

Series de tiempo univariadas - Presentación 2

Mauricio Alejandro Mazo Lopera
Universidad Nacional de Colombia
Facultad de Ciencias
Escuela de Estadística
Medellín



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Una visión muy **práctica** de lo que es una serie de tiempo consiste en definirla como **una secuencia ordenada de observaciones**, aunque más adelante veremos una definición más formal.

Una visión muy **práctica** de lo que es una serie de tiempo consiste en definirla como **una secuencia ordenada de observaciones**, aunque más adelante veremos una definición más formal.

El **análisis de series de tiempo** se refiere al conjunto de **metodologías estadísticas** para describir y analizar una serie de tiempo.

Son múltiples las áreas del conocimiento donde encontramos series de tiempo. Algunos ejemplos son:

Son múltiples las áreas del conocimiento donde encontramos series de tiempo. Algunos ejemplos son:

- **Agricultura:** Donde podemos encontrar **producción anual** de productos o cosechas, **precios mensuales promedio** de diferentes productos, entre otros.
- **Economía y finanzas:** **Precios de cierre** de acciones, **tasas de cambio** semanales, **Precios mensuales de índices** de la bolsa del mercado, etc.

Son múltiples las áreas del conocimiento donde encontramos series de tiempo. Algunos ejemplos son:

- **Agricultura:** Donde podemos encontrar **producción anual** de productos o cosechas, **precios mensuales promedio** de diferentes productos, entre otros.
- **Economía y finanzas:** **Precios de cierre** de acciones, **tasas de cambio** semanales, **Precios mensuales de índices** de la bolsa del mercado, etc.
- **Ingeniería:** **Señales** de **sonido** o **intensidad** de las mismas con respecto a la fuente que los genera, **señales eléctricas** y de **voltaje**, etc.

- **Meteorología:** Velocidad del viento por hora en un día, temperaturas diarias, precipitaciones semanales, etc.

- **Meteorología:** Velocidad del viento por hora en un día, temperaturas diarias, precipitaciones semanales, etc.
- **Medicina:** Electroencefalogramas, electrocardiogramas, etc.

- **Meteorología:** Velocidad del viento por hora en un día, temperaturas diarias, precipitaciones semanales, etc.
- **Medicina:** Electroencefalogramas, electrocardiogramas, etc.
- **Control de calidad:** Monitoreos de un proceso a lo largo del tiempo, etc.

- **Meteorología:** Velocidad del viento por hora en un día, temperaturas diarias, precipitaciones semanales, etc.
- **Medicina:** Electroencefalogramas, electrocardiogramas, etc.
- **Control de calidad:** Monitoreos de un proceso a lo largo del tiempo, etc.
- **Ciencias sociales:** Tasas de mortalidad, natalidad, crecimiento, desempleo, etc.

Una serie de tiempo puede ser:

- **Continua:** Cuando se mide continuamente, como por ejemplo una señal eléctrica que se mide sin parar por un periodo de una hora.

Una serie de tiempo puede ser:

- **Continua:** Cuando se mide continuamente, como por ejemplo una señal eléctrica que se mide sin parar por un periodo de una hora.
- **Discreta:** Cuando se mide en ciertos intervalos, como por ejemplo tasas de cambio diarias, las cuales se miden o se registran cada día.

Con respecto al **caso discreto**, los intervalos suelen ser iguales o equidistantes cada unidad de tiempo, es decir, cada día, cada mes, cada año, etc.

También podrían considerarse **otras unidades distintas al tiempo**, como por ejemplo la longitud y en este caso los intervalos equidistantes podrían ser cada 10 centímetros, cada metro, etc.

NOTA: En este curso nos enfocaremos en el caso de las series discretas.

El estudio de las series de tiempo puede abarcar diversos **objetivos**, entre los cuales podemos considerar:

El estudio de las series de tiempo puede abarcar diversos **objetivos**, entre los cuales podemos considerar:

- Entender y describir el mecanismo que genera la serie de tiempo.

El estudio de las series de tiempo puede abarcar diversos **objetivos**, entre los cuales podemos considerar:

- Entender y describir el mecanismo que genera la serie de tiempo.
- Pronosticar valores futuros.

El estudio de las series de tiempo puede abarcar diversos **objetivos**, entre los cuales podemos considerar:

- Entender y describir el mecanismo que genera la serie de tiempo.
- Pronosticar valores futuros.
- Control óptimo del sistema que está detrás de la serie de tiempo.

El estudio de las series de tiempo puede abarcar diversos **objetivos**, entre los cuales podemos considerar:

- Entender y describir el mecanismo que genera la serie de tiempo.
- Pronosticar valores futuros.
- Control óptimo del sistema que está detrás de la serie de tiempo.

En este curso nos enfocaremos en ver las distintas técnicas para alcanzar estos objetivos y otros más que iremos proponiendo.

Las técnicas descriptivas en el contexto de series de tiempo consisten resúmenes estadísticos y herramientas gráficas para extraer información referente a **patrones, tendencias, distribuciones, ciclos y relaciones** con otras variables que permitan aprender más acerca del pasado y la evolución de la serie.

Las técnicas descriptivas en el contexto de series de tiempo consisten resúmenes estadísticos y herramientas gráficas para extraer información referente a **patrones, tendencias, distribuciones, ciclos y relaciones** con otras variables que permitan aprender más acerca del pasado y la evolución de la serie.

Así, cuando realizamos un análisis descriptivo de una serie de tiempo buscamos identificar los siguientes componentes:

PATRONES ESTRUCTURALES:

- **Tendencia:** Vista como el componente de largo plazo que constituye la base del crecimiento o (declinación) de una serie de tiempo. Lo denotamos por T_t .

PATRONES ESTRUCTURALES:

- **Tendencia:** Vista como el componente de largo plazo que constituye la base del crecimiento o (declinación) de una serie de tiempo. Lo denotamos por T_t .
- **Variación cíclica:** Entendida como el conjunto de fluctuaciones en forma de onda o ciclos a medida que avanza la serie de tiempo. Lo denotamos por C_t .

PATRONES ESTRUCTURALES:

- **Tendencia:** Vista como el componente de largo plazo que constituye la base del crecimiento o (declinación) de una serie de tiempo. Lo denotamos por T_t .
- **Variación cíclica:** Entendida como el conjunto de fluctuaciones en forma de onda o ciclos a medida que avanza la serie de tiempo. Lo denotamos por C_t .
- **Variación estacional:** Se refiere a un patrón o cambio regularmente recurrente a través del tiempo. Lo denotamos por S_t .

PATRONES NO ESTRUCTURALES:

- **Fluctuación irregular:** Está compuesto por fluctuaciones causadas por sucesos impredecibles o no periódicos. Lo denotamos por I_t .

Si denotamos nuestra serie de tiempo en el instante t por X_t , podemos expresar nuestra serie de tiempo como un modelo:

- **Aditivo:**

$$X_t = T_t + S_t + C_t + I_t$$

- **Multiplicativo:**

$$X_t = T_t \times S_t \times C_t \times I_t$$

Para modelar la tendencia de una serie temporal es posible considerar:

- Modelos de **regresión lineal** considerando la serie de tiempo X_t como variable respuesta y el tiempo t como variable explicativa. En este caso se pueden utilizar modelos tales como: regresión lineal, polinomios, smoothing splines, splines cúbicos, etc.
- Métodos no paramétricos como por ejemplo, **medias móviles**.

Considere una serie de tiempo X_1, X_2, \dots, X_n . Las medias móviles pueden verse como promedios en “ventanas” de datos y las fórmulas son las siguientes:

- **Ventana unilateral:** Se define el tamaño de las ventanas, denotado por r , y se calcula como

$$\hat{X}_t = \frac{X_{t-r+1} + X_{t-r+2} + \dots + X_t}{r}$$

donde $t = r, r+1, \dots, n$ y $r \leq n$.

- **Ventana bilateral:** Se define el tamaño de las ventanas, denotado por r , y se calcula como

$$\hat{X}_t = \frac{X_{t-r_1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+r_2}}{r}$$

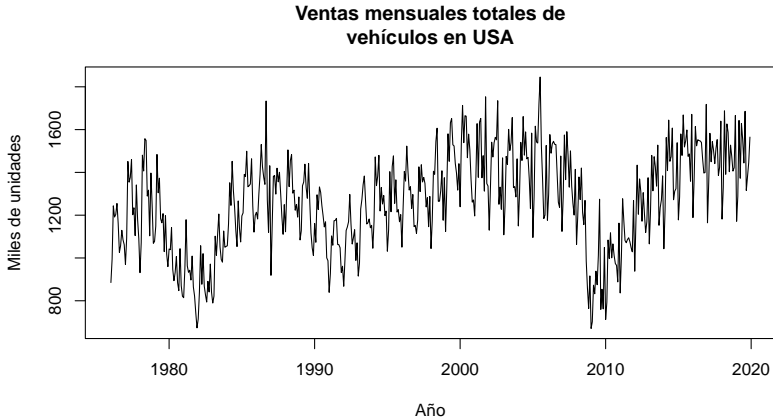
donde $r_1 + r_2 + 1 = r$, $t = r_1 + 1, \dots, n - r_2$ y $r \leq n$.

La programación de las fórmulas anteriores está dada en la siguiente función llamada **med_mov**:

```
med_mov<-function(x,unilateral=TRUE,r=2){  
  v1<-rep(NA,length(x))  
  if(unilateral==TRUE){  
    k<-1  
    for(i in r:length(x)){  
      v1[i]<-mean(x[k:i])  
      k<-k+1 } }  
  else{  
    r1<-ifelse(r%%2==0,r/2,(r-1)/2)  
    r2<-ifelse(r%%2==0,(r/2-1),(r-1)/2)  
    for(i in (r1+1):(length(x)-r2)){  
      v1[i]<-mean(x[(i-r1):(i+r2)]) } }  
  return(v1)}
```

Técnicas descriptivas: Tendencia - Medias Móviles

```
require(TSstudio)
plot(USVSales, main = "Ventas mensuales totales de
vehículos en USA",ylab= "Miles de unidades",xlab= "Año")
```



Apliquemos las funciones de medias móviles unilateral y bilateral con $r = 12$:

Apliquemos las funciones de medias móviles unilateral y bilateral con $r = 12$:

- **Unilateral:** Consideramos además darle el formato de series de tiempo al vector de medias móviles unilateral con la función **ts**

```
ma_uni1 <- med_mov(USVSAles,r=12, unilateral = TRUE)
ts_ma_uni1 <- ts(ma_uni1, start=1976, frequency=12)
```

Apliquemos las funciones de medias móviles unilateral y bilateral con $r = 12$:

- **Unilateral:** Consideramos además darle el formato de series de tiempo al vector de medias móviles unilateral con la función **ts**

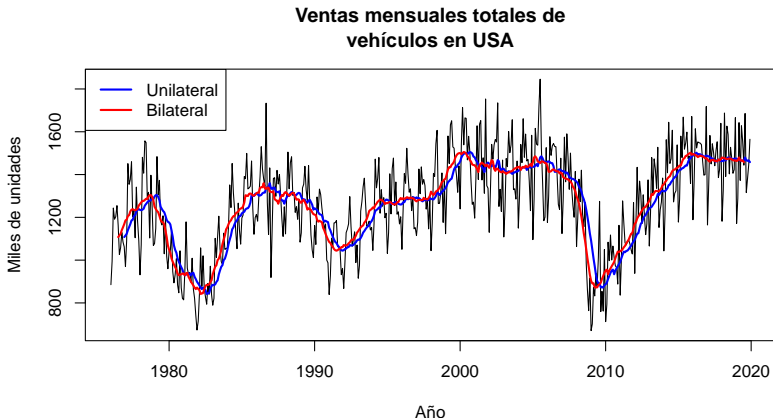
```
ma_uni1 <- med_mov(USVSAles,r=12, unilateral = TRUE)
ts_ma_uni1 <- ts(ma_uni1, start=1976, frequency=12)
```

- **Bilateral:** También le damos formato de series de tiempo al vector de medias móviles bilateral con la función **ts**

```
ma_bi1 <- med_mov(USVSAles,r=12, unilateral = FALSE)
ts_ma_bi1 <- ts(ma_bi1, start=1976, frequency=12)
```


Técnicas descriptivas: Tendencia - Medias Móviles

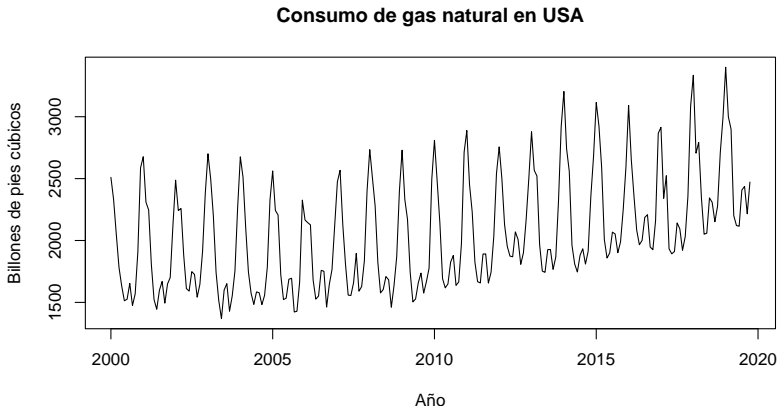
```
plot(USVSales, main = "Ventas mensuales totales de  
vehículos en USA", ylab= "Miles de unidades", xlab= "Año")  
lines(ts_ma_uni1 , col="blue", lwd=2)  
lines(ts_ma_bi1 , col="red", lwd=2)  
legend("topleft", c("Unilateral", "Bilateral"), col=c("blue", "red"),  
      lwd=2, lty=1)
```



- La tendencia busca describir cuándo una serie de tiempo crece o decrece y describir en qué periodos se presentaron dichas tendencias.
- Como vimos, existen muchos tipos de tendencias, desde lineal, polinómica, exponencial, etc., las cuales pueden ser modeladas con técnicas vistas en regresión lineal.
- Las medias móviles son muy utilizadas en series de tiempo para describir tendencias y se suele preferir las medias móviles bilaterales en lugar de las unilaterales, ya que estas están afectadas por una ventana que considera observaciones pasadas y futuras.

Técnicas descriptivas: Ciclos

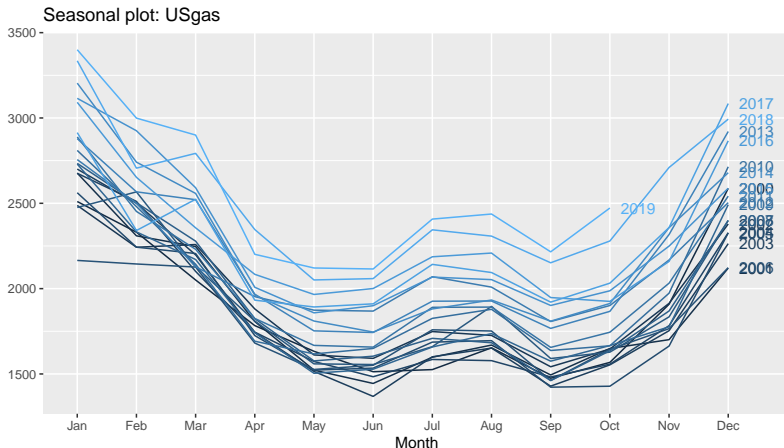
```
require(forecast)
plot(USgas, main = "Consumo de gas natural en USA",
     ylab= "Billones de pies cúbicos", xlab= "Año")
```



Técnicas descriptivas: Ciclos

Para una visualización más clara de los ciclos consideremos los datos de tres años seleccionados aleatoriamente:

```
ggseasonplot(USgas, year.labels=TRUE, continuous=TRUE)
```

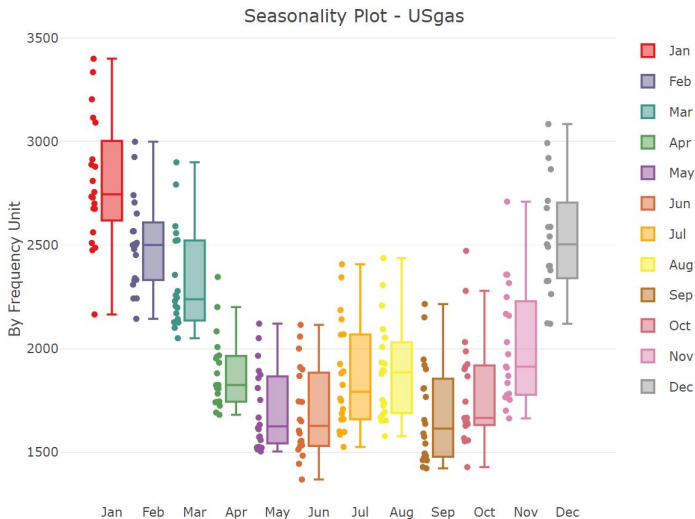


- Como vimos en el ejemplo, los ciclos pueden entenderse como eventos que se repiten a lo largo del tiempo y, por lo general, un ciclo comienza con un mínimo local de la serie y termina en el siguiente mínimo local, el cual a su vez es el punto de partida del siguiente ciclo.

- Como vimos en el ejemplo, los ciclos pueden entenderse como eventos que se repiten a lo largo del tiempo y, por lo general, un ciclo comienza con un mínimo local de la serie y termina en el siguiente mínimo local, el cual a su vez es el punto de partida del siguiente ciclo.
- Los ciclos no necesariamente ocurren en intervalos de la misma longitud y pueden variar de ciclo en ciclo. Es decir, podemos ver que en el ejemplo anterior del consumo de gas en USA se presentan ciclos anuales, sin embargo, en el ejemplo de las ventas de vehículos, vemos que un ciclo comienza aproximadamente en 1981 y termina aproximadamente en 1991 y otro que comienza aproximadamente en 1991, termina aproximadamente en 2009.

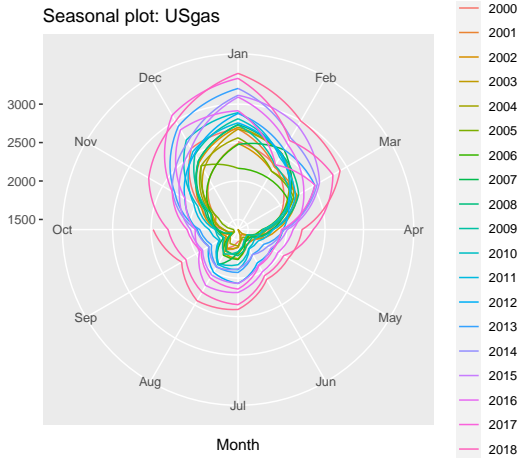
Técnicas descriptivas: Estacionalidad

```
ts_seasonal(USgas, type="box")
```



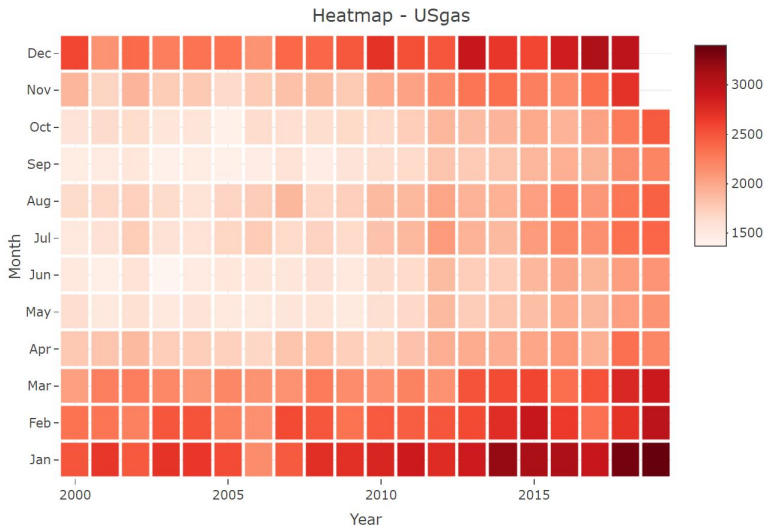
Técnicas descriptivas: Estacionalidad

```
require(forecast)
ggseasonplot(USgas, polar = TRUE)
```



Técnicas descriptivas: Estacionalidad

```
ts_heatmap(USgas, color = "Reds")
```



- La estacionalidad podría ser entendida como una variación repetida en la serie, que está relacionada con las unidades de frecuencia de la serie (por ejemplo, los meses del año para una serie mensual).

- La estacionalidad podría ser entendida como una variación repetida en la serie, que está relacionada con las unidades de frecuencia de la serie (por ejemplo, los meses del año para una serie mensual).
- Se asocia con eventos estacionales y, por lo general, describen comportamientos o situaciones que se repiten de manera reiterativa, por ejemplo, los hábitos de consumo son diferentes a mitad de año que en diciembre, o los fines de semana se observan comportamientos diferentes a los que se observan en semana, etc.

Trabajando con objetos **ts** y **data frame**:

En R, podemos crear objetos tipo series de tiempo o **ts** (por sus siglas en inglés), los cuales facilitan el trabajo a la hora de graficar o definir frecuencias.

Trabajando con objetos **ts** y **data frame**:

En R, podemos crear objetos tipo series de tiempo o **ts** (por sus siglas en inglés), los cuales facilitan el trabajo a la hora de graficar o definir frecuencias.

Como ejemplo consideremos la siguiente base de datos relacionada con tasas de desempleo en USA entre 1948 y 2019:

```
require(TSstudio) # Paquete con la base de datos  
class(USUnRate) # La base de datos es tipo ts
```

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

```
dim(USUnRate) # No tiene dimensiones porque NO es matriz
```

```
## NULL
```

```
length(USUnRate) # Es un vector y por eso tiene longitud
```

```
## [1] 864
```

Trabajando con objetos ts y data frame:

USUnRate

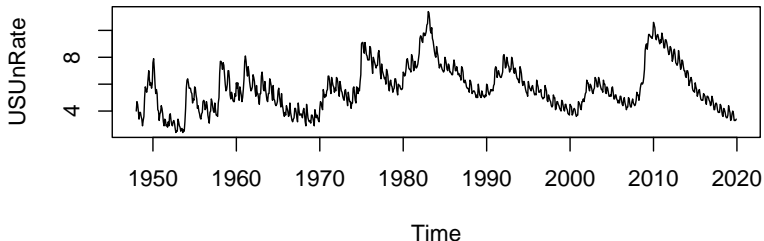
| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1948 | 4.0 | 4.7 | 4.5 | 4.0 | 3.4 | 3.9 | 3.9 | 3.6 | 3.4 | 2.9 | 3.3 | 3.6 |
| 1949 | 5.0 | 5.8 | 5.6 | 5.4 | 5.7 | 6.4 | 7.0 | 6.3 | 5.9 | 6.1 | 5.7 | 6.0 |
| 1950 | 7.6 | 7.9 | 7.1 | 6.0 | 5.3 | 5.6 | 5.3 | 4.1 | 4.0 | 3.3 | 3.8 | 3.9 |
| 1951 | 4.4 | 4.2 | 3.8 | 3.2 | 2.9 | 3.4 | 3.3 | 2.9 | 3.0 | 2.8 | 3.2 | 2.9 |
| 1952 | 3.7 | 3.8 | 3.3 | 3.0 | 2.9 | 3.2 | 3.3 | 3.1 | 2.7 | 2.4 | 2.5 | 2.5 |
| 1953 | 3.4 | 3.2 | 2.9 | 2.8 | 2.5 | 2.7 | 2.7 | 2.4 | 2.6 | 2.5 | 3.2 | 4.2 |
| 1954 | 5.7 | 6.3 | 6.4 | 6.1 | 5.7 | 5.7 | 5.7 | 5.4 | 5.3 | 4.6 | 4.9 | 4.8 |
| | : | | | : | | : | | : | | : | | |
| 2013 | 8.5 | 8.1 | 7.6 | 7.1 | 7.3 | 7.8 | 7.7 | 7.3 | 7.0 | 7.0 | 6.6 | 6.5 |
| 2014 | 7.0 | 7.0 | 6.8 | 5.9 | 6.1 | 6.3 | 6.5 | 6.3 | 5.7 | 5.5 | 5.5 | 5.4 |
| 2015 | 6.1 | 5.8 | 5.6 | 5.1 | 5.3 | 5.5 | 5.6 | 5.2 | 4.9 | 4.8 | 4.8 | 4.8 |
| 2016 | 5.3 | 5.2 | 5.1 | 4.7 | 4.5 | 5.1 | 5.1 | 5.0 | 4.8 | 4.7 | 4.4 | 4.5 |
| 2017 | 5.1 | 4.9 | 4.6 | 4.1 | 4.1 | 4.5 | 4.6 | 4.5 | 4.1 | 3.9 | 3.9 | 3.9 |
| 2018 | 4.5 | 4.4 | 4.1 | 3.7 | 3.6 | 4.2 | 4.1 | 3.9 | 3.6 | 3.5 | 3.5 | 3.7 |
| 2019 | 4.4 | 4.1 | 3.9 | 3.3 | 3.4 | 3.8 | 4.0 | 3.8 | 3.3 | 3.3 | 3.3 | 3.4 |

Trabajando con objetos ts y data frame:

```
ts_info(USUnRate)
```

```
## The USUnRate series is a ts object with 1 variable and 864 observations  
## Frequency: 12  
## Start time: 1948 1  
## End time: 2019 12
```

```
plot(USUnRate)
```



Trabajando con objetos **ts** y **data frame**:

Para crear un objeto **ts** desde cero, podemos considerar lo siguiente:

```
datos_1<-c(1.65,2.74,1.25,1.52,1.99,1.13,1.22,1.10,2.49,  
1.48,1.07,1.34,1.13,1.81,1.73,2.97,1.81,1.52,1.70,2.22,  
2.56,2.98,1.95,1.35)
```

Para una serie de datos mensuales:

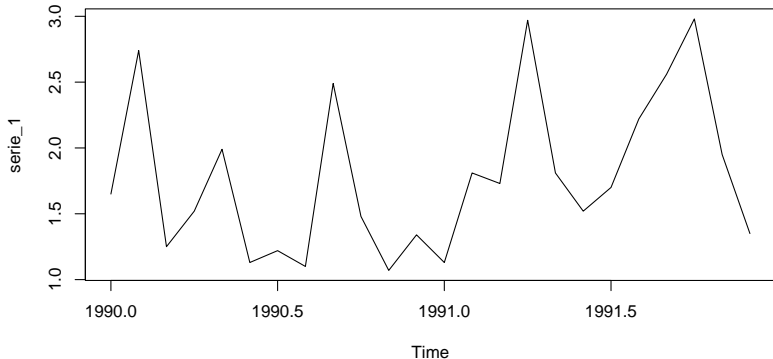
```
serie_1 <- ts(datos_1, start=1990, frequency = 12)
```

Para una serie de datos trimestrales:

```
serie_2 <- ts(datos_1, start=1990, frequency = 4)
```

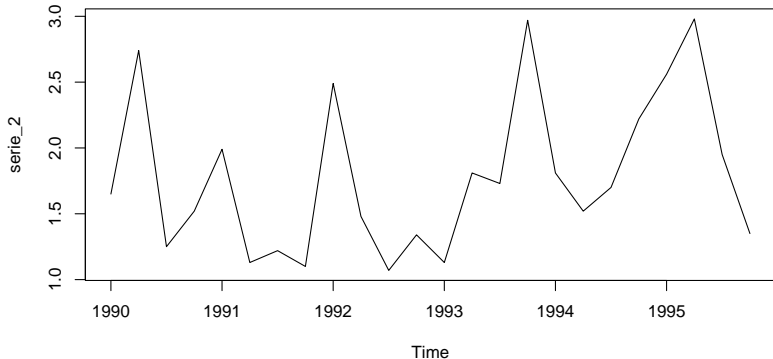

Trabajando con objetos ts y data frame:

```
plot(serie_1)
```



Trabajando con objetos ts y data frame:

```
plot(serie_2)
```



Trabajando con objetos **ts** y **data frame**:

Para obtener otros estadísticos descriptivos como por ejemplo medias mensuales o trimestrales, entre otros, debemos convertir los objetos **ts** a **data frame**.

Trabajando con objetos **ts** y **data frame**:

Para obtener otros estadísticos descriptivos como por ejemplo medias mensuales o trimestrales, entre otros, debemos convertir los objetos **ts** a **data frame**. Consideremos el siguiente ejemplo:

```
# Convertimos USgas de objeto ts a data.frame:
USgas_df <- data.frame(year = floor(time(USgas)),
                       month = cycle(USgas),
                       USgas = as.numeric(USgas))

# Transformamos los meses en factor:
USgas_df$month <- factor(month.abb[USgas_df$month],
                         levels = month.abb)
```

Trabajando con objetos ts y data frame:

```
head(USgas_df)
```

```
##   year month  USgas
## 1 2000   Jan 2510.5
## 2 2000   Feb 2330.7
## 3 2000   Mar 2050.6
## 4 2000   Apr 1783.3
## 5 2000   May 1632.9
## 6 2000   Jun 1513.1
```

```
tail(USgas_df)
```

```
##   year month  USgas
## 233 2019   May 2121.0
## 234 2019   Jun 2115.2
## 235 2019   Jul 2407.5
## 236 2019   Aug 2437.2
## 237 2019   Sep 2215.6
## 238 2019   Oct 2472.3
```

Trabajando con objetos **ts** y **data frame**:

Para obtener resúmenes, como por ejemplo medias, por mes consideramos lo siguiente:

```
require(magrittr)

require(tidyverse)

resum_1 <- USgas_df %>% group_by(month) %>%
  summarise(med_month = mean(USgas),
            sd_month = sd(USgas))
```

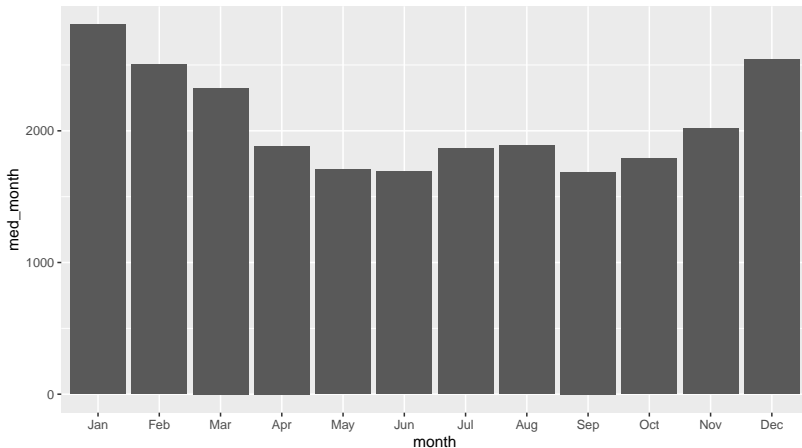
Trabajando con objetos ts y data frame:

```
resum_1
```

```
## # A tibble: 12 x 3
##   month med_month sd_month
##   <fct>      <dbl>    <dbl>
## 1 Jan      2806.      309.
## 2 Feb      2502.      223.
## 3 Mar      2325.      241.
## 4 Apr      1886.      175.
## 5 May      1708.      194.
## 6 Jun      1691.      216.
## 7 Jul      1864.      261.
## 8 Aug      1889.      237.
## 9 Sep      1687.      242.
## 10 Oct      1788.      260.
## 11 Nov      2015.      284.
## 12 Dec      2543.      278.
```

Trabajando con objetos ts y data frame:

```
ggplot(resum_1, aes(x=month, y=med_month))+  
  geom_bar(stat="identity")
```



- En el ejemplo anterior podemos observar que los meses con mayor consumo de gas promedio son enero y diciembre y los de menor consumo promedio corresponden a los meses de mayo, junio y septiembre.
- En el ejemplo anterior calculamos los promedios(`mean`) y las desviaciones estándar (`sd`), pero también es posible calcular todo tipo de estadísticos muestrales como medianas (`median`), mínimos (`min`), máximos (`max`), etc.

Actividad Presentación 2:

- 1 Repetir los códigos de esta presentación en sus computadores mediante el uso del RStudio.
- 2 Vaya al sitio web del DANE haciendo click [AQUÍ](#), luego hacer click en **Estadísticas por tema**, busque en la columna de **Economía** el item **Construcción** y hacer click en **Estadísticas de cemento gris (ECG)**. Vaya a **Información adicional** y hacer click en **Información Histórica**, luego hacer click en **Anexos** de Mayo en la tabla del Año 2022. En este punto se descarga automáticamente una base de datos en Excel.
 - a. Usando el RStudio, explore las hojas de la base de datos y describa de qué se trata cada una.
 - b. Realice un análisis descriptivo de al menos 4 series de tiempo que encuentre en esta base de datos.