

CLASE 01 - PROGRAMACIÓN CON R

Diplomado en Análisis de datos con R para la Acuicultura.

Dr. José Gallardo Matus

Pontificia Universidad Católica de Valparaíso

03 April 2022

PLAN DE CLASE

1. Introducción

- ▶ ¿Qué es R y Rstudio?
- ▶ ¿Por qué usar R para el análisis de datos en acuicultura?
- ▶ Investigación reproducible.

2. Práctica con R y Rstudio (cloud)

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

¿QUÉ ES R?

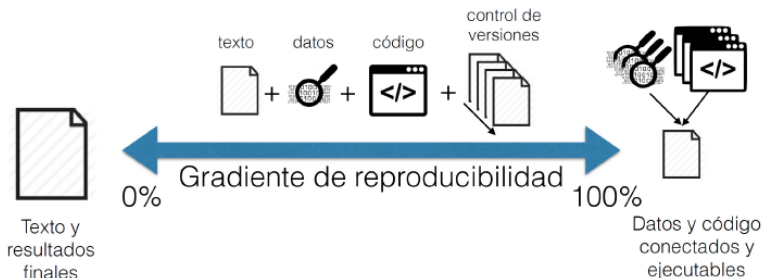
1. **R** es un lenguaje y entorno de programación de código abierto o libre creado por Ross Ihaka y Robert Gentleman en 1993 (University of Auckland) para realizar análisis estadísticos y gráficos.
2. Los usuarios de R tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar 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”?

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

INVESTIGACIÓN REPRODUCIBLE

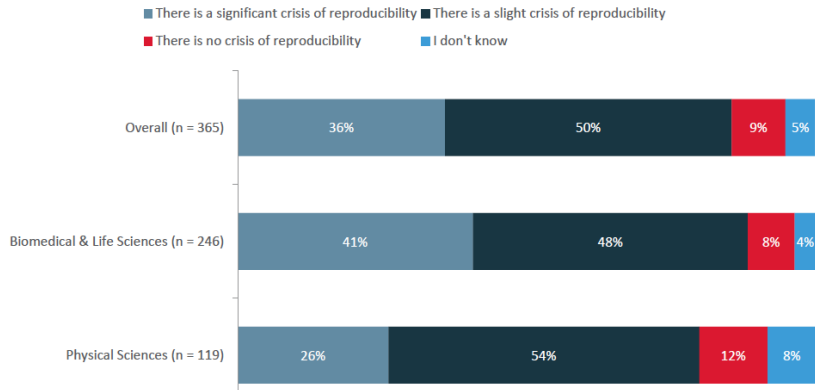
La investigación reproducible implica que desde los mismos datos y códigos se generarán los mismos resultados.



Peng. 2011

CRISIS DE REPRODUCIBILIDAD

70 % (1103/1,576) de los investigadores declaran que quisieron pero no pudieron reproducir un experimento realizado por otro científico.



Baker. 2016

CRISIS DE RERPRODUCIBILIDAD EN ACUICULTURA

North American
Journal of Aquaculture

Retraction | [Free Access](#)

Retraction

 This article retracts the following: 

First published: 05 September 2019 | <https://doi.org/10.1002/naaq.10119>

 PDF  TOOLS  SHARE

Abstract

The following article has been retracted at the request of the author, Nagaraj G. Chatakondi, the author's employer, U.S. Department of Agriculture (USDA), Agricultural Research Service, with the agreement of the journal editors, Reginal M. Harrell and Christopher C. Kohler, the American Fisheries Society, and John Wiley and Sons, Ltd:

Volume 79(4), October 2017: "Effective Dose of Salmon GnRH α for Induction of Ovulation in Channel Catfish," by Nagaraj G. Chatakondi, pages 310–316. DOI: [10.1080/15222055.2017.1356407](https://doi.org/10.1080/15222055.2017.1356407).

An internal USDA review identified issues with the data and underlying research record, and the USDA is unable to assure the scientific validity of the reported data and resulting conclusions.

Chatakondi, 2017

Pridgeon and Klesius, 2016

ALGUNOS CRITERIOS DE REPRODUCIBILIDAD

- ▶ 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.

Sánchez et al. 2016

BENEFICIOS 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 por pares.
- ▶ **Ahorro de tiempo y esfuerzo al reutilizar código en diferentes proyectos.**

RUTA DEL ANÁLISIS DE DATOS REPRODUCIBLE CON R

1. Toma de datos.

Es importante estandarizar y mantener estructura.

2. Manipulación de datos.

Es importante cuidar los datos originales.

Trabajaremos con R + Rstudio

3. Análisis datos integrado con texto.

Facilita la colaboración.

Trabajaremos con RMarkdown.

4. Publicar resultados y control de versiones.

Es importante comunicar de forma efectiva.

Trabajaremos Github.

CONCEPTOS BÁSICOS DE PROGRAMACIÓN

Metáfora de la maquina expendedora de bebidas

1. La máquina tiene una función específica.
2. Los productos son objetos almacenados de forma ordenada.
3. Los objetos tienen características (Nombre, precio, ubicación).
4. Para comprar debo seguir una secuencia de pasos (similar a un programa = códigos en secuencia).



¿QUÉ ES UN SCRIPT?

1. Los scripts son documentos de texto con una secuencia de comandos que permiten ejecutar programas.
2. Estos archivos son iguales a cualquier documentos de texto, pero R puede leer y ejecutar el código que contienen.
3. Los códigos de R están contenidos en librerías o packages o aplicaciones.
4. Algunos script que usaremos en este diplomado tienen extensión de archivo .R, por ejemplo mi_script.R.

EJEMPLO R SCRIPT

Nombre script

Ejecutar

Clase_01-script.R*

Source on Save

Profesores

Next Prev All Peces Replace All

☐ In selection ☐ Match case ☐ Whole word ☐ Regex ☒ Wrap

```
1 # -----
2 # Clase 01 - Programación con R
3 # Dr. José Gallardo Matus
4 # 05 abril 2022
5 # Diplomado en Análisis de datos con R para la Acuicultura.
6 # -----
7
8 #R reconoce funciones matemáticas
9 29+29
10 29*29
11 29==29
12
13 # Error en R
14 29 + diez
15
16 # Crear un objeto
17 diez <- 10
18 29 + diez
19
```

Metadata

Comentarios

Códigos de programación

5:60 (Untitled) R Script

R ES UN LENGUAJE ORIENTADO A OBJETOS

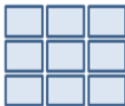
Tipos de objetos para trabajar con R

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



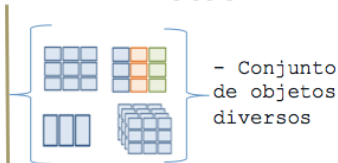
- multiple columns and/or rows of data
- multiple types

Array



- 3 dimensiones
- 1 tipo: numérico
- o caracter

Listas



- Conjunto de objetos diversos

OBJETO: DATA.FRAME

Principales características.

- ▶ 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, no es una serie de tiempo.

¿QUÉ ES R STUDIO?

1. **Rstudio** es el más popular entorno de desarrollo integrado (integrated development environment, IDE) para trabajar con **R**.
2. **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.
3. Actualmente es mantenido por la Corporación de Beneficio Público **Rstudio PCB**, la que ha creado otros software como Rmarkdown.

EJEMPLO RSTUDIO - VERSION CLOUD

The screenshot displays the RStudio Pro interface with the following components:

- Script Editor:** Contains the following R code:

```
1 library(ggplot2)
2 library(readxl)
3 Fraser_et_al_data <- read_excel("LIBRO_R_ACUICULTURA/Estudios_de
4 Fraser_et_al_data <- na.omit(Fraser_et_al_data)
5 Fraser_et_al_data$Temp <- as.factor(Fraser_et_al_data$Temp)
6 ggplot(Fraser_et_al_data, aes(x=Length, y=Mass, colour=Temp))+
7 geom_point()
```
- Console:** Shows the execution of the script:

```
> ggplot(Fraser_et_al_data, aes(x=Length, y=Mass, colour=Temp))+
+ geom_point()
> |
```
- Environment:** Displays the loaded data frame: "Fraser_et_al_data" with 1311 observations and 21 variables.
- Plots:** A scatter plot titled "Utilidades: Archivos, Gráficos, Librerías y Ayuda" showing "Mass" on the y-axis (0 to 7500) and "Length" on the x-axis (40 to 80). The points are colored by "Temp", with a legend indicating two categories: 10 (red) and 16 (teal).

Consola: Muestra los resultados de lo que se ejecuta/compila en el Script.

PRÁCTICA PROGRAMACIÓN CON R

Guía de trabajo programación con R en Rstudio.cloud.



0. RUN



1. STUDY



3. SHARE



4. IMPROVE

RESUMEN DE LA CLASE

- ▶ Investigación reproducible.
- ▶ Ruta del análisis de datos reproducible con **R**.
- ▶ Iniciamos un proyecto de análisis de datos con **R**.
- ▶ Escribimos un código de programación de **R** con **Rstudio cloud**.
- ▶ Nos familiarizamos con la manipulación de objetos y datos: vector, matrices, data.frame y listas.