

# Introduccion a R

## Transformando Datos en Conocimiento

Emilio Aguilar Cubilla

Diciembre, 2023

# Que es R?

R es un lenguaje de programación y un entorno de software diseñado específicamente para estadísticas y análisis de datos. Fue desarrollado por Ross Ihaka y Robert Gentleman en la Universidad de Auckland, Nueva Zelanda, en la década de 1990. Desde entonces, R se ha convertido en uno de los lenguajes más populares en el ámbito de la estadística y la ciencia de datos

## Características Clave

1. **Estadísticas y Análisis de Datos:** R es potente para realizar análisis estadísticos y exploración de datos, se pueden realizar tareas como regresión, análisis de varianza, machine learning, y más.
2. **Gráficos:** R tiene excelentes capacidades para la visualización de datos. Puedes crear una variedad de gráficos de alta calidad para explorar y comunicar patrones en los datos.
3. **Comunidad y Paquetes:** R tiene una comunidad activa de usuarios y desarrolladores. Además, cuenta con un extenso sistema de paquetes que amplían sus capacidades.

# Que es GIT?

**Git** es un sistema de control de versiones distribuido ampliamente utilizado. Fue creado por Linus Torvalds en 2005 y se ha convertido en una herramienta esencial para el desarrollo de software colaborativo.

## Beneficios de Git:

1. **Trabajo Colaborativo:** Git permite a varios desarrolladores trabajar en un proyecto simultáneamente sin temor a conflictos. Cada desarrollador puede trabajar en su propia rama y fusionar los cambios de manera eficiente.
2. **Historial y Auditoría:** El historial completo de cambios en un proyecto proporciona una trazabilidad clara y facilita la identificación de errores o problemas.
3. **Ramificación:** La capacidad de crear ramas fácilmente facilita el desarrollo de nuevas características y la corrección de errores sin afectar la rama principal del proyecto.
4. **Flexibilidad:** Git es flexible y se adapta a diversos flujos de trabajo, pequeños o grandes

# ¿Por qué debemos instalar los paquetes?

## Ayuda

En R, el signo de interrogación (?) se utiliza para acceder a la documentación de una función o de un paquete. Al colocar el signo de interrogación seguido del nombre de una función o paquete y ejecutar esa línea de código, R abrirá la documentación correspondiente en la ventana de ayuda.

```
?print()  
?lm #Busqueda Simple  
??lm #Busqueda Amplia
```

## Operadores

Los operadores son símbolos o palabras clave que realizan operaciones en variables y valores.

- ▶ Operadores de asignación:
- ▶ Operadores aritméticos
- ▶ Operadores de comparación

# ¿Por qué se muestra la segunda matriz y no la primera?

## Objetos | Matrices

### Manipulacion de Matrices

```
C.4 <- c(100, 1000) #crear vector para 4ta columna  
X1 <- cbind(X, C.4) #Agregar 4ta columna a X  
fila3 <- c(10, 100, 1000) #crear fila 3  
X2 <- rbind(X, fila3) #agregar fila 3 a X
```

X2

```
##           C.1 C.2  C.3  
## fila1    10  10   10  
## fila2    10  10   10  
## fila3    10 100 1000
```

## Objetos | Matrices

### Manipulacion de Matrices

# ¿Que tipos de datos hay en la matriz?

## Objetos | Manipulaciones

```
# Convertir objetos
```

```
data
```

```
##      id  peso tamaño  
## [1,] "1"  "50" "pequeño"  
## [2,] "2"  "90" "grande"  
## [3,] "3"  "75" "mediano"
```

```
data_1 <- as.numeric(data)
```

```
## Warning: NAs introduced by coercion
```

```
#Para convertir en Matriz
```

```
data_1
```

# tidyverse

tidyverse es un conjunto de paquetes de R diseñados para trabajar de manera integrada y coherente en el análisis de datos.



## Paquetes | tidyverse

Principales paquetes incluidos en tidyverse:

1. **ggplot2**: Para la creación de gráficos y visualizaciones.
2. **dplyr**: Para manipulación y transformación de datos.
3. **tidyr**: Para trabajar con datos en formato “tidy” (ordenados y estructurados).

# Practica

## Leer Datos

```
## La separación decimal en R es . y no ,  
## Datos en Millones de Guaranies  
Ingreso <- read.csv("Datos/IngresoPromPoblacion_py.csv")  
Ingreso <- read.table("Datos/IngresoPromPoblacion_py.csv",  
                      na.strings = "4/", sep= ";", check.names = FALSE)
```

Ingreso

	Departamento	1998	1999	2001	2002	2003	2004	2005
## 1	Alto Paraguay	NA	NA	NA	NA	NA	NA	NA
## 2	Alto Parana	3.467	2.716	2.829	2.345	2.392	2.512	2.512
## 3	Amambay	NA	NA	NA	NA	1.545	1.633	1.633
## 4	Asuncion	3.964	3.772	3.660	2.978	3.248	3.251	2.978
## 5	Boqueron	NA	NA	NA	NA	NA	NA	NA
## 6	Caaguazu	1.398	1.488	1.484	1.480	1.734	1.444	1.398
## 7	Caazapa	NA	NA	NA	NA	1.334	1.195	1.195



¿Que departamento tiene el mayor ingreso promedio?

¿Que departamento tiene el mayor ingreso promedio?

```
#Cambiar a datos largos
```

```
dato_largo <- Ingreso %>%
```

```
  gather(key = "year", value = "Cantidad", -Departamento) %>%
```

```
  mutate(year = as.integer(year))
```

```
dato_largo
```

##	Departamento	year	Cantidad
## 1	Alto Paraguay	1998	NA
## 2	Alto Parana	1998	3.467
## 3	Amambay	1998	NA
## 4	Asuncion	1998	3.964
## 5	Boqueron	1998	NA
## 6	Caaguazu	1998	1.398
## 7	Caazapa	1998	NA
## 8	Canindeyu	1998	NA