

DIPLOMADO EN ANÁLISIS DE DATOS CON R PARA LA  
ACUICULTURA

CLASE 2  
PROGRAMACIÓN CON R  
Dr. José Gallardo

# PLAN DE LA CLASE

## 1.- Introducción

- Preguntas al curso
- Estudio de caso
- Investigación reproducible
- Software para el análisis de datos

## 2.- Práctica con R y Rstudio cloud.

- Escribir un código de programación o scripts.
- Familiarizarse con manipulación de objetos y datos.

# Introducción

*Clase 2 – Programación con R*

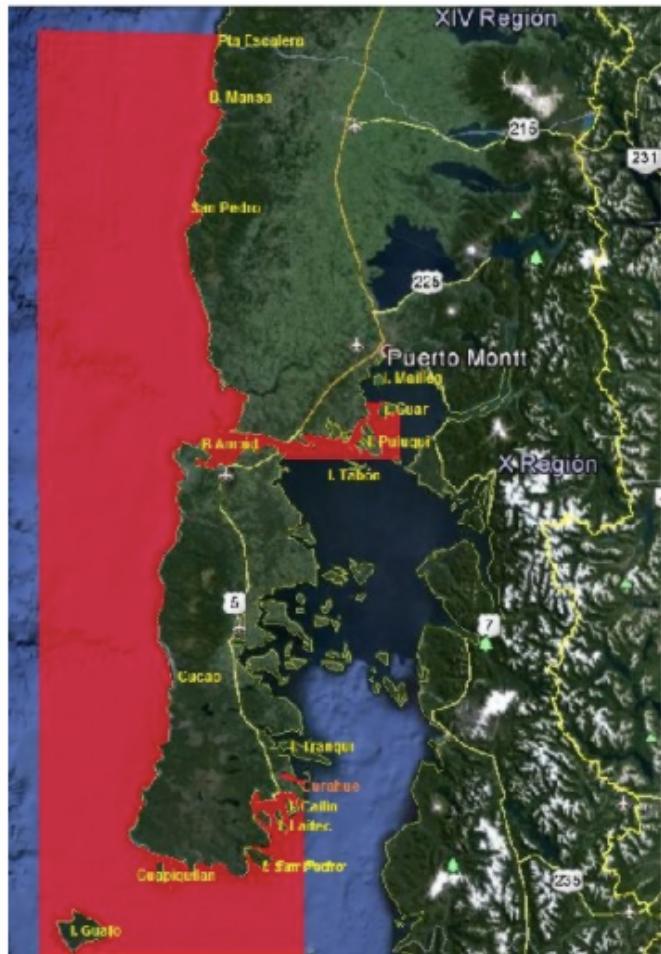


# PREGUNTAS AL CURSO

**¿Por qué hacer análisis de datos con R?**

**¿Qué problemas has tenido al hacer análisis de datos?**

# ESTUDIO DE CASO



## CONSECUENCIAS ACUICULTURA

### Acuicultura

45 centros de cultivo  
14 empresas

### Mortalidad

+25 millones de peces  
40 mil toneladas

**Vertimiento de mortalidad al mar**  
5.000 toneladas

# ESTUDIO DE CASO

## Reporte Crisis Social Ambiental en Chiloé GREENPEACE

### Resumen Ejecutivo Septiembre 2016

La principal conclusión de esta investigación es que si bien el **bloom de algas** estaba presente en la zona con anterioridad al vertimiento, **el salmón vertido actuó como un "fertilizante"** de éste aumentando su magnitud, intensidad y alcance.

# TRABAJO EN GRUPO

**¿Quién fue el analista?**

**¿Qué datos están disponibles?**

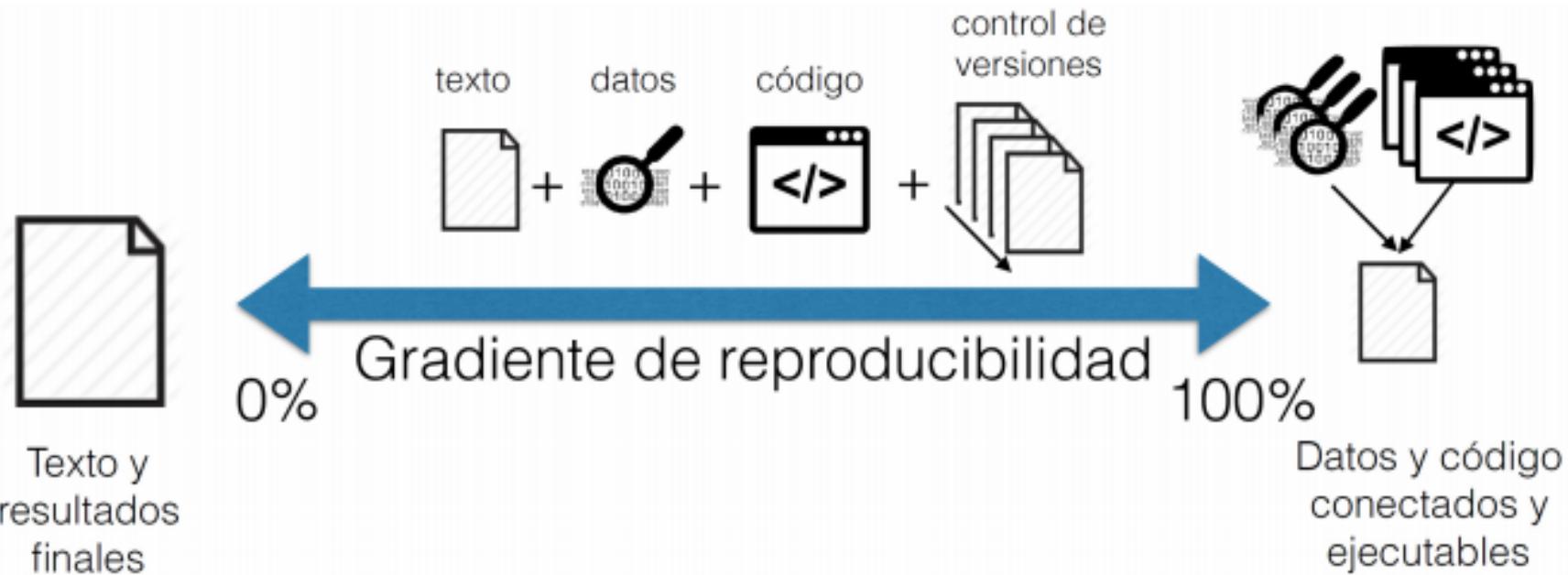
**¿Qué test estadístico se utilizó y cuál fue su significancia?**

**¿Qué software utilizaron?**

**¿Podría usted llegar a la misma conclusión?**

# INVESTIGACIÓN REPRODUCIBLE

**Investigación reproducible** implica que desde los mismos datos y códigos se generarán los mismos resultados.



Peng. 2011. Science 334 (6060). Sánchez et al. 2016 Ecosistemas 25(2): 83-92

# 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.

# BENEFICIOS DE LA CIENCIA REPRODUCIBLE 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.
- Facilitan la colaboración.
- Mayor facilidad para escribir reportes.
- Facilita el proceso de revisión.
- Ahorro de tiempo y esfuerzo al reutilizar código en otros proyectos.

Adaptado de Sánchez et al. 2016 Ecosistemas 25(2): 83-92

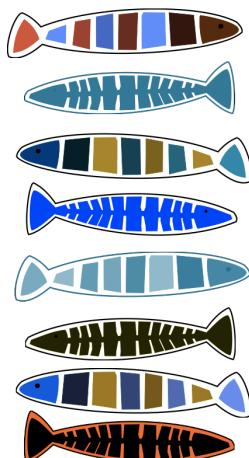
# RUTA DEL ANÁLISIS DE DATOS REPRODUCIBLE

Toma de datos

Manipulación de datos

Análisis  
datos  
integrado  
con texto

Publicar  
resultados



Adaptado de Sánchez et al. 2016 Ecosistemas 25(2): 83-92



# ¿QUÉ ES R?

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 estadístico y gráficos.

Los usuarios de R tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar el software.

Utilizar R supone un ahorro económico para los estudiantes, las instituciones educativas o incluso las empresas que decidan usarlo.

[www.r-project.org/](http://www.r-project.org/)

Adaptado de los principios del software libre [www.gnu.org](http://www.gnu.org)

# ¿POR QUÉ USAR “R” EN CIENCIA REPRODUCIBLE?

- 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 en R para trabajar en ámbitos tan diferentes como la economía, la astronomía, la ingeniería entre otras.
- R permite entonces difundir el conocimiento a toda la sociedad y no solo a los que pueden pagar por ella.

Adaptado de los principios del software libre [www.gnu.org](http://www.gnu.org)

## ¿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.

<http://www.rstudio.com>

# Práctica con R y Rstudio

*Clase 2 – Programación con R*

# EXPLORAR R STUDIO CLOUD

The screenshot shows the R Studio Cloud interface. On the left, there's a sidebar with various links: Spaces, Your Workspace (highlighted), Diplomado R, New Space, Learn, Guide, What's New, Primers, Cheat Sheets, Help, and Current System Status. The main area is titled 'All Projects' and shows a single project named 'Clase 02 Programación con R' created by 'Jose Gallardo' on April 15, 2021, at 10:26 PM. There are buttons for Assignment, Delete, Move, and Export. At the bottom, there's a footer with the R Studio Cloud logo, links to Terms and Status, and social media icons for LinkedIn, Facebook, Twitter, and GitHub.

R Studio Cloud

Diplomado R  
Jose Gallardo

Projects Members About

All Projects

New Project

List All projects Sort By name

ASSIGNMENT Clase 02 Programación con R

jG Jose Gallardo

Created Apr 15, 2021 10:26 PM

Delete Move Export

R Studio Cloud

Terms Status

in f t g

© 2020 RStudio, PBC

**Your workspace:** Puedes hacer tus propios proyectos durante el diplomado.

**Diplomado R:** Aquí las tareas del diplomado.

# EXPLORAR R STUDIO

Diplomado R / 02 Programación con R



Jose Gallardo

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R 4.0.3

Script\_example.R x

```
1 # -----
2 # R script example
3 # Jose Gallardo
4 # April 20, 2021
5 # Diplomado en Análisis de datos con R para la acuicultura
6 #
7
8 # Version R
9 R.version.string
10
```

40:1 # (Untitled) R Script

Console Terminal Jobs

~/

>

**Script**

Environment History Connections Tutorial

Import Dataset

List

R Global Environment

Environment is empty

**Object and history**

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Cloud > project

	Name	Size	Modified
<input type="checkbox"/>	..		
<input type="checkbox"/>	.Rhistory	0 B	Apr 15, 2021, 10:30 AM
<input type="checkbox"/>	project.Rproj	205 B	Apr 18, 2021, 10:30 AM
<input type="checkbox"/>	Script_example.R	789 B	Apr 18, 2021, 10:30 AM

**Files, plots, help**

# ¿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 los puede leer y ejecutar** el código que contienen.

Algunos script que usaremos en este curso tienen extensión de archivo **.R**, por ejemplo **mi\_script.R** o **.Rmd** (Rmarkdown) por ejemplo **reporte.Rmd**.

# EJEMPLO DE SCRIPT R

The screenshot shows an RStudio interface with a script file named "Script\_example.R". The file contains the following R code:

```
1 # -----
2 # R script example
3 # Jose Gallardo
4 # April 20, 2021
5 # Diplomado en Análisis de datos con R para la acuicultura
6 #
7
8 # Version R
9 R.version.string
10
11 # ¿Como citar R?
12 citation()
13
14 # Remover objetos de la sesión de trabajo
15 rm(list = ls())
16
17 # En que directorio estoy
18 getwd()
```

Annotations highlight specific parts of the code:

- A red box surrounds the file tab "Script\_example.R" in the top-left corner.
- A red box surrounds the "Run" button in the top-right toolbar.
- A blue bracket on the right side groups the first five lines of code (metadata) under the label "Metadata".
- A blue bracket on the right side groups the remaining lines of code (comments and codes) under the label "# Comentarios Códigos".

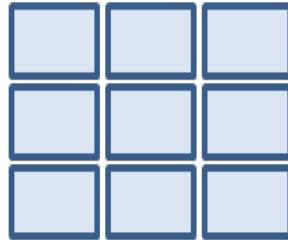
# TIPOS DE OBJETOS PARA TRABAJAR COM R.

## Vector



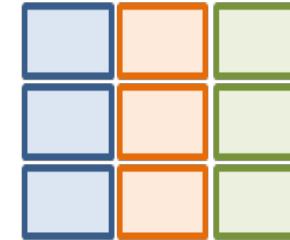
- 1 column or row of data
- 1 type (numeric or text)

## Matrix



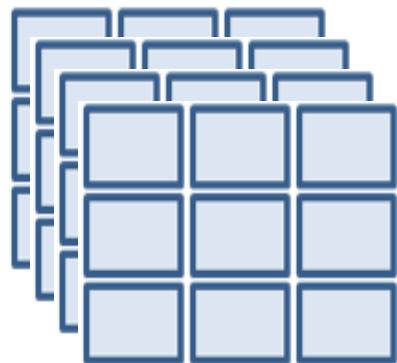
- multiple columns and/or rows of data
- 1 type (numeric or text)

## Data Frame



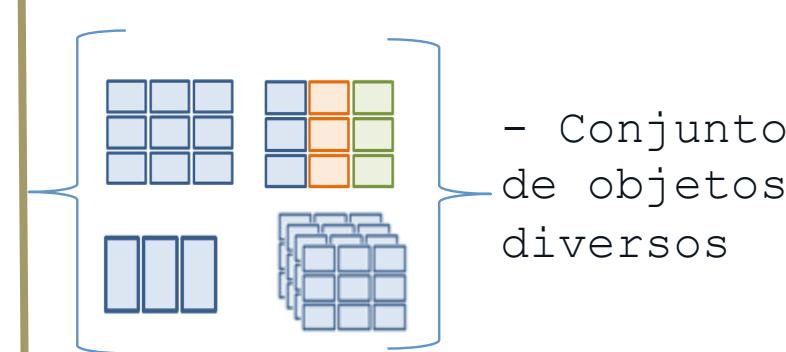
- multiple columns and/or rows of data
- multiple types

## Array



- 3 dimensiones
- 1 tipo: numérico
- o carácter

## Listas



- Conjunto de objetos diversos

# TRABAJO EN SALAS

**1.- Abir y ejecutar Script\_example.R desde el proyecto**

<https://rstudio.cloud/spaces/135178/project/2447826>

**2.- Realizar guía de trabajo clase 2.**

# AUTOAPRENDIZAJE CON DATACAMP

## Tarea para clase 3

The screenshot shows a DataCamp course page for 'Introduction to Importing Data in R'. The page has a dark header with the text 'INTERACTIVE COURSE' and the course title 'Introduction to Importing Data in R'. Below the title are three buttons: 'Practice Now' (green), 'Replay Course' (white), and 'Bookmark' (white). At the bottom of the page, there is a summary bar with the following information: 3 hours, 11 Videos, 42 Exercises, 139,608 Participants, and 3,550 XP.

### 3 Importing Excel data

100%

Excel is a widely used data analysis tool. If you prefer to do your analyses in R, though, you'll need an understanding of how to import .csv data into R. This chapter will show you how to use `readxl` and `gdata` to do so.

- |   |          |
|---|----------|
| ▶ <code>readxl</code> (1)               | ✓ 50 xp  |
| ◀/▶ List the sheets of an Excel file    | ✓ 100 xp |
| ◀/▶ Import an Excel sheet               | ✓ 100 xp |
| ◀/▶ Reading a workbook                  | ✓ 100 xp |
| ▶ <code>readxl</code> (2)               | ✓ 50 xp  |
| ◀/▶ The <code>col_names</code> argument | ✓ 70 xp  |
| ◀/▶ The <code>skip</code> argument      | ✓ 100 xp |

## Solo capítulo 3

# RESUMEN DE LA CLASE

- Revisión de ventajas de la investigación reproducible.
- Escribimos un código de programación con R y Rstudio cloud.
- Iniciamos un proyecto de análisis de datos.