

# R Markdown, investigación reproducible y aplicaciones educativas

Lino AA Notarantonio (lino@tec.mx)

12/17/2019

# Presentación

R Markdown,  
investigación  
reproducibile y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

La pueden bajar en el repositorio de Github. Sigán [Jump to file](#) en la página del repositorio.



# R Markdown

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

**R Markdown** proporciona un marco para académicos que trabajan en ciencias de datos.

R Markdown permite

- guardar y ejecutar código;
- generar reportes de alta calidad que se puedan compartir.

Más detalles dando click en la siguiente [liga](#).

# Investigación reproducible: Antecedentes

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Los métodos estadísticos juegan un papel fundamental en las ciencias empíricas con baja relación señal-ruido.

En los últimos años se han presentados problemas de reproducibilidad de muchos resultados en ciencias sociales y ciencias de la salud. Entre otros, mencionamos

- [Reproducibility Project in Psychology](#) (Science, 2015). Una de las conclusiones es una evidencia débil para los resultados en los artículos originales, aún con datos proporcionados por los mismos autores.

# Investigación reproducible: Antecedentes

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

- [Social Sciences Replication Project](#) (2018) De los 21 experimentos considerados (todos publicados en *Nature* o *Science*, en el periodo 2010-2015), sólo 13 pudieron ser replicados. Además, el tamaño del efecto en el estudio es aproximadamente la mitad de los que se publicó.

# Investigación reproducible: Ejemplo ilustrativo

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Datos perdidos ([L. Weitzman](#), 1996)

- Los datos modificados que sirvieron para el análisis de un artículo anterior (Weitzman, 1985) se perdieron. sólo quedaban los datos originales (*raw data*).
- Un investigador (Peterson, 1996) pudo reproducir parcialmente los hallazgos del 1985.

# Investigación reproducible

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Investigación que permite la presentación de los resultados con la manipulación **transparente** y **reproducible** de los datos correspondientes.

## Beneficios

- Investigación: facilidad de réplica de resultados empíricos.
- Aplicaciones educativas (análisis de datos): manipulación de los datos asociados visualizando el código y los resultados.

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

- Los datos a continuación se bajaron del sitio NOAA (National Oceanic and Atmospheric Administration) de EEUU.
- Los datos abarcan observaciones de CO2 en el periodo marzo 1958 a marzo 2017.
- La columna (variable) *average* contiene las medias mensuales de CO2 (ppm, partes per millones).
- Datos ausentes para la variable *average* se denotan con  $-99.99$ ; para la variable *no. days* con  $-1$ .



# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Visualización de las primeras cuatro observaciones del dataset:

```
co2_mm_mlo <-  
  read.table("co2_mm_mlo.txt", quote="\")  
head(co2_mm_mlo, 4)
```

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

##		V1	V2		V3	V4	V5	V6	V7
##	1	1958	3	1958.208	315.71	315.71	314.62	-1	
##	2	1958	4	1958.292	317.45	317.45	315.29	-1	
##	3	1958	5	1958.375	317.50	317.50	314.71	-1	
##	4	1958	6	1958.458	-99.99	317.10	314.85	-1	

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Se usará la variable *average* (V4).

Cambiamos el valor numérico  $-99.99$  (que indica observaciones ausentes) a *NA*:

```
co2_mm_mlo$V4[co2_mm_mlo$V4 == -99.99] <- NA  
head(co2_mm_mlo, 4)
```

##		V1	V2		V3	V4	V5	V6	V7
##	1	1958	3	1958.208	315.71	315.71	314.62	-1	
##	2	1958	4	1958.292	317.45	317.45	315.29	-1	
##	3	1958	5	1958.375	317.50	317.50	314.71	-1	
##	4	1958	6	1958.458		NA	317.10	314.85	-1

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Cálculo de estadísticos descriptivos:

```
x <- co2_mm_mlo$V4  
summary(x, na.rm = TRUE)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	313.2	328.5	350.2	352.5	373.7	407.7

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

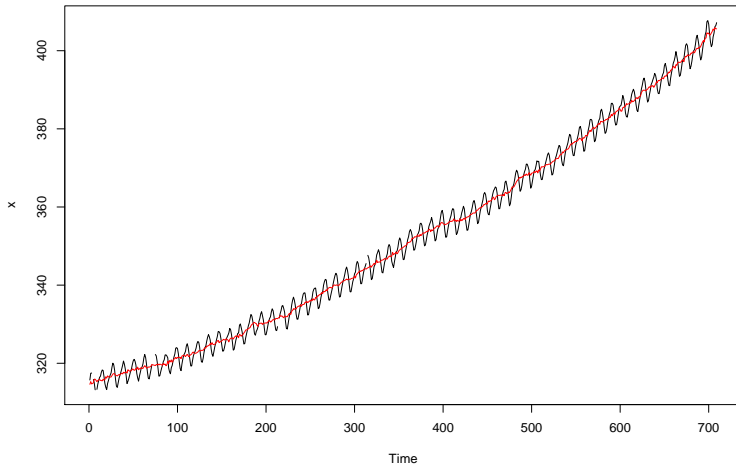
Ploteo de la variable *average* como series de tiempo, incluyendo la tendencia:

```
plot.ts(x)
y <- co2_mm_mlo$V6
lines(y, col = "red")
```

# Ejemplo 1

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)



## Ejemplo 2

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

Simulación de 100 muestras tomadas de una población normal estándar con histograma y las densidades de probabilidad teórica y empírica correspondiente.

```
set.seed(123)
y <- rnorm(100)
hist(y, probability = TRUE)
curve(dnorm(x), col = "red", add = TRUE)
d <- density(y)
lines(d, col = "blue", add = TRUE)
```

## Ejemplo 2

R Markdown,  
investigación  
reproducible y  
aplicaciones  
educativas

Lino AA  
Notarantonio  
(lino@tec.mx)

```
## Warning in plot.xy(xy.coords(x, y), type = type, .  
## graphical parameter
```

