

Clase 2 - Programación con R y Rstudio

Curso Introducción al Análisis de datos con R para la
acuicultura

Dr. José Gallardo Matus y Dra. María Angélica Rueda.
Pontificia Universidad Católica de Valparaíso
jose.gallardo@pucv.cl

02 July 2021

PLAN DE CLASE

1). Introducción

- ▶ Preguntas al curso.
- ▶ ¿Qué es R y Rstudio?
- ▶ ¿Por qué usar R para el análisis de datos?
- ▶ R en investigación reproducible.
- ▶ ¿Qué es un script?

2). Práctica con R y Rstudio (cloud)

- ▶ Iniciar un proyecto de análisis de datos con R.
- ▶ Escribir un script.
- ▶ Familiarizarse con manipulación de objetos y datos.

PREGUNTAS AL CURSO

- 1). ¿ Por qué deseas hacer análisis de datos con R?**
- 2). ¿ Qué problemas has tenido cuando quieres rehacer un análisis de datos a partir de un set de datos antiguo o en el que no has trabajado por mucho tiempo?**

¿QUÉ ES R?

- 1). **R** es un lenguaje y entorno de programación de código abierto o libre creado por Ross Ihaka y Robert Gentleman en 1993 (University of Auckland) para realizar análisis estadísticos y gráficos.
- 2). Los usuarios de R tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar el **software**.
- 3). Utilizar **R** supone un ahorro económico para los estudiantes, las instituciones educativas o incluso las empresas que decidan usarlo.

¿POR QUÉ USAR “R”?

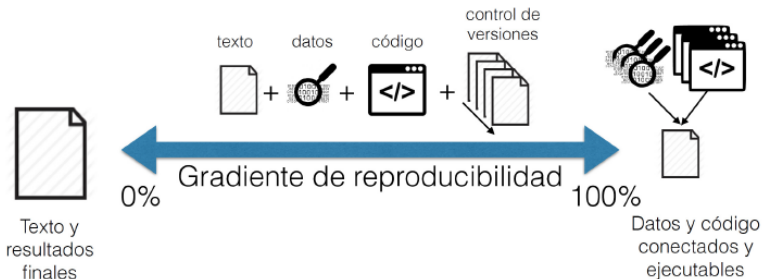
- ▶ Aprender a usar **R** te da ***independencia digital***, te permite ***cooperar con otros*** y ***beneficiarte de la ayuda de otro***.
- ▶ Actualmente existen cerca de **17.000 librerías** o aplicaciones disponibles de forma gratuita para trabajar con R en ámbitos tan diferentes como las ciencias sociales, la economía, la astronomía, la ingeniería y la acuicultura.
- ▶ **R** permite entonces difundir el conocimiento a toda la sociedad y no solo a los que pueden pagar por ella.

BENEFICIOS USAR R PARA EL ANALISTA DE DATOS

- ▶ **Permite la ejecución de tareas de análisis repetitivo sin esfuerzo.**
- ▶ **Muy fácil corregir y regenerar resultados, tablas y figuras.**
- ▶ **Reducción drástica del riesgo de errores.**
- ▶ **Facilita la colaboración.**
- ▶ **Mayor facilidad para escribir reportes.**
- ▶ **Facilita el proceso de revisión.**
- ▶ **Ahorro de tiempo y esfuerzo al reutilizar código en otros proyectos.**

R EN INVESTIGACIÓN REPRODUCIBLE

Trabajar con R permite realizar **Investigación reproducible**. La investigación reproducible implica que desde los mismos datos y códigos se generarán los mismos resultados.



ALGUNOS CRITERIOS DE REPRODUCIBILIDAD

- ▶ **Los datos originales están disponibles en la nube.**
- ▶ Los datos están almacenados en formato abierto (texto) .
- ▶ **Todo el análisis y manejo de datos se hace mediante código.**
- ▶ El código genera las tablas y figuras finales.
- ▶ **Los datos brutos están separados de los datos derivados.**
- ▶ Existe un '*script*' maestro que ejecuta todos los pasos del análisis ordenadamente.
- ▶ **Existe un documento README que explica los objetivos y organización del proyecto.**
- ▶ Tanto el reporte, como los datos y código son públicos.

RUTA DEL ANÁLISIS DE DATOS REPRODUCIBLE CON R

- ▶ Toma de datos.
Es importante estandarizar y mantener estructura.
- ▶ Manipulación de datos.
Es importante cuidar los datos originales, trabajaremos con R + Rstudio
- ▶ Análisis datos integrado con texto.
Facilita la colaboración, trabajaremos R + Rstudio + Markdown
- ▶ Publicar resultados.
Es importante comunicar de forma efectiva

¿QUÉ ES R STUDIO?

- ▶ **Rstudio** es el más popular entorno de desarrollo integrado (integrated development environment, IDE) para trabajar con **R**.
- ▶ **Rstudio** es un *software* libre y de código abierto creado por **Joseph J. Allaire en 2009** para la ciencia de datos, la investigación científica y la comunicación técnica.
- ▶ Actualmente es mantenido por la Corporación de Beneficio Público **Rstudio PCB**, la que ha creado otros software como Rmarkdown.

R ES UN LENGUAJE ORIENTADO A OBJETOS

Ejemplo del objeto “Bombero”

- ▶ En Chile trabajan *ad honoren*.
- ▶ Apagan incendios.
- ▶ Pueden usar diferentes equipos.
- ▶ Reciben instrucciones sobre como apagar un incendio.
- ▶ Pero, no pueden detener a ladrones, ni dirigir el transito (Objeto policía).

R ES UN LENGUAJE ORIENTADO A OBJETOS 2

Objeto `data.frame`

- ▶ Objeto similar a una tabla de datos.
- ▶ Almacenan texto o números.
- ▶ Primera fila contiene el nombre de las variables.
- ▶ Puedo unir con otro **`data.frame`**.
- ▶ Puedo aplicar funciones para calcular estadísticos.
- ▶ Pero, no tiene atributos de una matriz, ni de un vector, ni es una serie de tiempo.

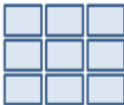
TIPOS DE OBJETOS PARA TRABAJAR CON R

Vector



- 1 column or row of data
- 1 type (numeric or text)

Matrix



- multiple columns and/or rows of data
- 1 type (numeric or text)

Data Frame



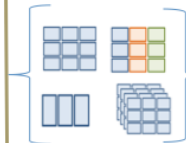
- multiple columns and/or rows of data
- multiple types

Array



- 3 dimensiones
- 1 tipo: numérico
- o caracter

Listas



- Conjunto de objetos diversos

¿QUÉ ES UN SCRIPT?

Los ***scripts*** son documentos de texto con una secuencia de comandos que permiten ejecutar programas.

Estos archivos son iguales a cualquier documentos de texto, pero **R puede leer y ejecutar** el código que contienen.

Los códigos de **R** están contenidos en librerías o packages.

Algunos ***script*** que usaremos en este curso tienen extensión de archivo **.R**, por ejemplo mi_script.R o **.Rmd** (R+markdown) por ejemplo reporte.Rmd.

EJEMPLO DE SCRIPT R

Práctica con R y Rstudio

Clase 2 – Programación con R

EXPLORAR R STUDIO CLOUD

EXPLORAR R STUDIO

TRABAJO EN SALAS

- 1). Guía de trabajo programación con R disponible en drive.**
- 2). La tarea se realiza en Rstudio.cloud.** Ingresa al siguiente proyecto de **Rstudio.Cloud**

RESUMEN DE LA CLASE

- ▶ Paradigma de la investigación reproducible.
- ▶ Ruta del análisis de datos reproducible con **R**.
- ▶ Iniciamos un proyecto de análisis de datos.
- ▶ Escribimos un código de programación con **R** y **Rstudio cloud**.
- ▶ Nos familiarizamos con la manipulación de objetos.