

# Series de tiempo univariadas - Presentación 1

Mauricio Alejandro Mazo Lopera

Universidad Nacional de Colombia  
Facultad de Ciencias  
Escuela de Estadística  
Medellín



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

En **Análisis de Regresión** se planteó que:

- Se toma una **muestra aleatoria** de tamaño  $n$ .

En **Análisis de Regresión** se planteó que:

- Se toma una **muestra aleatoria** de tamaño  $n$ .
- El interés se centra en una **variable aleatoria dependiente**, cuyas mediciones para los  $n$  individuos de la muestra, se ubica en un vector  $(n \times 1)$ :  $\mathbf{Y}_n$

En **Análisis de Regresión** se planteó que:

- Se toma una **muestra aleatoria** de tamaño  $n$ .
- El interés se centra en una **variable aleatoria dependiente**, cuyas mediciones para los  $n$  individuos de la muestra, se ubica en un vector ( $n \times 1$ ):  $\mathbf{Y}_n$
- Se tienen un conjunto de  $p$  covariables, medidas en los  $n$  individuos y que se ubica en una matriz conocida como **matriz de diseño** de dimensiones ( $n \times (p + 1)$ ):  $\mathbf{X}_{n \times (p+1)}$

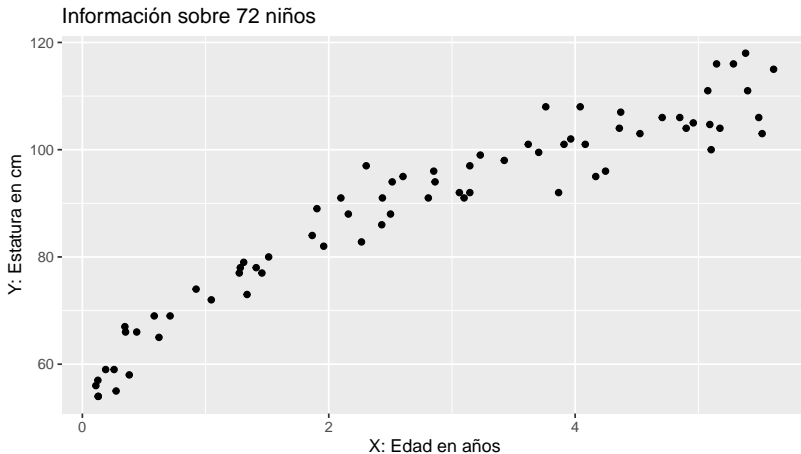
En **Análisis de Regresión** se planteó que:

- Se toma una **muestra aleatoria** de tamaño  $n$ .
- El interés se centra en una **variable aleatoria dependiente**, cuyas mediciones para los  $n$  individuos de la muestra, se ubica en un vector ( $n \times 1$ ):  $\mathbf{Y}_n$
- Se tienen un conjunto de  $p$  covariables, medidas en los  $n$  individuos y que se ubica en una matriz conocida como **matriz de diseño** de dimensiones ( $n \times (p + 1)$ ):  $\mathbf{X}_{n \times (p+1)}$
- Se plantea **el modelo**:  $\mathbf{Y}_n = \mathbf{X}_{n \times (p+1)}\beta_{p+1} + \varepsilon_n$ . Donde  $\beta_{p+1}$  es el vector de parámetros y  $\varepsilon_n$  es el vector de errores.

En **Análisis de Regresión** se planteó que:

- Se toma una **muestra aleatoria** de tamaño  $n$ .
- El interés se centra en una **variable aleatoria dependiente**, cuyas mediciones para los  $n$  individuos de la muestra, se ubica en un vector ( $n \times 1$ ):  $\mathbf{Y}_n$
- Se tienen un conjunto de  $p$  covariables, medidas en los  $n$  individuos y que se ubica en una matriz conocida como **matriz de diseño** de dimensiones ( $n \times (p + 1)$ ):  $\mathbf{X}_{n \times (p+1)}$
- Se plantea **el modelo**:  $\mathbf{Y}_n = \mathbf{X}_{n \times (p+1)}\beta_{p+1} + \varepsilon_n$ . Donde  $\beta_{p+1}$  es el vector de parámetros y  $\varepsilon_n$  es el vector de errores.
- Los **supuestos** son:
  - a. El vector de errores  $\varepsilon_n \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ .
  - b. Los errores de los individuos en la muestra  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  son **INDEPENDIENTES**.

# Ejemplo introductorio: 72 niños (puntos) diferentes e independientes



En **Series de Tiempo Univariadas** tendremos en cuenta que:

- A menudo NO se puede garantizar la independencia entre las observaciones, sobre todo cuando se tienen variables cuya evolución se mide a lo largo del tiempo, esto es, cuando se tienen de por medio variables que son **series de tiempo**.



En **Serie de Tiempo Univariadas** tendremos en cuenta que:

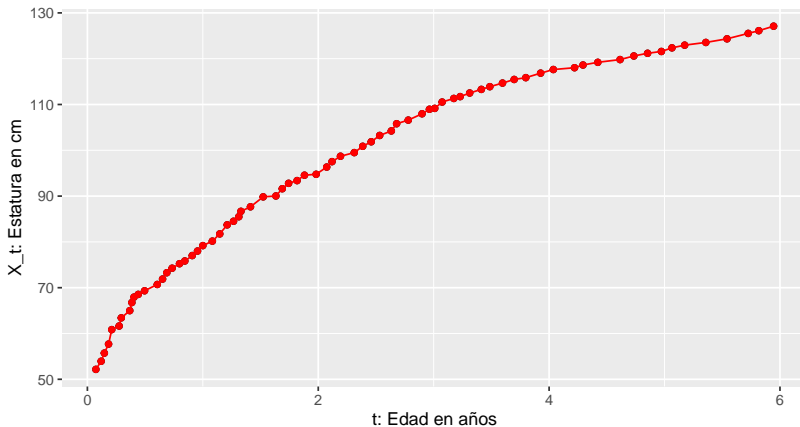
- A menudo NO se puede garantizar la independencia entre las observaciones, sobre todo cuando se tienen variables cuya evolución se mide a lo largo del tiempo, esto es, cuando se tienen de por medio variables que son **series de tiempo**.
- Explicaremos la serie de tiempo a través de “ella misma” y posiblemente del tiempo.

En **Series de Tiempo Univariadas** tendremos en cuenta que:

- A menudo NO se puede garantizar la independencia entre las observaciones, sobre todo cuando se tienen variables cuya evolución se mide a lo largo del tiempo, esto es, cuando se tienen de por medio variables que son **series de tiempo**.
- Explicaremos la serie de tiempo a través de “ella misma” y posiblemente del tiempo.
- Un análisis descriptivo sobre el comportamiento de la serie a lo largo del tiempo es fundamental para poder desarrollar un método que la modele.

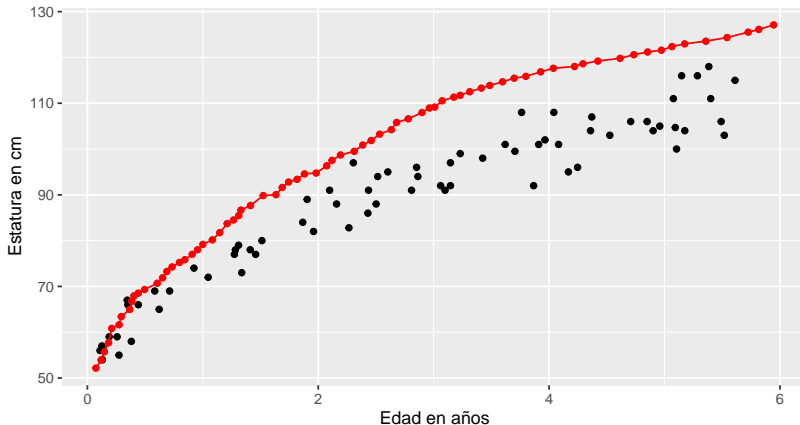
# Ejemplo introductorio: 72 puntos (no independientes) de un solo niño

Información sobre 72 mediciones de un solo niño



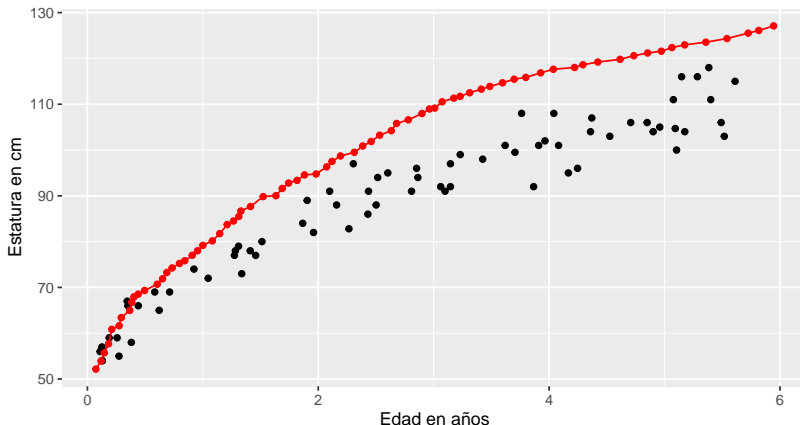
# Ejemplo introductorio: Contrastando ambos casos

Información sobre 72 niños versus 72 datos de un solo niño



# Ejemplo introductorio: Contrastando ambos casos

Información sobre 72 niños versus 72 datos de un solo niño



**A continuación veremos algunos ejemplos de series de tiempo y describiremos sus características:**

# Introducción: Ejemplo 1 - Precios de acciones



# Introducción: Ejemplo 1 - Precios de acciones

Considere los precios diarios de la acción Bancolombia entre el 3 de junio de 2020 al 2 de julio de 2020 (para obtener estos datos basta con escribir Bancolombia Yahoo Finance en Google):

```
BD1<-read.csv("../DATOS/BANC_2020_JUN_JUL.csv", header=TRUE)
```

Las variables en esta base de datos son:

- **Date:** Fecha en formato Año-Mes-Día.
- **Open:** Precio de la acción en el momento en que abre el mercado.
- **High:** Máximo precio alcanzado en un día.
- **Low:** Mínimo precio alcanzado en un día.
- **Close:** Precio de cierre al cerrar el mercado. Esta variable suele ser la más importante.
- **Adj. Close:** Precio de cierre ajustado por “splits” y dividendos.
- **Volume:** Número de transacciones realizadas.

# Introducción: Ejemplo 1 - Precios de acciones

```
require(magrittr) # Paquete para usar los pipes
```

```
BD1 %>% dim()
```

```
## [1] 22 7
```

```
BD1 %>% head(3)
```

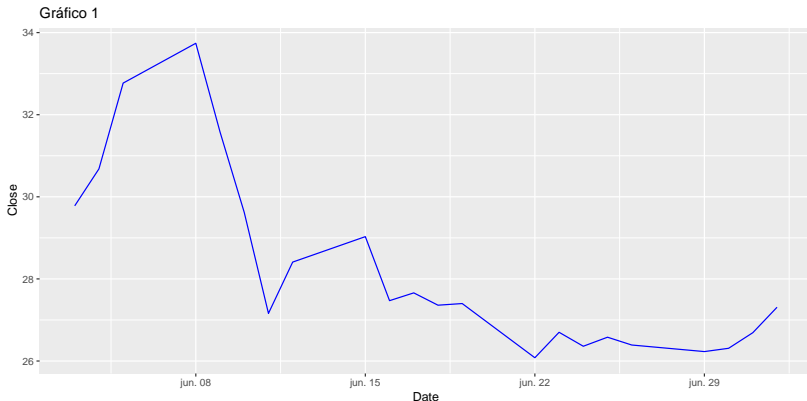
```
##           Date  Open  High   Low Close Adj.Close  Volume
## 1 2020-06-03 28.36 29.85 28.36 29.78 29.42115 575900
## 2 2020-06-04 29.60 30.72 29.23 30.68 30.31030 575200
## 3 2020-06-05 32.37 33.19 31.59 32.77 32.37512 1068200
```

```
BD1 %>% tail(3)
```

```
##           Date  Open  High   Low Close Adj.Close  Volume
## 20 2020-06-30 25.96 26.51 25.73 26.31 26.31 397400
## 21 2020-07-01 26.51 27.26 26.22 26.69 26.69 598400
## 22 2020-07-02 27.54 27.76 26.99 27.31 27.31 351200
```



# Introducción: Ejemplo 1 - Precios de acciones

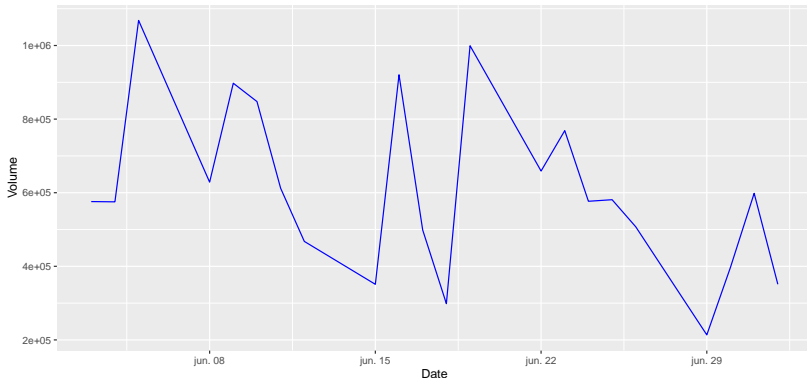


## CÓDIGOS

En este gráfico podemos observar que el precio de la acción Bancolombia presenta un crecimiento hasta el 8 de junio de 2020 y luego una **tendencia** decreciente muy marcada con una leve recuperación al final.

# Introducción: Ejemplo 1 - Precios de acciones

Gráfico 2



## CÓDIGOS

Se observa un **cambio en la variabilidad** de los volúmenes de la acción siendo mayor entre el 15 y el 22 de junio de 2020.

## Introducción: Ejemplo 2 - Precipitaciones



**Fuente:** Corporación Autónoma Regional de Cundinamarca - CAR. Base de datos original disponible [AQUÍ](#)

# Introducción: Ejemplo 2 - Precipitaciones

```
BD2<-read.csv("../..DATOS/Precipitaciones_Totales_Mensuales.csv",  
              header=TRUE,fileEncoding = "utf8")
```

```
BD2 %>% head()
```

```
##      ESTACION CODIGO   X.N   Y.E MUNICIPIO   CUENCA ANIO ENERO FEBRERO  
## 1 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 1995   0.0    0.0  
## 2 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 1996 295.9  127.2  
## 3 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 1997 294.2   76.9  
## 4 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 1998 144.3  118.0  
## 5 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 1999  78.3  256.6  
## 6 Pajas Blancas 2119022 973320 944710      Nilo R. Paguey 2000 112.6  304.8  
##   MARZO ABRIL  MAYO  JUNIO  JULIO  AGOSTO SEPTIEMBRE OCTUBRE NOVIEMBRE DICIEMBRE  
## 1   0.0   0.0   0.0   0.0   0.0   0.0       31.5   190.0       153.6       76.2  
## 2 247.5 149.2   0.0   0.0   0.0   0.0       0.0    0.0       0.0      115.2  
## 3 160.2 296.3 144.1 134.3  20.1   3.6       75.3   174.0      182.0      110.1  
## 4 209.1 438.3 292.6  30.6  35.2  30.7      103.2  385.9      105.4      235.7  
## 5  83.0 339.3 108.2 236.2  17.6  23.7      322.9  337.9      234.4      199.5  
## 6 201.8 322.7 338.9 199.5  37.5  74.8      218.3  244.3      190.9      115.5
```

```
BD2 %>% dim()
```

```
## [1] 6094   19
```

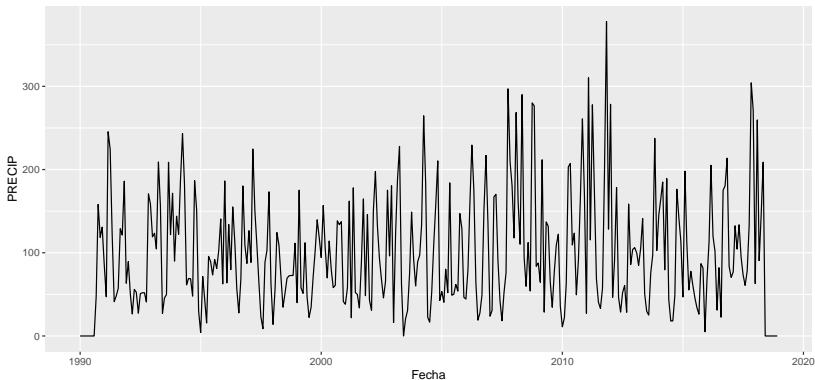
# Introducción: Ejemplo 2 - Precipitaciones - Mesitas

```
BD2 %>% filter(ESTACION=="Mesitas") %>% head(5)
```

|      | ESTACION | CODIGO  | X.N    | Y.E    | MUNICIPIO     | CUENCA     | ANIO    | ENERO     | FEBRERO   | MARZO |
|------|----------|---------|--------|--------|---------------|------------|---------|-----------|-----------|-------|
| ## 1 | Mesitas  | 2120646 | 997800 | 959500 | El Colegio R. | Bogotá     | 2005    | 53.4      | 40.5      | 80.2  |
| ## 2 | Mesitas  | 2120646 | 997800 | 959500 | El Colegio R. | Bogotá     | 2001    | 38.1      | 58.0      | 162.1 |
| ## 3 | Mesitas  | 2120646 | 997800 | 959500 | El Colegio R. | Bogotá     | 2007    | 23.8      | 30.6      | 167.0 |
| ## 4 | Mesitas  | 2120646 | 997800 | 959500 | El Colegio R. | Bogotá     | 2013    | 106.2     | 99.6      | 84.9  |
| ## 5 | Mesitas  | 2120646 | 997800 | 959500 | El Colegio R. | Bogotá     | 1993    | 119.1     | 123.5     | 104.6 |
| ##   | ABRIL    | MAYO    | JUNIO  | JULIO  | AGOSTO        | SEPTIEMBRE | OCTUBRE | NOVIEMBRE | DICIEMBRE |       |
| ## 1 | 52.0     | 184.0   | 49.0   | 50.1   | 61.9          | 53.9       | 147.1   | 129.0     | 46.5      |       |
| ## 2 | 21.9     | 178.0   | 52.2   | 50.0   | 33.7          | 88.7       | 164.7   | 48.4      | 146.0     |       |
| ## 3 | 170.2    | 90.8    | 42.8   | 18.3   | 52.0          | 76.2       | 297.0   | 209.5     | 180.7     |       |
| ## 4 | 105.3    | 141.2   | 50.6   | 29.8   | 25.1          | 75.8       | 97.4    | 237.5     | 102.7     |       |
| ## 5 | 209.1    | 156.5   | 27.1   | 45.5   | 49.9          | 208.7      | 121.7   | 171.5     | 90.0      |       |

# Introducción: Ejemplo 2 - Precipitaciones - Mesitas

Gráfico 3

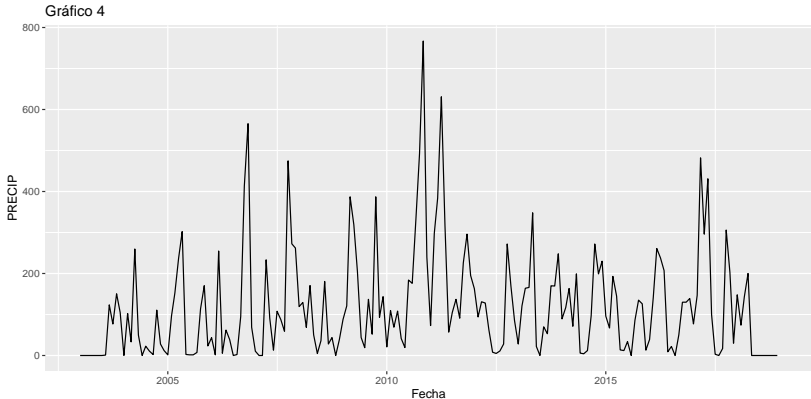


## CÓDIGOS

Vemos que los valores de precipitación promedio mensual **oscilan alrededor** de 100 mm, presentando contrastes de menor y mayor **variabilidad** en algunos periodos. Además de lo anterior, se “presume” que hay cierta repetición anual o **ciclos** de la precipitación:

# Introducción: Ejemplo 2 - Precipitaciones - Mesitas

# Introducción: Ejemplo 2 - Precipitaciones - San Isidro



## CÓDIGOS

Vemos un comportamiento muy marcado por picos que se presentan luego de estar en cero. Además, se presumen ciertos **ciclos** anuales:



# Introducción: Ejemplo 2 - Precipitaciones - San Isidro

## Introducción: Ejemplo 3 - Número de pasajeros anualmente

Una base de datos clásica para ilustrar los distintos componentes que pueden aparecer en una serie de tiempo es la que se relaciona con el número de pasajeros (dados en miles) mensuales que se presentaron en cierta aerolínea desde 1949 hasta 1960. Para obtener información sobre esta BD:

```
?AirPassengers
```

Esta BD tiene además una estructura especial en R, conocida como **ts** (time series, para obtener más información escriba en R **?ts**):


```
AirPassengers %>% class()
```

```
## [1] "ts"
```

En el siguiente gráfico se observa una clara **tendencia** creciente y un comportamiento **cíclico** o **periódico** anual:

# Introducción: Ejemplo 3 - Número de pasajeros anualmente

# Introducción: Ejemplo 4 - Tasas de suicidios mundiales



The screenshot shows the Kaggle dataset page for "Suicide Rates Overview 1985 to 2016". The header features a dark background with a forest image. The title "Suicide Rates Overview 1985 to 2016" is prominently displayed, followed by the subtitle "Compares socio-economic info with suicide rates by year and country". The dataset is credited to "Rusty" and was updated 2 years ago (Version 1). Below the header, there are tabs for "Data", "Tasks (9)", "Kernels (343)", "Discussion (24)", "Activity", and "Metadata". A "Download (3 MB)" button and a "New Notebook" button are also visible. At the bottom, the "Usability" is rated 8.2, the "License" is "World Bank Dataset Terms of Use", and the "Tags" include "news, sensitive subjects, economics, world, demographics".

Dataset

## Suicide Rates Overview 1985 to 2016

Compares socio-economic info with suicide rates by year and country

Rusty • updated 2 years ago (Version 1)

Data Tasks (9) Kernels (343) Discussion (24) Activity Metadata

Download (3 MB) New Notebook

Usability 8.2 License World Bank Dataset Terms of Use

Tags news, sensitive subjects, economics, world, demographics

**Fuente:** Organización mundial de la salud. Base de datos original disponible AQUÍ

# Introducción: Ejemplo 4 - Tasas de suicidios mundiales

```
# read_csv es una función del paquete readr  
BD3<-read_csv("../..//DATOS/master.csv")
```

```
BD3 %>% names()
```

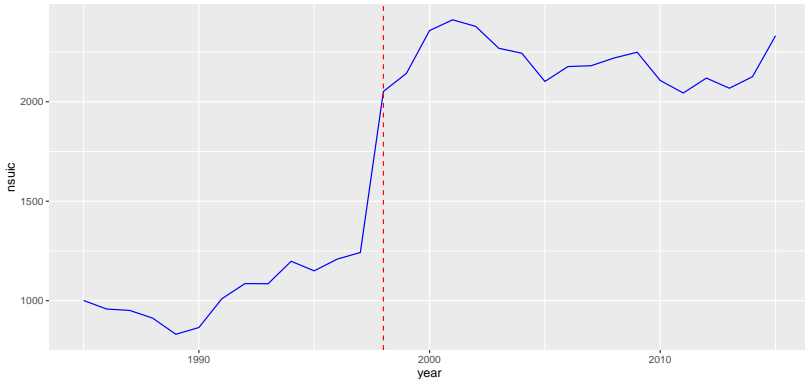
```
## [1] "country"          "year"              "sex"  
## [4] "age"              "suicides_no"        "population"  
## [7] "suicides/100k pop" "country-year"        "HDI for year"  
## [10] "gdp_for_year ($)" "gdp_per_capita ($)" "generation"
```

```
BD3 %>% dim()
```

```
## [1] 27820    12
```

# Introducción: Ejemplo 4 - Tasas de suicidios mundiales

Gráfico 5 – Colombia



## CÓDIGOS

Se observan dos momentos con **punto de ruptura** cerca del año 1998. Allí se ve un aumento en el total de suicidios de casi mil, cuando se observan los datos antes y después de 1998, evidenciando que los datos **oscilan alrededor de** dos **valores diferentes** antes y después de dicho punto.

# Actividad Presentación 1:

- 1 Repetir los códigos de esta presentación en sus computadores mediante el uso del RStudio.
- 2 Graficar las series de tiempo de al menos otras tres estaciones en la base de datos de precipitaciones del ejemplo 2.
- 3 Graficar la evolución del total de suicidios y de la población de al menos otros tres países en la base de datos de tasas de suicidios del ejemplo 3.
- 4 Ingrese a [Yahoo Finance](#) y descargue la cotización diaria de Bancolombia (en el mercado de EEUU) de los últimos 2 años. Mediante RStudio, realice un análisis descriptivo sobre el comportamiento de dicha serie. Saque al menos tres conclusiones al respecto.





```
require(magrittr) # Paquete para usar los pipes
BD1$Date %<>% as.Date() # Date a formato Fecha:

require(tidyverse) # Cargamos el tidyverse
ggplot(BD1, aes(x=Date, y=Close))+
  geom_line(col="blue") + labs(title="Gráfico 1")
```

VOLVER

```
ggplot(BD1, aes(x=Date, y=Volume))+  
  geom_line(col="blue") + labs(title="Gráfico 2")
```

VOLVER

```
BD2_1 <- filter(BD2, ESTACION=="Mesitas")
BD2_2 <- gather(BD2_1, "ENERO", "FEBRERO", "MARZO",
  "ABRIL", "MAYO", "JUNIO", "JULIO", "AGOSTO",
  "SEPTIEMBRE", "OCTUBRE", "NOVIEMBRE",
  "DICIEMBRE", key="MES", value="PRECIP")

head(BD2_2)
BD2_2 <- gather(BD2_1, "ENERO", "FEBRERO", "MARZO", "ABRIL",
  "MAYO", "JUNIO", "JULIO", "AGOSTO",
  "SEPTIEMBRE", "OCTUBRE", "NOVIEMBRE",
  "DICIEMBRE", key="MES", value="PRECIP")

head(BD2_2)
dim(BD2_2)
BD2_3 <- BD2_2[order(BD2_2$ANIO),]
head(BD2_3)
BD2_3$Fecha <- paste(rep(1,nrow(BD2_3)),BD2_3$MES,BD2_3$ANIO)
BD2_3$Fecha %<>% as.Date(format="%d %B %Y")
head(BD2_3)
```

VOLVER

```
BD2_1 <- filter(BD2, ESTACION=="San Isidro")

BD2_2 <- gather(BD2_1, "ENERO", "FEBRERO", "MARZO", "ABRIL",
               "MAYO", "JUNIO", "JULIO", "AGOSTO",
               "SEPTIEMBRE", "OCTUBRE", "NOVIEMBRE",
               "DICIEMBRE", key="MES", value="PRECIP")

BD2_3 <- BD2_2[order(BD2_2$ANIO),]

BD2_3$Fecha <- paste(rep(1,nrow(BD2_3)),
                    BD2_3$MES,BD2_3$ANIO)

BD2_3$Fecha %<>% as.Date(format="%d%B%Y")
ggplot(BD2_3, aes(x=Fecha, y=PRECIP))+
  geom_line(col="black") + labs(title="Gráfico 4")
```

VOLVER

```
BD3_1<-filter(BD3, country=="Colombia")
BD3_2 <- BD3_1 %>% group_by(year) %>%
  summarise(nsuic=sum(suicides_no), pob=sum(population))
ggplot(BD3_2, aes(x=year, y=nsuic))+
  geom_line(col="blue") + labs(title="Gráfico 5")+
  geom_vline(xintercept = 1998, col="red", lty=2)
```

VOLVER