

Curso introductorio de R

Manuel Toral

Octubre 2018

¿Dónde es el curso?

El curso se llevará a cabo en las instalaciones *Cubic Idea* en Colima 158, Colonia Roma Norte, en la Ciudad de México.

¿Cómo contactar al instructor?

- Correo electrónico: jmtoralcruz@gmail.com
- Teléfono: Será provisto el primer día del curso.

Objetivo del curso

El objetivo del curso es que aprendas a utilizar efectivamente R para su vida profesional diaria. Así pues, al final del curso se espera desarrolles habilidades con cuatro objetivos particulares.

1. *Ahorrar tiempo en el procesamiento de información:* El talento de las personas no debe perderse en procesar y talachear bases de datos. R es una herramienta útil para hacer esta tarea de manera más rápida que con una hoja de cálculo.
2. *Pensar de manera estructurada:* Pensar nuestros procesos de análisis desde una perspectiva estructurada nos da la oportunidad de encontrar respuestas a nuestras preguntas de manera más sencilla.
3. *Aumentar los límites de acceso a la información:* Muchas de las bases de datos que se usan (y se necesitan) hoy en día están en formatos como JSON o SHP. R es capaz de leerlos y transformarlos en información útil.
4. *Automatizar tareas:* Las computadoras fueron inventadas para repetir procesos. Una y otra vez. Cada vez que un ser humano repite una actividad frente a una computadora, está haciendo algo mal. La idea es evitar ese tipo de repeticiones e instruir a la computadora a que las haga.
5. *Aprovechar a la comunidad:* Las computadoras fueron inventadas para repetir procesos. Una y otra vez. Cada vez que un ser humano repite una actividad frente a una computadora, está haciendo algo mal. La idea es evitar ese tipo de repeticiones e instruir a la computadora a que las haga.

Material guía

Cheatsheets

- Base R
- Variables de tiempo
- Transformación de datos con dplyr
- RMarkdown

- `ggplot2`

Libros

(Da click en el nombre del libro para abrirlo.)

- La guía pirata de R - Phillips
- R for Data Science - Golemund & Wickham
- Geocomputation with R - Lovelace, Nowosad & Muenchow

Páginas webs

- Stack Overflow
- STHDA

Requisitos

- Es conveniente que elijas un problema que quieras resolver en tu día a día y poder hacer ejercicios con información de ese tipo. Te recomiendo que me mandes un día antes alguna base con la que quieras trabajar o algún tema que quieras explorar con datos.
- El único requisito es tener instaladas las tres herramientas necesarias. Aquí unas breves instrucciones. Si necesitas ver el paso a paso, consulta el archivo `Instalar_Paso_a_Paso.pdf` en el repositorio.
- Haber instalado R en la computadora.
 1. Da clic **aquí**.
 2. Seleccionar el archivo según tu sistema operativo.
 3. Seleccionar el link de **install R for the first time**.
 4. Descargar el archivo y abrirlo para seguir las instrucciones.
- Haber instalado *RStudio* en la computadora.
 1. Da clic **aquí**.
 2. Descarga el instalador de la sección **Installers for Supported Platforms** de acuerdo a tu sistema operativo.
 3. Da clic en el archivo descargado y sigue las instrucciones.
- Haber instalado \LaTeX . El link está aquí.
 1. Este está más complicado. **Aquí** están las instrucciones, pero sí está muy engorroso, en la clase lo resolvemos.
- *Estar preparadx para la frustración.*

Uso profesional de R y RStudio

Gran parte de las cosas que uno hace en R forman parte de reportes, presentaciones y textos. En esta primera sección aprenderemos cómo usar R en *RStudio* de manera profesional a través de *R Markdown*. Los temas son los siguientes:

1. Crear archivos en *R Markdown* y “knitearlos” para generar archivos estéticamente atractivos y claros.
2. Entender el uso de los *chunks*.
3. Entender *RStudio* y sus componentes.

4. Uso de la consola y la terminal.
5. Uso de comandos `echo`, `warning`, `fig.*` y `eval` en la escritura de *R Markdown*.
6. Exportar *PDFs*, archivos en *HTML* y presentaciones en *Beamer*.
7. ¿Cómo instalar librerías?

Brevísima introducción a R base

En R hay tres sintaxis: *Signo de dólar*, *Formula* y *La sintaxis del tidyverso*. Mi favorita y, a mi parecer, la más útil y fácil es la del tidyverso. No obstante, repasaremos algunas funciones de la sintaxis básica del lenguaje para que no te agarren en curva.

1. Como crear objetos (y por qué este método de trabajo es superior a hojas de cálculo y ventanas de software)
2. Tipos de datos (vectores, listas, data frames, matrices, strings)
3. El método de la *indexación* y cómo usar el famoso `$`.
4. Manipulaciones básicas para explorar bases.
5. Funciones básicas para obtener estadística descriptiva.
6. Uso de variables de tiempo.

El tidyverse

1. ¿Qué es y cómo usar un pipe `%>%`?
2. Filtrar y filtrar con condiciones
3. Seleccionar variables y modificar sus nombres
4. Convertir de *long* a *wide* y viceversa.
5. Agrupar, convertir y “colapsar” bases de datos
6. Uso de `grep` y una muy breve introducción a *regular expressions*.

Importar (y usar efectivamente) bases de datos

1. Importar datos en formatos `.xls` (*MS Excel*), `.dta` (*SATA*), `.sav` (*SPSS*), `.dta` (*SATA*), `.txt` y `.csv`.
2. Funciones básicas para el análisis estadístico (`lm`, `predict`, `stepAIC`, `anova`).
3. Manipulación de datos: hacer muestras y subsets, unir bases con `merge`, creación de variables binomiales, entre otras.
4. Manipulación de *strings*
5. Introducción a la visualización de datos con `ggplot`.

Visualización

1. Introducción a la gramática de `ggplot2` y al uso de `aes()`.
2. Introducción a la lógica de capas.
3. Elementos estructurales de las visualizaciones.
4. Modificación de elementos guía: títulos, subtítulos, títulos de ejes, `thicks`, escalas, límites, grids y guías.
5. `qplot`, esa salida fácil.
6. Tipos de visualización con respecto a la naturaleza de los datos.

Últimas fronteras

1. Breve introducción a la minería de texto.
2. Breve introducción a los datos geográficos.
3. Breve introducción al scrapeo de información.