

Clase 1

Alexis Zúñiga & Ernesto Barrios

Capítulo 1: Instalar R

Vamos a empezar con los pasos para la instalación de R en los sistemas operativos Windows, MacOS y Linux.

Descargar R

1. Ve al sitio web oficial de R en <https://www.r-project.org/>
2. haz clic en donde diga **Download R**.
3. En la lista mirrors, busca <https://cran.itam.mx/> y hacer click.

continuamos en la sección correspondiente a tu sistema operativo:

Instalar R en Windows

1. Hacer click donde diga **Download for Windows** y guarda el archivo ejecutable
2. Corre el archivo .exe y seguir las instrucciones de instalación.

Instalar R en macOS

1. Hacer click donde diga **Download for macOS**
2. Hacer click en el link del archivo de la versión mas reciente de R
3. Correr el archivo .pkg y seguir las instrucciones de instalación.

Instalar R en Linux

1. Hacer click donde diga **Download for Linux**
 2. Selecciona tu distribución de Linux y sigue las instrucciones para instalar desde la terminal
-

Capítulo 2: ¿Para qué usar R?

Breve historia y usos

R es un software estadístico basado en los lenguajes *S* y *Scheme*. Por un lado, *S* es un lenguaje de programación desarrollado por Rick Bercker, John Chambers y colegas en Bell Labs en los años ochentas. Tiene como principal objetivo alentar y permitir el buen análisis estadístico.

En cuanto a R, todo comenzó en agosto de 1993 como un experimento realizado por Ross Ihaka y Robert Gentleman, ellos reconocieron la necesidad de un mejor ambiente de cálculo del que tenían. Ninguno de los productos comerciales les convencían por lo que decidieron crear una herramienta que les funcionara. Esto los llevo a tomar ideas de *S*, donde les resultó natural usar una sintaxis parecida. Por lo que, en el uso diario, *S* y R son muy similares, pero R no es *S*.

Las diferencias entre ellos son resultado de la herencia de *Scheme*, fundamentalmente el manejo de la memoria y el acceso a las variables dependiendo de dónde fueron definidas. También se distingue el manejo del color, áreas de graficación, rotulación matemática, entre otros.

Algunas fechas importantes en el desarrollo de R: en febrero de 2000 sale finalmente la versión 1.0 de R; en 2001 se publica el primer número de *R-News*, revista electrónica dedicada a la discusión y anuncios de nuevos procedimientos y paquetes de R; la revista es reemplazada por *R-Journal* en 2009. Finalmente, para garantizar que R sea siempre un software libre de código abierto, se creó *R-foundation* que entre sus objetivos están:

1. Avanzar el proyecto de R para cálculo estadístico que provee de software libre y código abierto para el análisis de datos y gráficas.
2. Guardar y administrar los derechos de copia de R y su documentación.¹

¿Qué podemos hacer con R?

- **Análisis y Visualización de Datos:** Procesamiento de datos, análisis estadísticos y visualización de datos.
- **Tareas Académicas:** Trabajos y tareas, y análisis de ejercicios estadísticos.
- **Programación y desarrollo:** Automatización de tareas y desarrollo web.
- **Educación y aprendizaje:** Aprendizaje de estadística y programación.

Cosas más avanzadas

- **Crea aplicaciones interactivas:** Puedes crear aplicaciones accesibles desde tu teléfono para interactuar con datos en tiempo real.
- **Análisis de Redes Sociales:** Analiza datos de redes sociales como Twitter utilizando paquetes como 'rtweet'. Puedes visualizar interacciones, hashtags populares y más.
- **Mapas Interactivos:** Utiliza 'leaflet' para crear mapas interactivos. Puedes trazar datos geoespaciales, añadir capas de información y más.
- **Uso de Sensores de tu Teléfono:** Utiliza aplicaciones como "Sensor Data" (Android) o "SensorLog" (iOS) para registrar datos de los sensores de tu teléfono y luego analiza esos datos en R.

¹Barrios, E.(2010). R: Un lenguaje para análisis de datos y graficación. Recuperado de: <https://gente.itam.mx/ebarrios/docs/porqueR.pdf>

Ejemplos

Grafico usando R básico

```
# Creación del Conjunto de Datos
clientes <- c(100, 150, 200, 130, 180, 170, 190, 210, 160, 120,
             130, 140, 230, 220, 240, 200, 210, 150, 180, 170,
             200, 190, 160, 180, 190, 210, 170, 140, 150, 180)

ventas <- c(20, 25, 30, 22, 27, 26, 28, 31, 24, 21,
           23, 24, 35, 34, 37, 31, 33, 25, 27, 26,
           30, 28, 25, 27, 28, 31, 26, 23, 25, 27)

# Estadísticas Descriptivas Básicas
promedio_clientes <- mean(clientes)
promedio_ventas <- mean(ventas)
desviacion_clientes <- sd(clientes)
desviacion_ventas <- sd(ventas)
resumen <- summary(data.frame(Clientes = clientes, Ventas = ventas))

# Imprimir resultados
cat("Promedio de clientes:", promedio_clientes, "\n")

## Promedio de clientes: 175

cat("Promedio de ventas:", promedio_ventas, "\n")

## Promedio de ventas: 27.3

cat("Desviación estándar de clientes:", desviacion_clientes, "\n")

## Desviación estándar de clientes: 33.80981

cat("Desviación estándar de ventas:", desviacion_ventas, "\n")

## Desviación estándar de ventas: 4.128538

cat("\nResumen de los datos:\n")

##
## Resumen de los datos:
print(resumen)

##      Clientes      Ventas
## Min.   :100   Min.   :20.0
## 1st Qu.:150   1st Qu.:25.0
## Median :180   Median :27.0
## Mean   :175   Mean   :27.3
## 3rd Qu.:200   3rd Qu.:30.0
## Max.   :240   Max.   :37.0

# Gráfico de Dispersión
plot(clientes, ventas,
     main = "Relación entre Número de Clientes y Ventas",
     xlab = "Número de Clientes",
     ylab = "Ventas (en miles de dólares)",
     pch = 19, col = "blue")
```

```
# Agregar una línea de tendencia  
abline(lm(ventas ~ clientes), col = "red", lwd = 2)
```

Relación entre Número de Clientes y Ventas

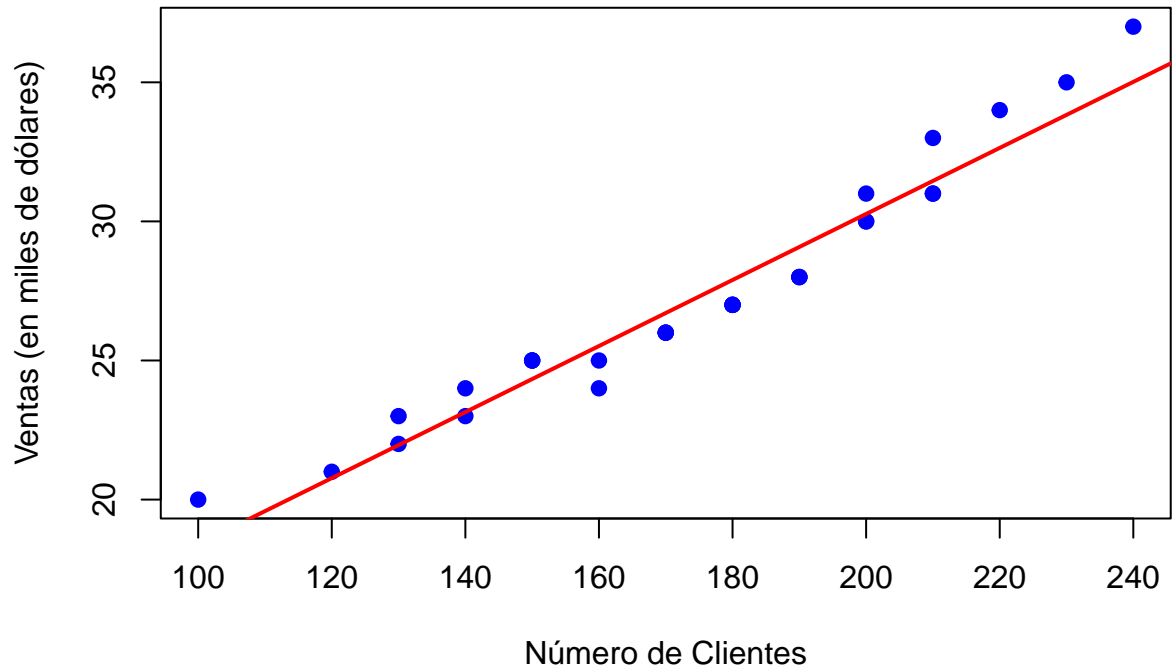


Gráfico de dispersión

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

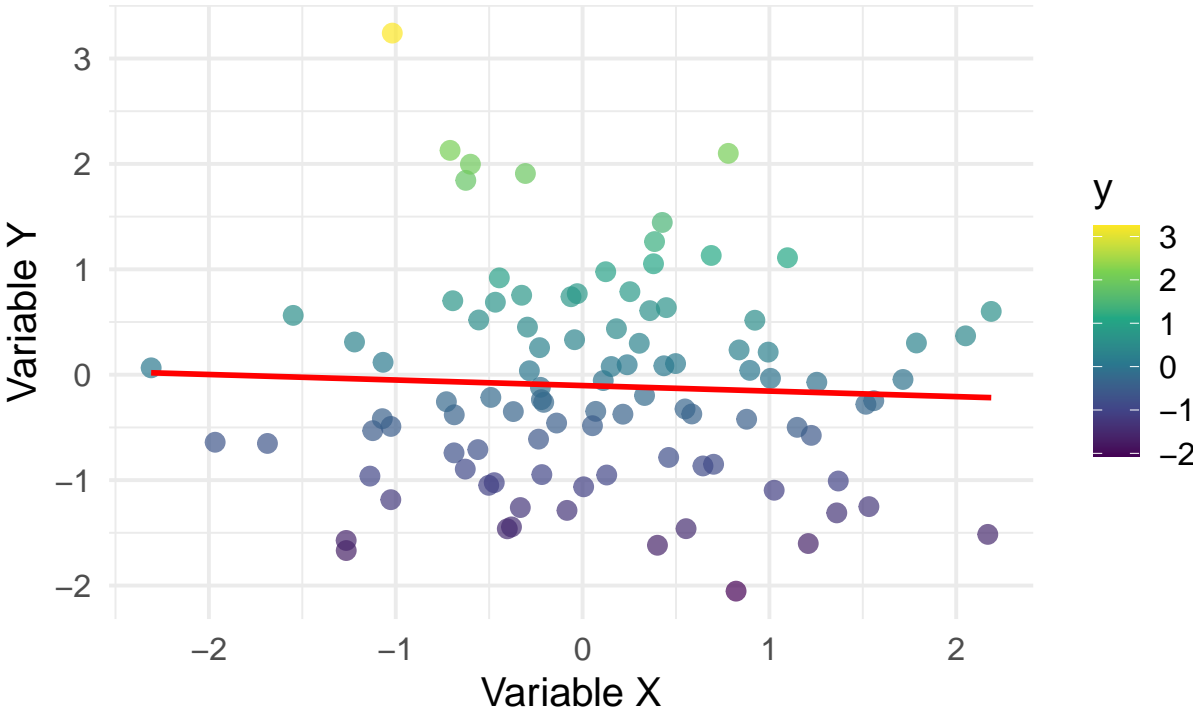
set.seed(123)
data <- data.frame(
  x = rnorm(100),
  y = rnorm(100)
)

p <- ggplot(data, aes(x = x, y = y)) +
  geom_point(aes(color = y), size = 3, alpha = 0.7) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(
    title = "Gráfico de Dispersión con Línea de Ajuste",
    x = "Variable X",
    y = "Variable Y",
    caption = "Fuente: Datos simulados"
  ) +
  theme_minimal(base_size = 15) +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 1)
  ) +
  scale_color_viridis_c(option = "D")

print(p)

## `geom_smooth()` using formula = 'y ~ x'
```

Gráfico de Dispersión con Línea de Ajuste



Fuente: Datos simulados

Gráfico de Violín

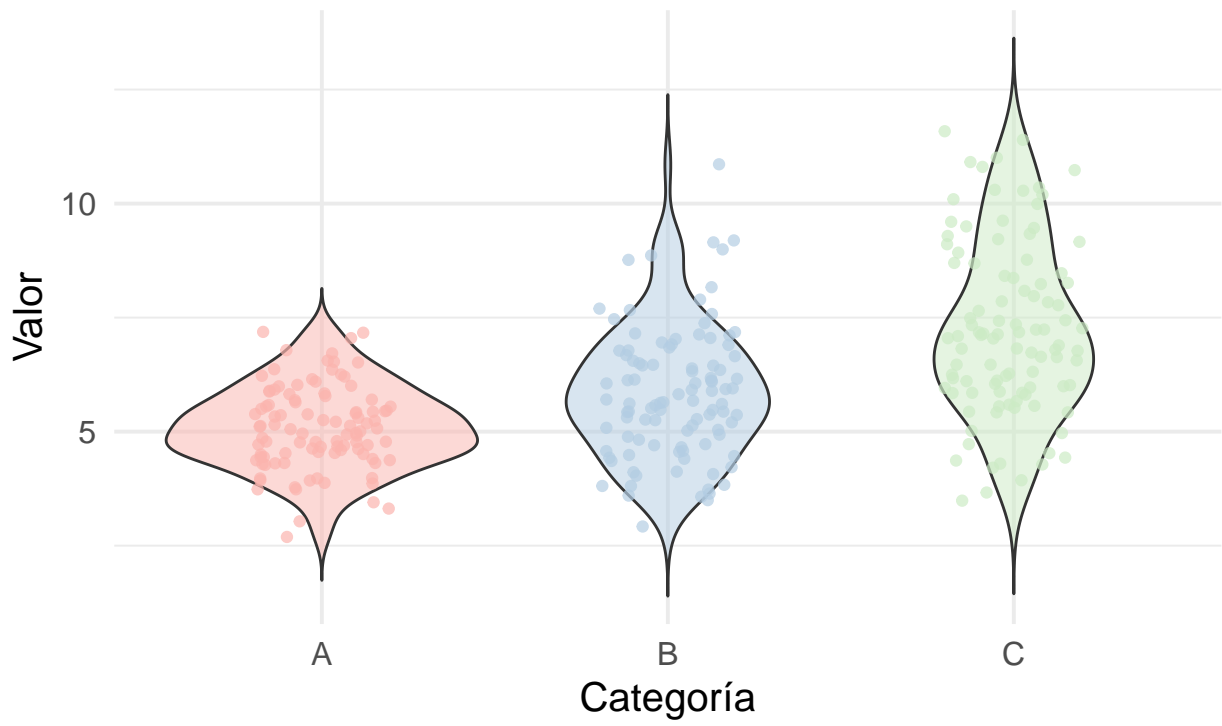
```
library(ggplot2)
library(dplyr)

set.seed(123)
data <- data.frame(
  category = factor(rep(c("A", "B", "C"), each = 100)),
  value = c(rnorm(100, mean = 5, sd = 1),
            rnorm(100, mean = 6, sd = 1.5),
            rnorm(100, mean = 7, sd = 2))
)

p <- ggplot(data, aes(x = category, y = value, fill = category)) +
  geom_violin(trim = FALSE, alpha = 0.5) +
  geom_jitter(width = 0.2, size = 1.5, alpha = 0.7, aes(color = category)) +
  labs(
    title = "Gráfico de Violín con Puntos Superpuestos",
    x = "Categoría",
    y = "Valor",
    caption = "Fuente: Datos simulados"
  ) +
  theme_minimal(base_size = 15) +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    plot.caption = element_text(hjust = 1),
    legend.position = "none"
  ) +
  scale_fill_brewer(palette = "Pastel1") +
  scale_color_brewer(palette = "Pastel1")

# Mostrar la gráfica
print(p)
```


Gráfico de Violín con Puntos Superpuestos



Fuente: Datos simulados

Gráfico de líneas

```
library(ggplot2)
library(dplyr)

set.seed(123)
date_seq <- seq(as.Date("2023-01-01"), by = "month", length.out = 12)
data <- data.frame(
  date = rep(date_seq, 3),
  value = c(cumsum(rnorm(12, mean = 5, sd = 2)),
            cumsum(rnorm(12, mean = 7, sd = 2)),
            cumsum(rnorm(12, mean = 6, sd = 2))),
  category = factor(rep(c("A", "B", "C"), each = 12))
)

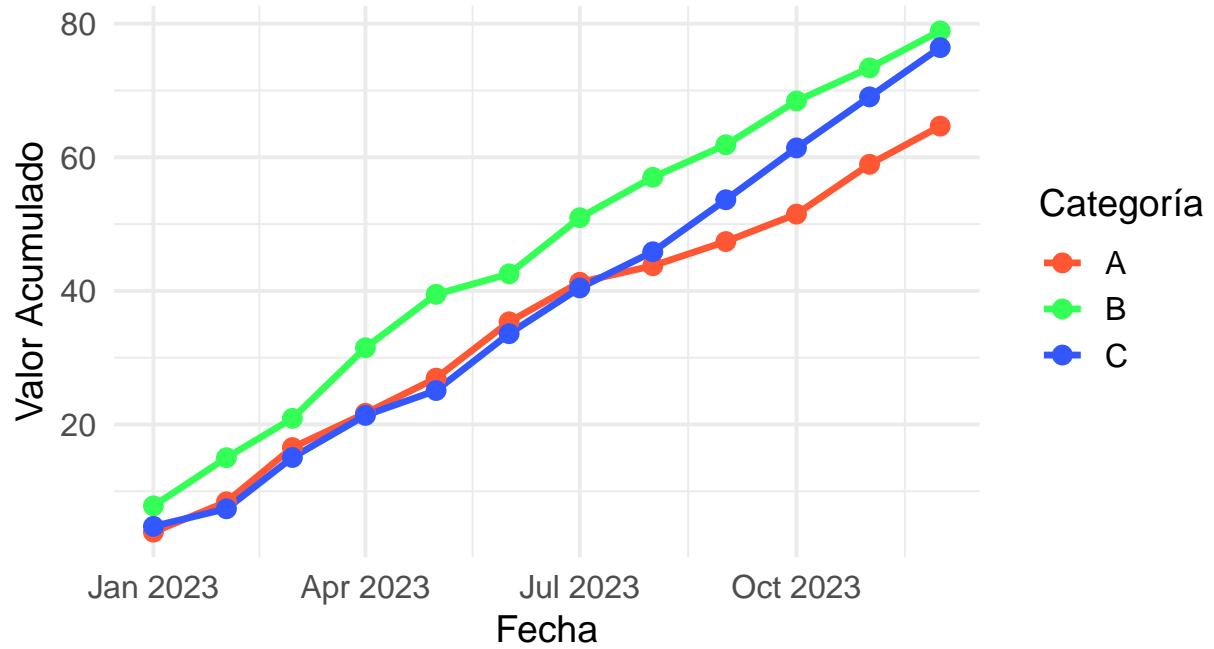
p <- ggplot(data, aes(x = date, y = value, color = category, group = category)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(
    title = "Tendencia Temporal de Múltiples Series",
    subtitle = "Ejemplo de visualización de datos temporales con ggplot2",
    x = "Fecha",
    y = "Valor Acumulado",
    color = "Categoría",
    caption = "Fuente: Datos simulados"
  ) +
  theme_minimal(base_size = 15) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, size = 16),
    plot.caption = element_text(hjust = 1, size = 12),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12),
    legend.title = element_text(size = 14),
    legend.text = element_text(size = 12)
  ) +
  scale_color_manual(values = c("A" = "#FF5733", "B" = "#33FF57", "C" = "#3357FF"))

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

print(p)
```

Tendencia Temporal de Múltiples Series

Ejemplo de visualización de datos temporales con ggplot2



Fuente: Datos simulados

Capítulo 3: Abrir R por primera vez.

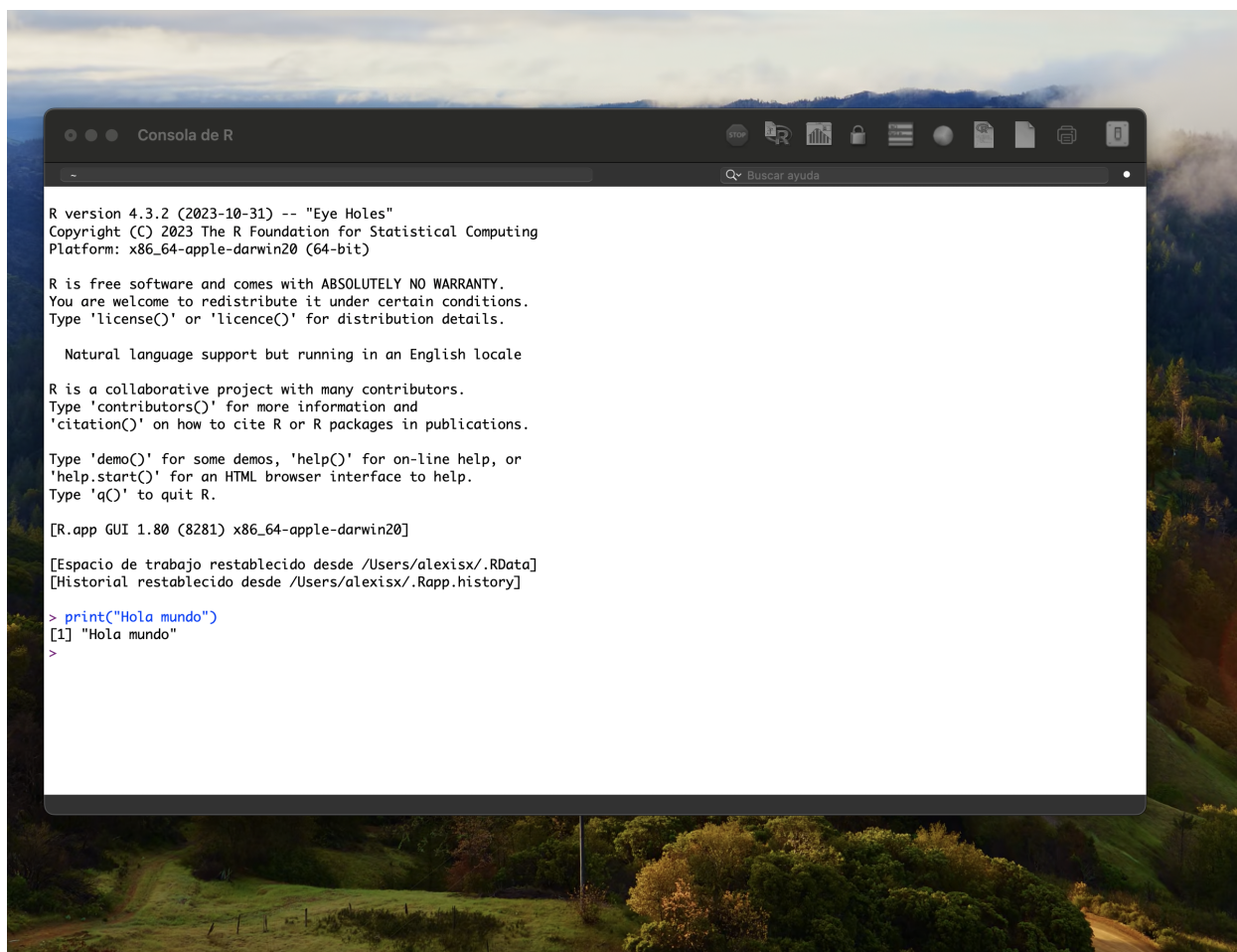
Una vez que hayamos instalado R en nuestro sistema operativo, abramos la consola y ejecutemos nuestro primer comando paso a paso.

1. Localiza donde está guardado el programa de R y ábrelo. Notese que una vez abierto nos abra una consola donde se puede empezar a escribir comandos y generar respuestas.
2. Una vez abierta la consola ejecute el siguiente comando:

```
print("Hola mundo")
```

```
## [1] "Hola mundo"
```

Note como en este ejemplo, al ejecutar la función `print()` y escribir el texto entre comillas *"Hola mundo"* la consola unicamente nos regresa el texto dentro de la función.



Con esto empezamos nuestro camino para usar R como una herramienta :)

Agradecimientos especiales y bibliografía

Quiero expresar mi sincero agradecimiento a Santiago Casanova por su contribución. Sus notas y sugerencias han sido fundamentales en la mejora de este documento.

Casanova, S. & Barrios, E. (2021) Introducción_a_R. Recuperado de:https://gente.itam.mx/ebarrios/cursoR/Agosto2022/Introduccion_a_R.pdf

Barrios, E.(2010). R: Un lenguaje para análisis de datos y graficación. Recuperado de:<https://gente.itam.mx/ebarrios/docs/porqueR.pdf>

Hernández, F. & Usuga, O.(2024) Manual de R. Recuperado de: <https://fhernanb.github.io/Manual-de-R/>