

Resumen del módulo de Manejo de Software

Programa Internacional de Estadística Aplicada a la Investigación Científica

Linda Cabrera

2022-10-20

Contents

Profesoras	1
Gabriela Castro	1
Linda Cabrera	1
Contenido del módulo	2
RStudio	2
Introducción a RStudio	3
Importación y orden de los datos en R	4
Manipulación de datos en R	4
Referencias	6

Profesoras

Gabriela Castro Centeno y Linda Cabrera Orellana son las profesoras que impartieron el módulo.

Gabriela Castro

Sociedad Ecuatoriana de Estadística

Directora Ejecutiva

Linda Cabrera

R-Ladies Guayaquil

Co-organizadora

Contenido del módulo

El módulo inició el lunes 3 de octubre de 2022 y finalizará el jueves 20 de octubre de 2022, abarcó los siguientes temas:

1. Introducción a R
2. Introducción a R Commander
3. Introducción a SPSS
4. RStudio
 1. Introducción a RStudio
 2. Importación y orden de los datos en R
 3. Manipulación de datos en R
 4. Análisis exploratorio de datos
 5. Comunicando resultados con R

El módulo tiene una duración de 25 horas y en cada capítulo de Rstudio se profundizaron temas relevantes para el contenido del Programa Internacional de Estadística Aplicada a la Investigación Científica. Por ejemplo en el capítulo de Importación y orden de los datos en R se revisaron los siguientes temas:

- Calidad de los datos
- Importar datos
 - Importar archivos `.csv`
 - Importar desde Excel
 - Importar desde SPSS y STATA
- Estructuras de datos
 - Vectores
 - Data.frames (Tibbles)
 - Funciones básicas
- Tipos de datos
 - Numéricos
 - Cadena de caracteres
 - Factores
 - Fechas y Horas

Horario del módulo:

Hora	Contenido
19:00 a 19:45	Bloque 1
19:45 a 20:50	Receso 1
20:50 a 20:35	Bloque 2
20:35 a 20:45	Receso 2
20:45 a 21:30	Bloque 3

RStudio

En la introducción de la parte de RStudio recordamos una frase de Hadley Wickhman para recordar que aprender R es aprender un nuevo idioma por lo que necesita práctica y paciencia:

“Cuando comienzas a programar, es fácil sentirte realmente frustrado... Es solo una parte natural de la programación, le sucede a todos y se vuelve cada vez menos con el tiempo. No te culpes. Solo tómate un descanso, haz algo divertido y luego regresa e inténtalo de nuevo más tarde.”

Hadley Wickham

Introducción a RStudio

1. Descargamos RStudio en el siguiente link
2. Instalamos RStudio en nuestras computadoras
3. Conocimos el entorno de RStudio
4. Aprendimos a instalar y cargar paquetes
5. Es posible usar R como una calculadora, donde podemos hacer operaciones desde muy sencillas a muy complicadas. Por ejemplo puedo calcular sin problema:

$$\sqrt{23^3 + 10} - \frac{2}{3} + \ln(1) - e^3$$

6. Aprendimos a crear objetos mediante asignaciones `<-` cuyo atajo de teclado es `Alt+-`.
7. Aprendimos qué es una función en R y que nos ahorran cálculos en nuestro análisis. R tiene funciones para casi todo lo que deseamos hacer en nuestros trabajos. Por ejemplo, para calcular la integral de una función puedo utilizar la función `integrate()`, en lugar de hacer el cálculo matemático paso a paso. Vamos a realizar el cálculo de una integral definida de forma analítica y a través de R:

Ejercicio: Calcular $\int_1^3 (x^2 + 4) dx$ Recordemos el *Teorema Fundamental del Cálculo*:

Teorema

Sea f una función integrable en el intervalo $[a, b]$ y $f = g'$ para alguna función g entonces:

$$\int_a^b f(x) dx = g(b) - g(a)$$

Resolvemos la integral definida:

$$\begin{aligned} \int_1^3 (x^2 + 4) dx &= \left| \frac{x^3}{3} + 4x \right|_1^3 \\ &= \left(\frac{3^3}{3} + 4(3) \right) - \left(\frac{1^3}{3} + 4(1) \right) \\ &= 9 + 12 - \frac{1}{3} - 4 \\ &= 17 - \frac{1}{3} = \frac{50}{3} = 16.67 \end{aligned}$$

Ahora utilizamos la función `integrate()` en R para calcular la misma integral. Si consultamos `?integrate` en la consola podemos observar que necesitamos 3 argumentos que son la función `function(x)`, el límite inferior `lower` y el límite superior `upper`: `integrate(function(x) (x^2 + 4), lower =1, upper = 3)`

El material de este capítulo se encuentra en Introducción a RStudio

Importación y orden de los datos en R



En este capítulo necesitamos instalar algunos paquetes:

- Imagen con tamaño definido:
- Imagen centrada con tamaño definido y caption:
Paquetes instalados en el capítulo 2 del Módulo

El material de este capítulo se encuentra en Importación y Orden de los Datos en R

Manipulación de datos en R

Este capítulo nos da una introducción práctica a los datos ordenados (o tidy data) y a las herramientas que provee el paquete `tidyr`, puedes leer más en *R para Ciencia de Datos* de Wickham & Grolemund (2017). Si deseas aprender más acerca de la teoría subyacente, puede que te guste el artículo Tidy Data de Wickham (2014) publicado en la revista Journal of Statistical Software, <http://www.jstatsoft.org/v59/i10/paper>.

```
profesores <- read_excel("Datos_R/r_profesores.xlsx")
```

Hicimos esta tablita:

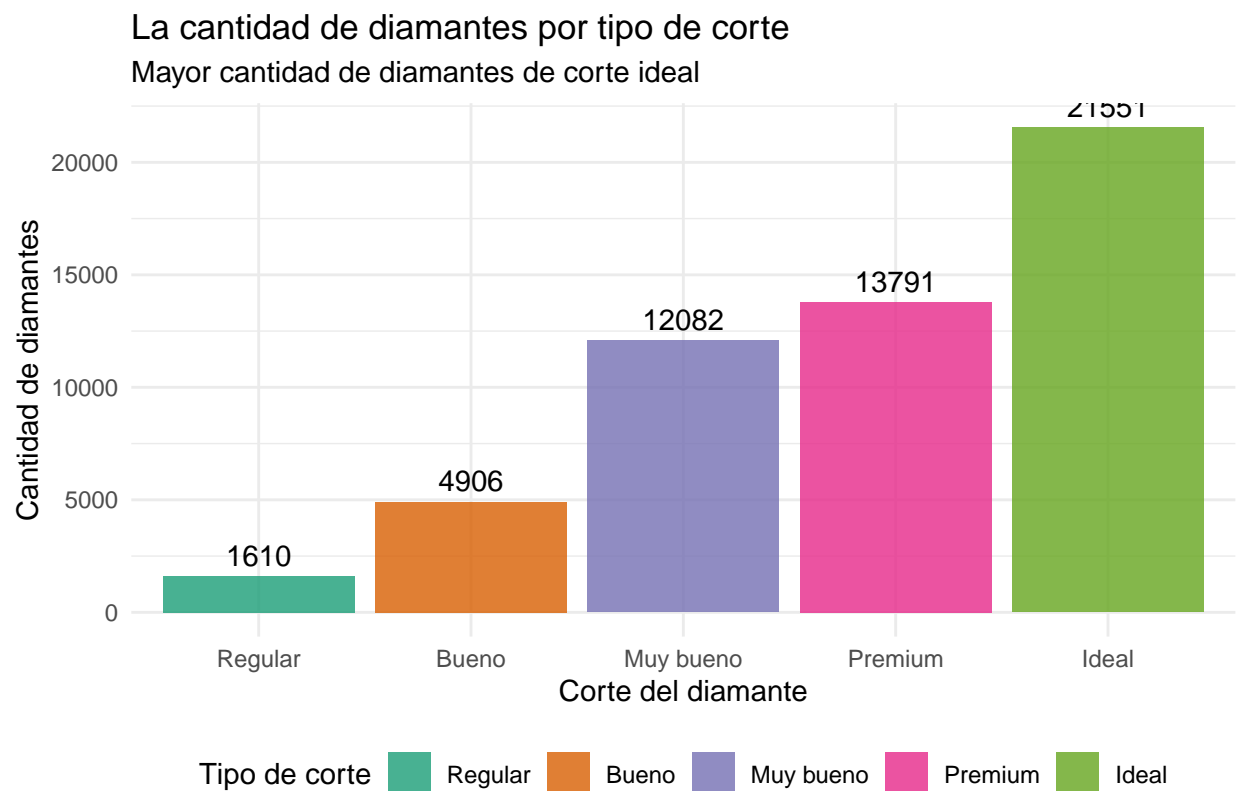
Table 2: Tabla 1. Resumen de indicadores

Nivel que imparte docencia	Profesores	Estudiantes	Estudiantes Promedio	Edad Promedio
Educación infantil, primaria o básica	98	3148	32	36
Educación secundaria y bachillerato	42	1398	33	42
Educación superior	35	1138	33	36

Usamos este código:

```
profesores %>%
  group_by(Nivel_docencia) %>%
  summarise(Profesores = n(),
            Estudiantes = sum(No_Alumnos),
            'Estudiantes Promedio' = round(Estudiantes/Profesores),
            'Edad Promedio' = round(mean(Edad)))%>%
  rename("Nivel que imparte docencia"=Nivel_docencia)
```

Se hace un gráfico con el paquete “Datos”:



Datos de R.com

El mismo gráfico pero solo el código:

```
diamantes %>% ggplot(aes(x = corte,
                        fill=corte)) +
  geom_bar(alpha=0.8) +
```

```

scale_fill_brewer(palette = "Dark2") +
geom_text(aes(label=..count..),
          stat = 'count',
          vjust=-0.5) +
theme_minimal() +
theme(legend.position = "bottom") +
labs(x = "Corte del diamante",
     y = "Cantidad de diamantes",
     title = "La cantidad de diamantes por tipo de corte",
     subtitle = "Mayor cantidad de diamantes de corte ideal",
     caption = "Datos de R.com",
     fill = "Tipo de corte")

```

Referencias

- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- Wickham, H., & Grolemund, G. (2017). *R for data science (2e)*. O'Reilly.