

Unidad 1 / Escenario 1

Lectura Fundamental

Introducción al *software* R

Contenido

- 1 Introducción
- 2 ¿Por qué utilizar R?
- 3 Ventajas de R
- 4 ¿Cómo instalar R?
- 5 ¿Cómo instalar Rstudio?
- 6 R es un proyecto colaborativo

7 R como calculadora

8 Funciones básicas de Estadística con R

9 Gráficos con R

10 Documentación disponible en Internet

Referencias

Palabras Claves: *software* estadístico R, Rstudio, computación estadística, tratamiento de datos, proyecto colaborativo.

1. Introducción

R es un lenguaje computacional para el análisis y modelamiento de datos (R *Core Team*, 2018). Utilizado como lenguaje de programación orientado a objetos, o como un entorno estadístico en el que el usuario pueden utilizar, crear o modificar librerías para diferentes metodologías. (Ihaka y Gentleman, s.f.).

R es un software libre con licencia GNU (*General Public Licence*). Esto a hecho que R tenga una capacidad y versatilidad increíble, ya que con esté *software* se pueden realizar desde análisis sencillos como también análisis complejos y muy utilitarios en diferentes ramas de las ciencias.

2. ¿Por qué utilizar R?

Según Jones (s.f.), algunas de las principales razones para utilizar R, son las siguientes:

R es usado por la mayoría de los estadísticos académicos. Actualmente en el software R se implementarán nuevos desarrollos, lo que lo hace muy utilizado por los estadísticos.

R es independiente de la plataforma (sea Linux, Windows, MacOS). Se puede instalar para cualquiera de las plataforma más utilizadas.

R dispone de muchas recursos de ayudas. En la Internet hay muchas ayudas disponibles de R. Incluso hay bibliografía especializada en R, como también revistas con publicaciones de código y programas en R.

Permite realizar gráficos de gran calidad. Existen muchos paquetes que para hacer gráficos de calidad lo que lo hace muy potente en esté entorno.

La interfaz de líneas de comandos es mejor para la enseñanza. El lenguaje de programación no es complicado, ya que es un lenguaje muy sencillo de entender.

R es un lenguaje de programación completo, no solo una aplicación en estadística. Con el auge de la cantidad de paquetes que tiene, se puede aplicar en muchas áreas lo que lo hace muy útil.

3. Ventajas de R

Entre las ventajas que más se destacan de R tenemos las siguientes:

- R es un software libre.

- R utiliza muchas librerías para diferentes metodológicas estadísticas.
- R contiene rutinas estadísticas avanzadas creadas por los usuarios que no están disponibles en otros paquetes.
- R es un software referente para la comunidad académica y científica.
- R es un programa que se puede manejar bajo comandos (aunque existen interfaz gráficas bastante útiles).

4. ¿Cómo instalar R?

Para instalar R, ingrese a la siguiente página <http://www.r-project.org/>, y haga clic en la opción *download R* (Figura 1).

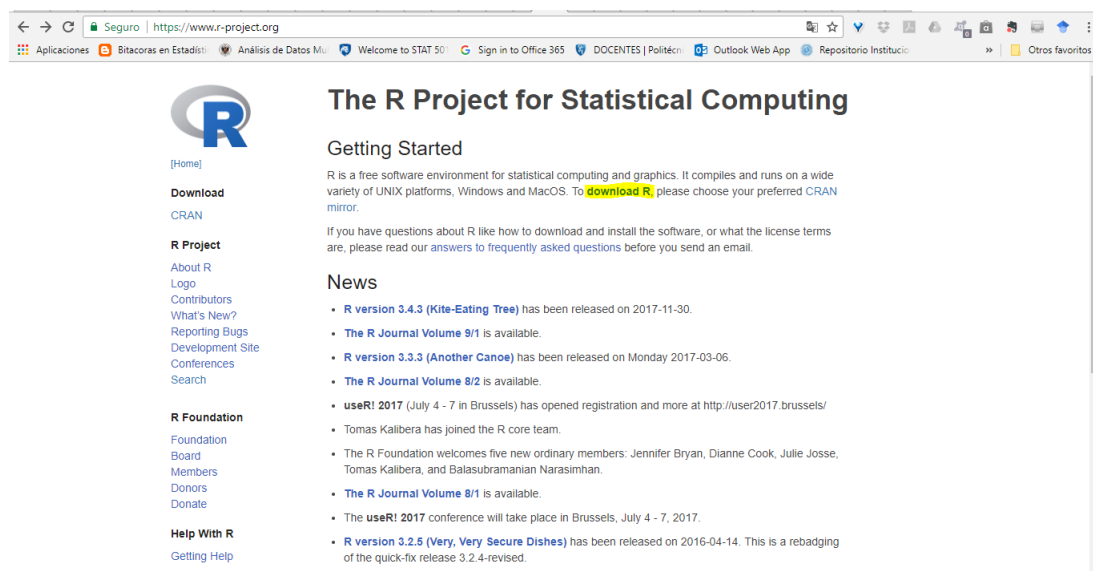


Figura 1: Imagen pagina web www.r-project.org para descargar R.

Fuente: elaboración propia.

Aparece un listado de servidores (CRAN *mirror*) ubicados en varios países, se escoge uno de su preferencia (Figura 2).

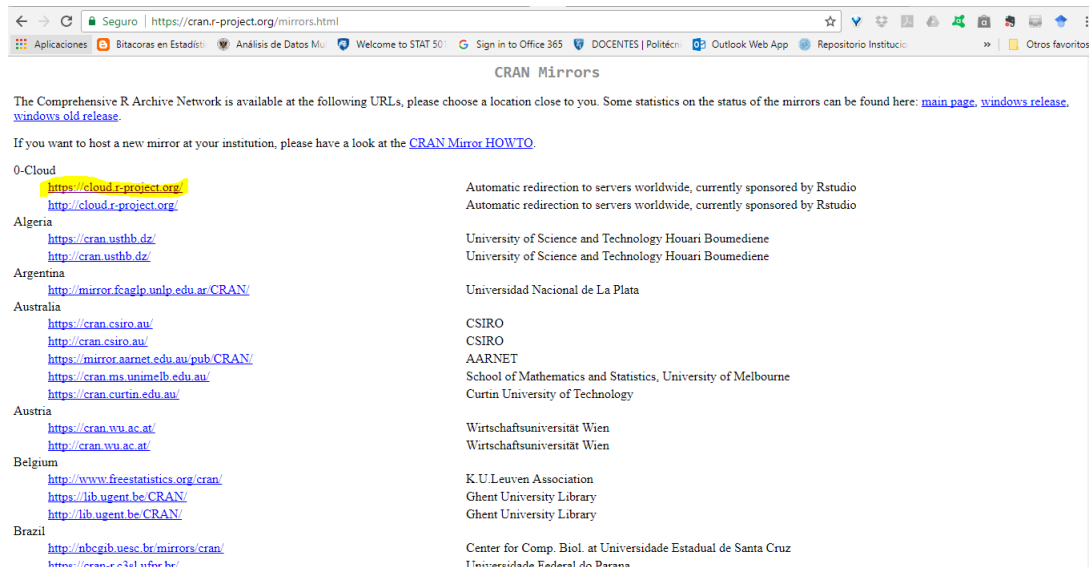


Figura 2: Imagen página web de los servidores para descargar R. `cran.r-project.org/mirrors.html`.
Fuente: elaboración propia.

Seleccione la opción R según su sistema operativo (Figura 3).

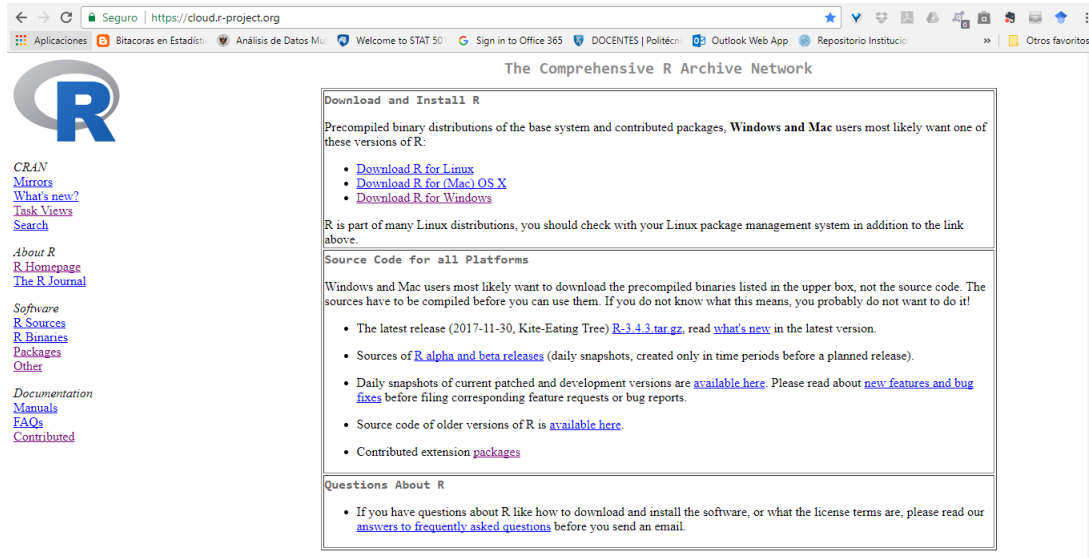


Figura 3: Imagen de la selección de R para descargar según su sistema operativo. `cloud.r-project.org`.
Fuente: elaboración propia.

Se descargará la aplicación para el sistema operativo que el usuario tenga en su computadora. Si es windows debe descargar la versión base (Figura 4).

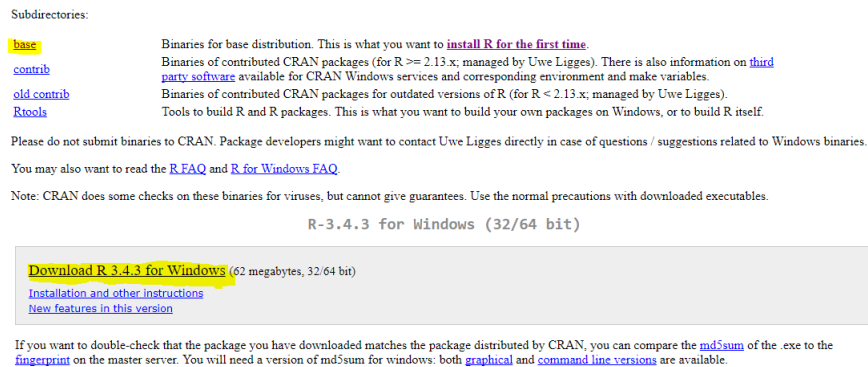


Figura 4: Imagen para descargar la versión base de R para windows.
Fuente: elaboración propia.

La instalación se hace mediante un ejecutable el cual se instala de forma usual siguiendo los pasos y listo (Figura 5).

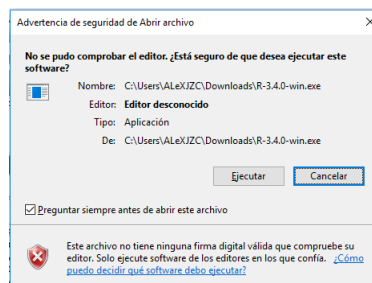



Figura 5: Imagen de la instalación del ejecutable de R para windows.
Fuente: elaboración propia.

En el escritorio deben tener el icono R . La apariencia de R al abrirlo por primera vez es muy simple (Figura 6).

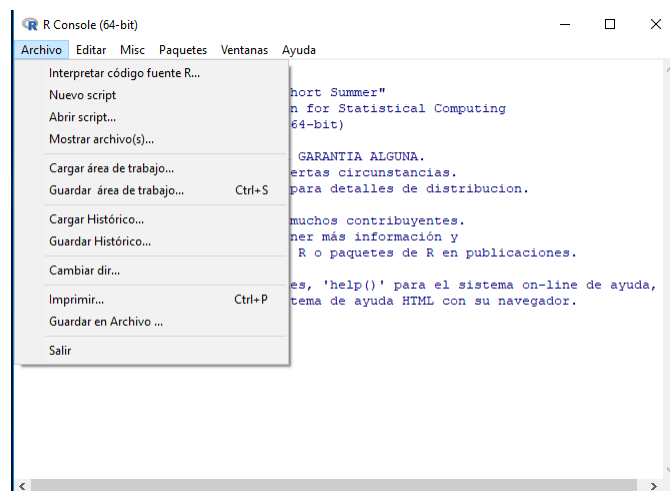


Figura 6: Imagen de la consola de R.
Fuente: elaboración propia.

5. ¿Cómo instalar Rstudio?

Algunos desarrolladores de software, han creado interfaces que permitan mejorar la apariencia de R. Uno de estos y en la actualidad el más usado es el programa Rstudio (“RStudio Team.(2016). Rstudio: Integrated Development environment for R [Manual de software informático].”, s.f.). Para instalarlo debe ingresar a la siguiente página web www.rstudio.com/ (Figura 7). Ingrese a la página y de clic en *download* para descargar la opción para escritorio. (Figura 8).

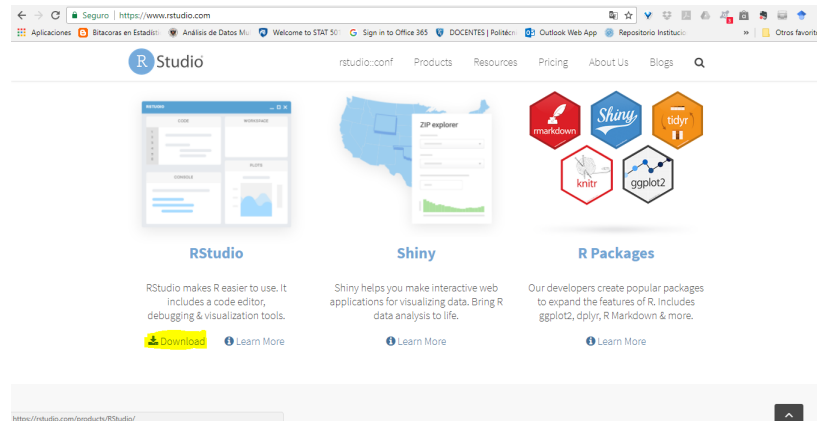


Figura 7: Imagen pagina web www.rstudio.com para descargar *Rstudio*.
Fuente: elaboración propia.

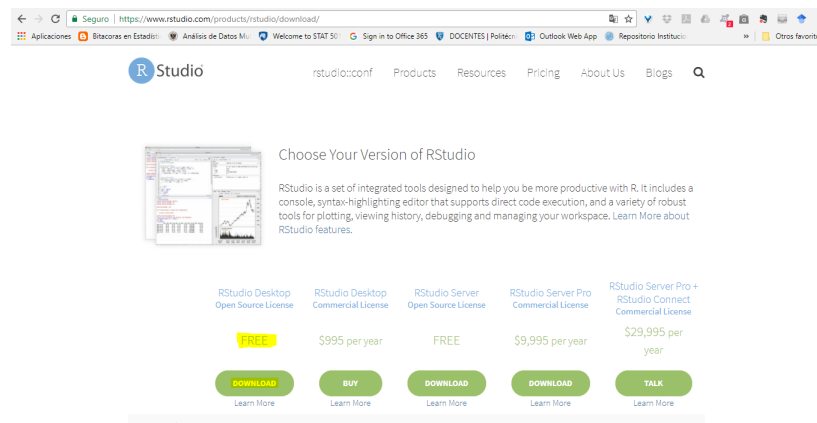


Figura 8: Imagen para descargar la versión libre de Rstudio para desktop. www.rstudio.com/products/rstudio/download/.
Fuente: elaboración propia.

Seleccione la opción Rstudio según su sistema operativo e instale de forma usual (Figura 9).

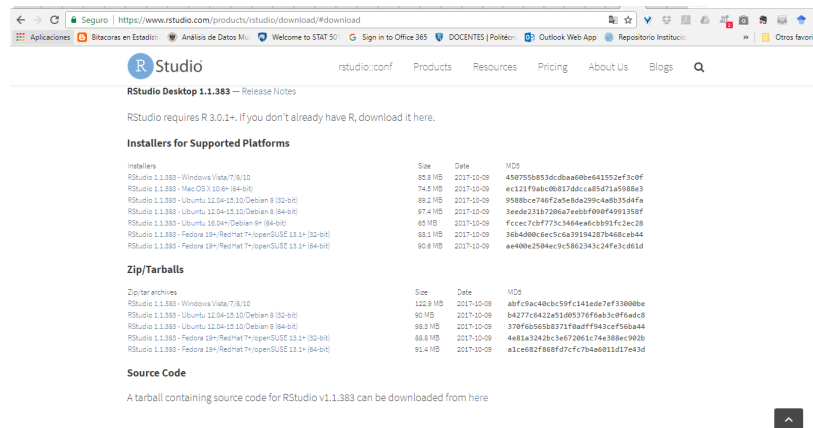


Figura 9: Imagen de la seleccionando Rstudio para descargar según su sistema operativo.
Fuente: elaboración propia.

Rstudio ofrece una interfaz que permite interactuar con R de manera muy simple. (Figura 10).

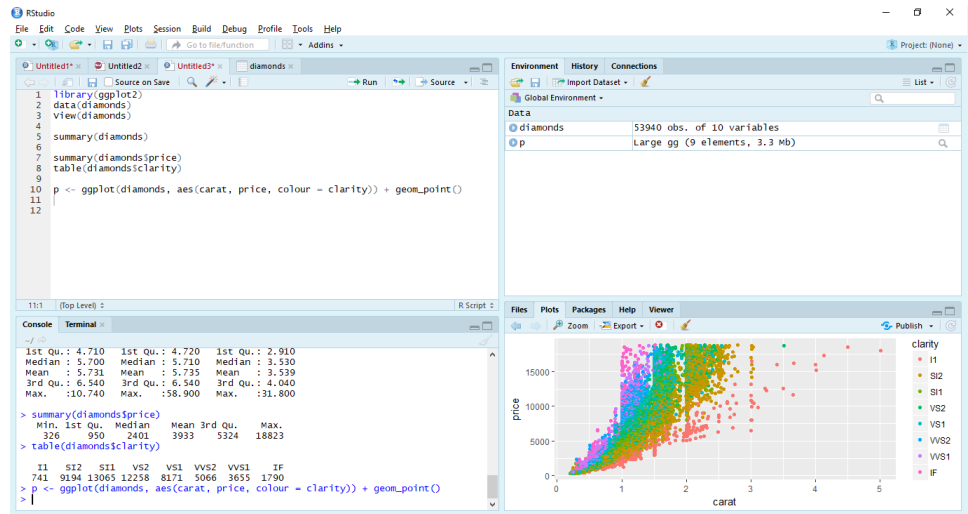


Figura 10: Imagen de Rstudio para *desktop*.
Fuente: elaboración propia.

6. R es un proyecto colaborativo

Una de las grandes riquezas de R es la cantidad de paquetes (más de 12000 actualmente) que amplían sus funcionalidades. La lista completa está en <http://cran.es.r-project.org/web/packages/>. Si desea verlas agrupadas por temáticas puede verlas en el CRAN *Task Views* en el siguiente link <http://cran.r-project.org/web/views/>

Algunos vienen instalados y se cargan al empezar el software.


```

sessionInfo()

## R version 3.4.4 (2018-03-15)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Spanish_Colombia.1252 LC_CTYPE=Spanish_Colombia.1252
## [3] LC_MONETARY=Spanish_Colombia.1252 LC_NUMERIC=C
## [5] LC_TIME=Spanish_Colombia.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  base
##
## other attached packages:
## [1] knitr_1.20
##
## loaded via a namespace (and not attached):
## [1] compiler_3.4.4 magrittr_1.5   tools_3.4.4   stringi_1.1.7
## [5] methods_3.4.4  stringr_1.3.0 evaluate_0.10.1

```

Otros vienen instalados pero hay que cargarlos:

```

library(lattice)
packageVersion('lattice')

## [1] '0.20.35'

packageDescription('lattice')

## Package: lattice
## Version: 0.20-35
## Date: 2017-03-23
## Priority: recommended
## Title: Trellis Graphics for R
## Author: Deepayan Sarkar <deepayan.sarkar@r-project.org>
## Maintainer: Deepayan Sarkar <deepayan.sarkar@r-project.org>
## Description: A powerful and elegant high-level data visualization
##              system inspired by Trellis graphics, with an emphasis on
##              multivariate data. Lattice is sufficient for typical
##              graphics needs, and is also flexible enough to handle most
##              nonstandard requirements. See ?Lattice for an
##              introduction.

```

```
## Depends: R (>= 3.0.0)
## Suggests: KernSmooth, MASS, latticeExtra
## Imports: grid, grDevices, graphics, stats, utils
## Enhances: chron
## LazyLoad: yes
## LazyData: yes
## License: GPL (>= 2)
## URL: http://lattice.r-forge.r-project.org/
## BugReports: https://github.com/deepayan/lattice/issues
## NeedsCompilation: yes
## Packaged: 2017-03-25 11:25:40 UTC; deepayan
## Repository: CRAN
## Date/Publication: 2017-03-25 18:05:47 UTC
## Built: R 3.4.0; x86_64-w64-mingw32; 2017-05-01 11:38:19 UTC;
##      windows
##
## -- File: C:/Users/AleXJZC/Documents/R/win-library/3.4/lattice/Meta/package.rds
```

Otros hay que instalarlos y después cargarlos:

```
install.packages('zoo')
library('zoo')
packageDescription('zoo')
```

Una forma bastante útil de instalar un paquete dando clic a la pestaña *Packages* y luego *Install* en la parte central a mano izquierda de los menús de Rstudio (Figura 11).

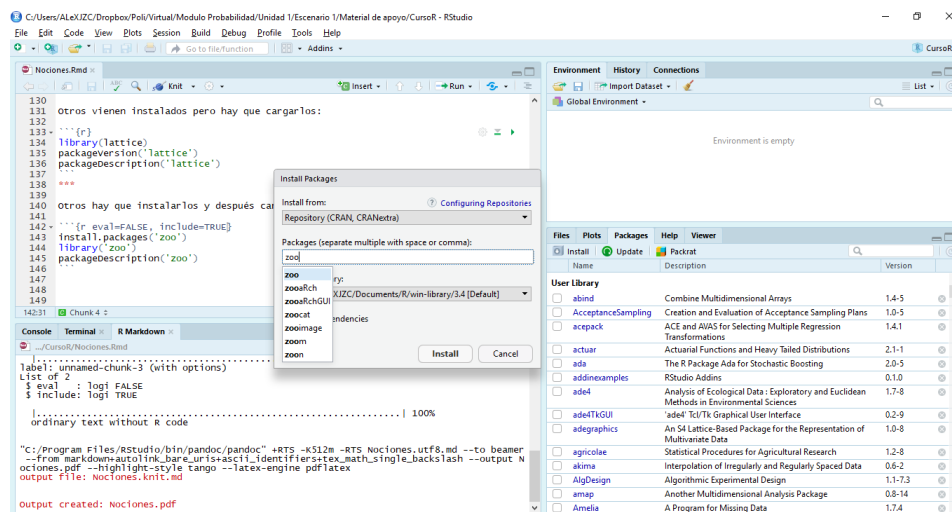


Figura 11: Instalando paquetes en Rstudio

Fuente: elaboración propia.

7. R como calculadora

En el software R se pueden realizar algunos cálculos muy sencillos utilizando los símbolos de $+$, $-$, $*$, $/$, $^$ para realizar operaciones básicas. Si el lector desea conocer algunos ejemplos sencillos y sus funciones se recomienda la lectura de Verzani (2004). A continuación se describen algunos ejemplos.

```
5+8
```

```
## [1] 13
```

```
2*4
```

```
## [1] 8
```

```
75-12
```

```
## [1] 63
```

```
(3-5)*4
```

```
## [1] -8
```

```
sqrt(16) # raíz cuadrada de 16
```

```
## [1] 4
```

```
exp(1) # número euler
```

```
## [1] 2.718282
```

```
log(3) # logaritmo natural de 3
```

```
## [1] 1.098612
```

```
log(1000,10) # logaritmo de 1000 en base 10
```

```
## [1] 3

pi # número pi

## [1] 3.141593

sin(pi/2)

## [1] 1

cos(pi)

## [1] -1

tan(pi)

## [1] -1.224647e-16

floor(4.8) # redondeo por debajo

## [1] 4

ceiling(4.3) # redondeo por arriba

## [1] 5

round(pi,digits=2) # redondear por 2 decimales

## [1] 3.14

trunc(12.48) # redondeo eliminando decimales

## [1] 12
```

7.1. Actividad

Realice los siguientes cálculos aritméticos y redondee los resultados a 2 cifras decimales, utilizando R

- $\frac{24}{\frac{-8+7}{9-6(2)}}$
- $2^3(-7) + 4 - \left(\frac{1}{3} + \frac{1}{2}\right)$
- $\frac{\sqrt{16}+5(4)-3^{-2}}{3(4-8)+1}$
- $\frac{1}{2} \left[\left(\frac{5}{4} - 2^{-1} \right) \frac{13}{4} - \sqrt{7} + 8 \div \left(2 - \frac{1}{5} \right) - 3^2 \right]$
- $2 \sin\left(\frac{\pi}{3}\right) + 5 \cos\left(\frac{\pi}{4}\right) - \frac{\tan\left(\frac{\pi}{6}\right)}{4}$
- $4 \log_2(5) - 3 \log_3(7) + \frac{1}{3} \ln^3(8)$

8. Funciones básicas de Estadística con R

Dado que el *software* R es un programa estadístico se puede realizar cálculos de todas las estadísticas descriptivas que se requieran. Para ello miremos el siguiente ejemplo.

```
notas=c(3.7,2.8,3.6,2,4.1,3.8,3.6,4.4,4.1,2.5,3.3,3.1) # se crea el vector de notas
sort(notas) # ordena los valores de menor a mayor
```

```
## [1] 2.0 2.5 2.8 3.1 3.3 3.6 3.6 3.7 3.8 4.1 4.1 4.4
```

```
sum(notas) # suma los valores de las notas
```

```
## [1] 41
```

```
length(notas) # longitud del vector de notas
```

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

```
min(notas) # valor mínimo de las notas
```

```
## [1] 2
```

```
max(notas) # valor máximo de las notas

## [1] 4.4

mean(notas) # valor promedio de las notas

## [1] 3.416667

median(notas) # valor de la mediana de las notas

## [1] 3.6

quantile(notas) # valor de los cuantiles de las notas

##      0%    25%    50%    75%   100%
## 2.000 3.025 3.600 3.875 4.400

sd(notas) # desviación estándar de las notas

## [1] 0.7094599

var(notas) # varianza muestras de las notas

## [1] 0.5033333
```

Cómo mejorar

Si el lector desea conocer más sugiere la lectura de Dalgaard (2008).

9. Gráficos con R

A continuación se describen algunos gráficos que se pueden realizar con R. Para ello se tendrá en cuenta el vector de notas de la sección anterior.

```
hist(notas)
```

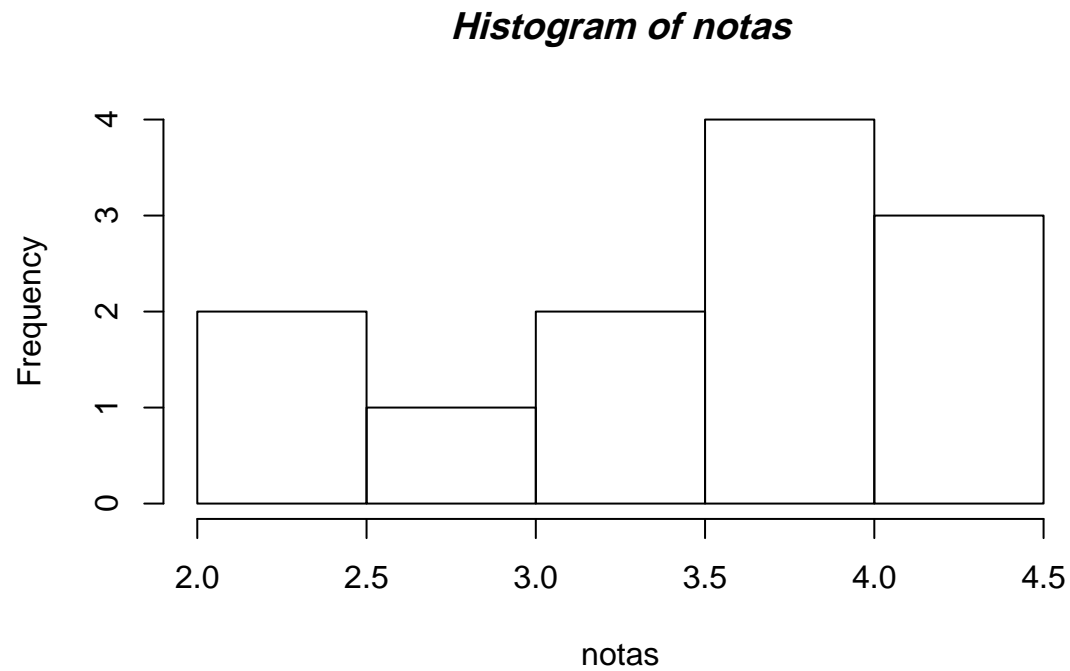


Figura 12: Histograma de las notas.
Fuente: elaboración propia.

```
boxplot(notas, horizontal = TRUE)
```

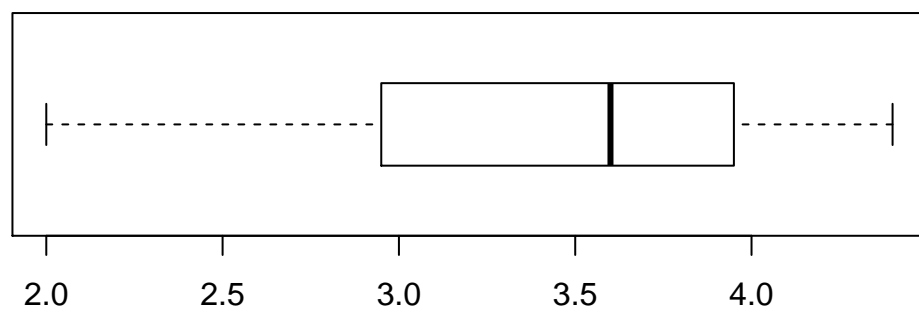


Figura 13: Boxplot de las notas.
Fuente: elaboración propia.

```
d <- density(notas)
plot(d)
```

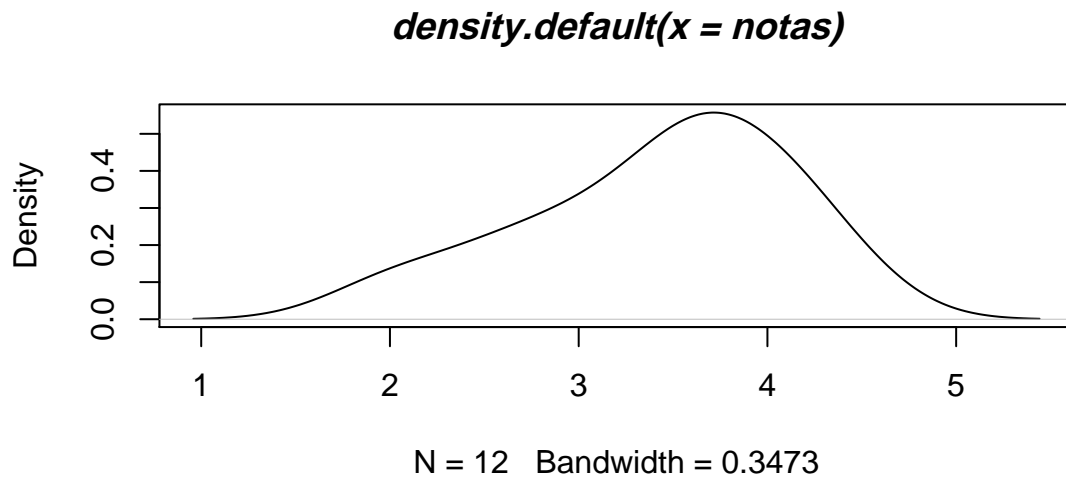


Figura 14: Diagrama de densidad de las notas.
Fuente: Elaboración propia.

```
dotchart(notas)
```

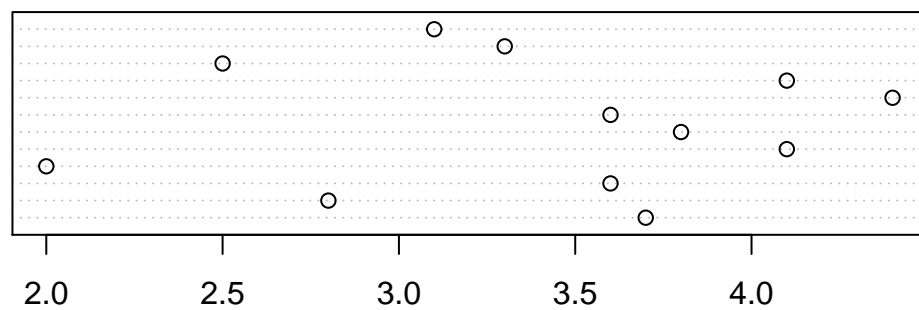


Figura 15: Diagrama de puntos de las notas.
Fuente: elaboración propia.

10. Documentación disponible en Internet

A continuación se describe algunas páginas en donde se podrá encontrar alguna documentación de interés.

- Manual introductorio: <http://cran.r-project.org/doc/manuals/R-intro.html>
- Manuales oficiales: <http://cran.r-project.org/manuals.html>
- Manuales externos: <http://cran.r-project.org/other-docs.html>
- R-help-es: <https://stat.ethz.ch/pipermail/r-help-es/>
- R-bloggers: <http://www.r-bloggers.com>
- Quick-R: <http://www.statmethods.net/>
- Introducción a R: <http://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- Libros: <http://www.r-project.org/doc/bib/R-books.html>

Cómo mejorar.

Si el lector desea en el Material del apoyo del presente modulo, podrá encontrar un documento con más ejemplos y actividades para realizar.

Referencias

- Dalgaard, P. (2008). *Introductory statistics with R*. New York: Springer.
- Ihaka, R., y Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Jones, J. H. (2009). *Why use r?* Descargado de <http://monkeysuncle.stanford.edu/?p=367>
- R Core Team. (2018). R: A language and environment for statistical computing [manual de software informático] [Manual de software informático]. Vienna, Austria. Descargado de <https://www.R-project.org/>
- Rstudio team.(2016).rstudio: Integrated development environment for r [manual de software informático]. [Manual de software informático]. (s.f.).
- Verzani, J. (2004). *Using r for introductory statistics*. Chapman & Hall/CRC.

INFORMACIÓN TÉCNICA



Módulo: Probabilidad

Unidad 1: Estadística descriptiva

Escenario 1: Nociones básicas de R y RStudio

Autor: Alex Johann Zambrano

Asesor Pedagógico: Diana Marcela Díaz Salcedo

Diseñador Gráfico: Jully Amanda Guzman

Corrector de estilo: Felipe Garán

Asistente: Ginna Paola Quiroga

*Este material pertenece al Politécnico Grancolombiano.
Por ende, es de uso exclusivo de las Instituciones
adscritas a la Red Ilumino. Prohibida su reproducción
total o parcial.*