibble)
- Factor (forcats package)
- ¿Cómo identificar la estructura del objeto?
- ¿Cómo cambiar la estructura del objeto?
- ¿Cómo enumerar los objetos creados?
- ¿Cómo guardar los objetos creados?
- ¿Cómo cargar los objetos creados?
- ¿Cómo quitar los objetos creados?
- Ejercicios
- **Funciones**
 - ¿Qué son?
 - ¿Por qué usarlo?
 - Argumentos de la función: valor del objeto y parámetros
 - Funciones base R
 - Funciones del paquete R
 - * ¿Qué son los paquetes?
 - * ¿Cómo descargar e instalar un paquete?
 - * ¿Cómo cargar uno o más paquetes?
 - * Paquetes R CRAN
 - * Paquetes de GitHub
 - * ¿Dónde se guardan mis paquetes?
- **Estructuras de Control**
 - ¿Qué es?
 - ¿Para que sirve?
 - “if-else”
 - Loopings de programación: for y while.
 - “repeat”, “next” y “break”
 - El operador pipe (%>%)
 - Buenas prácticas de programación en R
 - Ejercicios

Clase 3

- **Datos biológicos**
 - ¿Qué son?
 - Variable dependiente e independiente
 - Gestión de datos brutos
 - * Hojas de datos: ¿cómo estructurar la mía?
 - * Metadatos
 - Almacenes de datos
 - * Verificación de datos
 - * Valores atípicos
 - * Errores
 - * Datos perdidos
 - Ejercicios
- **Importación de bases de datos para el R**
 - Paquetes R
 - Importación de archivos de texto: .csv y .txt.
 - Importación de archivos de Excel: .xls y .xlsx.
 - Ejercicios
- **Manipulación de la base de datos en R**
 - Paquetes R
 - Una base de datos:
 - * Filtrar datos por columnas y/o filas
 - * Modificar datos en columnas y/o filas
 - * Unir datos de varias columnas y/o filas en una
 - * Separe los caracteres en varias columnas y/o filas
 - * Cambiar el nombre de columnas o filas
 - * Ordenar filas o columnas
 - * Seleccionar datos de columnas o filas
 - * Agregar nuevas columnas o filas
 - * Eliminar filas con valores repetidos
 - * Retirar valores NA
 - * Reemplazar valores NA

- * Cambiar el formato de la tabla (largo > ancho; ancho > largo)
- Dos o más bases de datos:
 - * Unir diferentes bases de datos por información en columnas
 - * Unir diferentes bases de datos por información en filas
- Ejercicios

Clase 4

- **Variables estadística**
 - Variables cuantitativas (discreta/continua)
 - Variables cualitativas (nominal/ordinal)
- **Distribución de probabilidad**
 - Paquetes R
 - Bernoulli / Binomial
 - Poisson
 - Normal
 - Ejercicios
- **Estadística descriptiva: medidas de posición**
 - Paquetes R
 - La media aritmética
 - Otras medias
 - La mediana
 - La moda
 - ¿Cuándo usar cada medida? Resumen
 - Ejercicios
- **Estadística descriptiva: medidas de dispersión**
 - Paquetes R
 - La varianza
 - La desviación estándar
 - El error estándar de la media
 - Asimetría, curtosis y momento central
 - Quantis
 - ¿Cómo utilizar las medidas de dispersión?

- Ejercicios

Clase 5

- **Estadísticas descriptivas: las frecuencias**

- Paquetes R
- Frecuencia absoluta
- Frecuencia relativa
- Frecuencia porcentual
- Frecuencia acumulada
- Frecuencia acumulada porcentual
- Ejercicios

- **Gráficos**

- Paquetes R
- Una variable cuantitativa
- Dos variables cuantitativas
- ¿Más de dos variables cuantitativas?
- Una variable cualitativa
- Dos variables cualitativas
- ¿Más de dos variables cualitativas?
- Una variable cuantitativa y cualitativa
- ¿Dos o más variables cuantitativas y cualitativas?
- Gráficos de tiempo
- Gráficos de red
- Ejercicios

Clase 6

- **Mapas con R**

- Paquetes R
- ¿Qué es un mapa?
- ¿Qué es un ráster?
- ¿Qué es un vector?
- ¿Qué es un punto?
- Sistema de información geográfica

- Elementos de un mapa
- Base de datos geoespacial
- Mapas para datos continuos
- Mapas para datos de recuento
- Mapas para datos binarios
- Mapas para series de tiempo
- Cómo crear un mapa animado
- Ejercicios

Clase 7

- **Informes dinámicos con RMarkdown**
 - El ‘markdown’ y el ‘RMarkdown’
 - La sintaxis de Markdown
 - Incluir tablas
 - Incluir imágenes
 - “R code chunks” y “inline R code”
 - Otros lenguajes de programación
 - YAML
 - Formatos de salida
 - Compilar un documento ‘RMarkdown’

Clase 8

- **Resolución de ejercicios y presentación de informe final**
-

Referencias

(Todos los libros de referencia de este curso tienen copias para distribución gratuita).

Aragón TJ. (2017) **Population Health Data Science with R: *Transforming data into actionable knowledge***.

Atlan (2019) **Introduction to GIS: Manipulating and Mapping Geospatial Data in R**.

Brunsdon C, Comber L. (2019) **An Introduction to Spatial Analysis and Mapping**. Sage.

Chang W. (2013) **R Graphics Cookbook: Practical Recipes for Visualizing Data**. O’Reilly Media.

DiMaggio C. (2014) **Spatial Epidemiology Notes: Applications and Vignettes in R**. Columbia University, NY

- Engel CA. (2019) **Using Spatial Data with R**.
- Gillespie C, Lovelace R. (2021) **Efficient R Programming: A practical guide to smarter programming**. O'Reilly Media.
- Grolemund G. (2014) **Hands-On Programming with R**. O'Reilly Media.
- Healy K. (2018) **Data visualization: A practical introduction**. Princeton University Press.
- Hijmans, R. 2019. **Spatial Data Analysis and Modeling with R**.
- Holmes S, Huber W. (2018) **Modern Statistics for Modern Biology**. Cambridge.
- Irizarry RA. (2021) **Introducción a la ciencia de datos: *Análisis de datos y algoritmos de predicción con R***. CRC Press.
- Irizarry RA, Love MI. (2015) **Data analysis for the life sciences**. Leanpub.
- Lovelace R, Nowosad J, Muenchow J. (2021) **Geocomputation with R**. CRC Press.
- Peng RD. (2016a) **Exploratory Data Analysis with R**. Leanpub.
- Peng RD. (2016b) **Report Writing for Data Science in R**. Leanpub.
- Venables WN, Smith DM. (2021) **An introduction to R**. R Core Team.
- Xie Y, Allaire JJ, Grolemund G. (2021) **R Markdown: The Definitive Guide**. CRC Press.
- Wickham H. (2016) **ggplot2: Elegant Graphics for Data Analysis (Use R)**. Spring.
- Wickham H, Grolemund G. (2014) **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**. O'Reilly Media.
- Wilke CO. (2019) **Fundamentals of Data Visualization**. O'Reilly Media.
- Wright C, Ellis SE, Hicks SC, Peng RD. (2021) **Tidyverse Skills for Data Science in R**. Johns Hopkins University.