

Propuesta de enseñanza para la estadística de 1º de Bachillerato de Ciencias Sociales a través del lenguaje de programación R

Teaching proposal for statistics for the 1st year of Bachillerato of Social Sciences through the programming language R

Luis Felipe Del Río López¹

¹ Universidad de Granada.

lfdelriolopez@gmail.com

Resumen

El presente trabajo surge con el objetivo de introducir y promover el uso del lenguaje de programación R en el aula, en particular el programa de software libre RStudio. Para contextualizar dicho uso mostraremos una propuesta de su aplicación en la enseñanza de la estadística de 1º de Bachillerato de Ciencias Sociales, donde veremos la utilidad del citado software a través de una serie de tareas que propondremos a nuestros alumnos.

Palabras Clave

Propuesta de enseñanza, material didáctico, estadística, gráficos estadísticos, bachillerato, software libre, TIC.

Abstract

The present work arises with the aim of introducing and promoting the use of the programming language R in the classroom, in particular the free software program RStudio. In order to contextualize this use, we will show a proposal of its application in the teaching of statistics in the first year of Bachillerato of Social Sciences, where we will see the usefulness of this software through a serie of tasks that we will propose to our students.

Keywords

Teaching proposal, didactic material, statistics, statistical graphs, secondary school, free software, TIC.

1. Introducción

En Gea et al. (2015) se afirma que la presencia de recursos tecnológicos en los libros de texto referentes al tema de estadística bidimensional es escasa. En consecuencia, esto implica que los conjuntos de datos utilizados en los ejercicios sean pequeños y, en consiguiente, que las estimaciones de la correlación y regresión sean poco fiables. Además de ello, los alumnos dedican mucho tiempo a realizar cálculo en vez de dedicarlo a interpretar y a reflexionar sobre los datos tratados (Galindo, 2017). Es por ello que utilizaremos como material didáctico el lenguaje de programación R ya que, gracias a él, podremos realizar dichos cálculos con grandes conjuntos de datos y, además, utilizar el tiempo que ahorraremos en hacer dichos cálculos para aumentar la comprensión, interpretación y el espíritu crítico del alumnado frente a los datos tratados.

2. Descripción y clasificación del material didáctico



Figura 1: Logo RStudio

El material en el cual centramos nuestro trabajo es el lenguaje de programación R, en particular el entorno de desarrollo integrado RStudio. R es uno de los softwares de referencia en el mundo de la estadística, de hecho, es el séptimo lenguaje de programación más popular según el índice PPYL en junio de 2022. El lenguaje de programación R fue desarrollado inicialmente por los profesores Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993, los cuales combinaron los puntos fuertes de dos lenguajes de programación ya existentes en la época, S y Scheme. En la actualidad su desarrollo está a cargo del *R Development Core Team*.

El programa que utilizaremos será RStudio, el cual utiliza R como lenguaje de programación. Su principal ventaja con respecto a R es que incluye un editor de sintaxis, múltiples opciones de trazado y depuración y una interfaz que facilita la gestión del espacio de trabajo. Está disponible para los principales sistemas operativos (Windows, Mac y Linux) e, incluso, está disponible para navegadores conectados a los servidores de RStudio.

Las principales características del programa, las cuales se pueden consultar en <https://www.rstudio.com/products/rstudio>, son las siguientes:

- El resaltado de sintaxis, auto completado de código y sangría inteligente.
- Ejecutar código R directamente desde el editor de código fuente.
- Salto rápido a las funciones definidas.
- Administración sencilla de múltiples directorios de trabajo mediante proyectos.
- Navegación en espacios de trabajo y visor de datos.
- Depurador interactivo para diagnosticar y corregir los errores rápidamente.

A continuación, mostramos una imagen del programa, donde podremos apreciar claramente las cuatro partes en las que está dividida su interfaz gráfica:

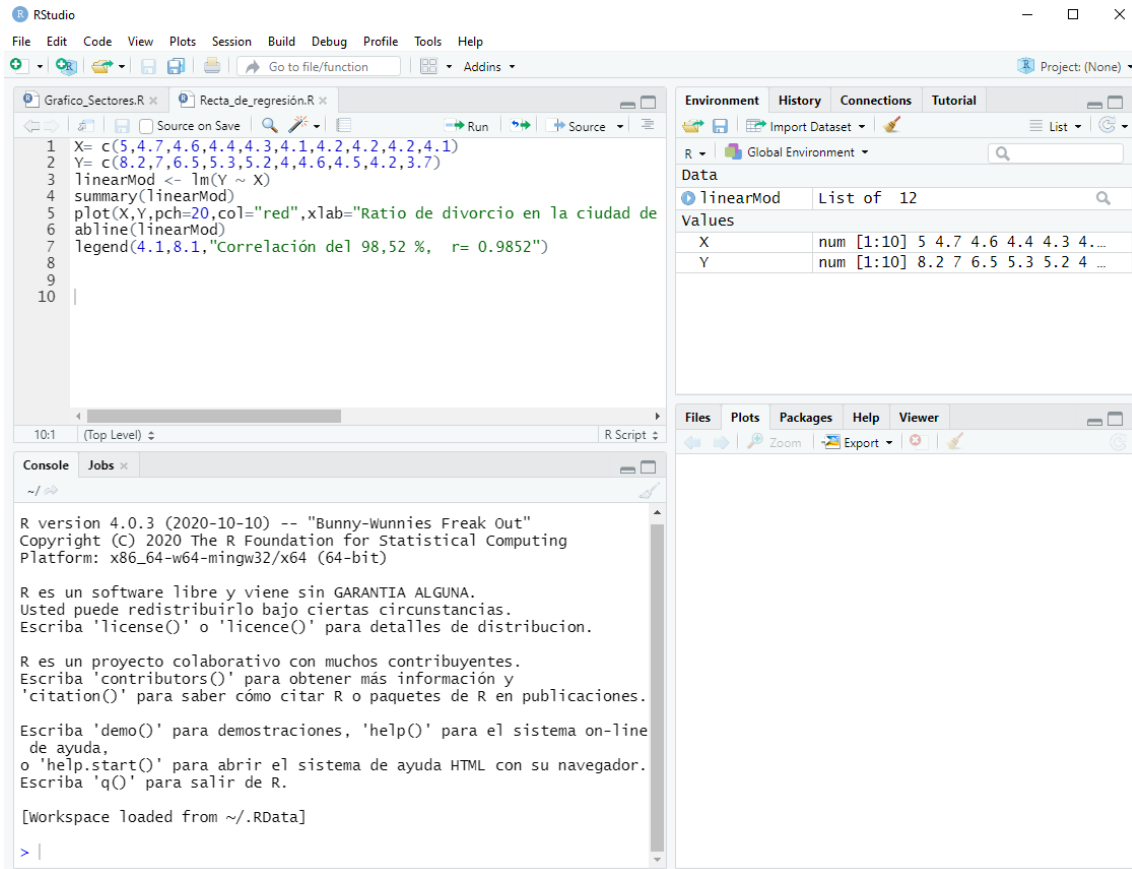


Figura 2: Interfaz de Rstudio

Para la clasificación de nuestro material (el software R) tomaremos como referencia los criterios descritos en Rico y Moreno (2016) y recogeremos dicha información en las siguientes tablas:

Utilidad	
Amplitud de su utilidad	Muy versátil.
Contenido	Álgebra, funciones, probabilidad, estadística.
Nivel educativo	ESO segundo ciclo, Bachillerato, Universidad.
Momento de utilización	Pre-instruccional, co-instruccional, post-instruccional
Tipo de tarea y actividad que facilita	Mostrar-observar, resolver problemas, proponer-manipular,

Tabla 1: Clasificación en función de la utilidad del material.

Formato	
Soporte	Informático
Accesibilidad	Fácil de encontrar en el mercado, software libre.
Grado de difusión	Muy específico

Tabla 2: Clasificación en función del formato del material.

3. ¿Por qué usar el lenguaje de programación R?

Una de las preguntas que, lícitamente, puede hacerse el lector es la siguiente: ¿por qué usar RStudio cuando, por ejemplo, programas como Excel también nos pueden ayudar en nuestro cometido y son mucho más conocidos por nuestros alumnos? Ciertamente es que muchas de las funciones que usaremos en RStudio están presentes en Excel y, además, de forma más intuitiva. Sin embargo, el uso de R en el aula tiene, a mi juicio, cuatro principales beneficios que justifican su elección en esta propuesta de enseñanza:

- 1) *R es poco conocido antes de la universidad:* R es, junto con SPSS, uno de los softwares más importantes en el ámbito de la estadística. Sin embargo, no es generalmente hasta la universidad cuando el alumno conoce de su existencia. Un primer contacto antes de ella mejorará el rendimiento de los alumnos en programación a la vez que trabajan la competencia digital.
- 2) *Es una apuesta de futuro:* Para problemas con muestras de pequeño tamaño, o que requieran poca potencia de cálculo, Excel es, a priori, mucho más cómodo. Sin embargo, en los problemas estadísticos reales se usan enormes cantidades de datos que requieren gran potencia de cálculo. Si el alumno aprende a usar R desde un primer momento podrá, más adelante, enfrentarse a problemas que no podría tratar de forma óptima con Excel. Además, R tiene muchas más aplicaciones, también fuera del ámbito de la estadística, lo cual lo hace un lenguaje de programación muy versátil.
- 3) *Permite presentar de forma fácil y elegante los resultados:* Gracias al paquete RMarkdown todos los gráficos, cálculos y comentarios que realice el alumno en RStudio podrán ser exportados en formato pdf, pudiendo crear así informes muy fácilmente. De esta forma, los alumnos podrán resolver los ejercicios en el propio programa y, sólo con darle a un botón, exportarlos en formato pdf para, por ejemplo, enviárselos a su profesor.
- 4) *Es Software Libre:* RStudio es Software Libre y, en consecuencia, es accesible a todo el alumnado de forma gratuita. Además, este software no necesita grandes requerimientos computacionales, por lo que podrá ejecutarse en prácticamente cualquier ordenador, eliminando una posible brecha económica-social.

4. Objetivos didácticos

Además de la estadística bidimensional propia de la unidad que estamos tratando se realizará, en primer lugar, un repaso de los contenidos de estadística unidimensional que nuestros alumnos han debido de tratar a lo largo de la ESO, especialmente en 4º. Gracias a dicho repaso se facilitará, en gran medida, el posterior aprendizaje de los conceptos de las distribuciones estadísticas bidimensionales, los cuales son totalmente nuevos para nuestro alumnado. Los objetivos que buscaremos en la siguiente unidad didáctica los hemos clasificado en tres focos, los cuales pasamos a presentar a continuación:

Foco 1: Leer y representar datos estadísticos

- O1.* Distinguir los tipos de fenómenos sociales que son susceptibles de ser tratados estadísticamente.
- O2.* Leer y confeccionar tablas estadísticas.
- O3.* Interpretar y confeccionar gráficos estadísticos.

Foco 2: Calcular parámetros estadísticos

- O4.* Hallar las medidas de centralización.
- O5.* Hallar las medidas de dispersión.
- O6.* Calcular el coeficiente de correlación.
- O7.* Calcular las rectas de regresión

Foco 3: Analizar e interpretar de forma crítica los resultados

- O8.* Interpretar las medidas de centralización.
- O9.* Interpretar las medidas de dispersión.
- O10.* Conocer el significado de la correlación entre variables.
- O11.* Interpretar el coeficiente de correlación.
- O12.* Hacer estimaciones a partir de las rectas de regresión.
- O13.* Analizar e interpretar de forma crítica informaciones estadísticas presentes en los medios de comunicación, publicidad y otros ámbitos, detectando posibles errores y manipulaciones tanto en la presentación de los datos como de las conclusiones.

Objetivo común a todos los focos

- O14.* Introducir y utilizar el software R en el aula.

La conexión de estos focos se puede apreciar en la siguiente figura:

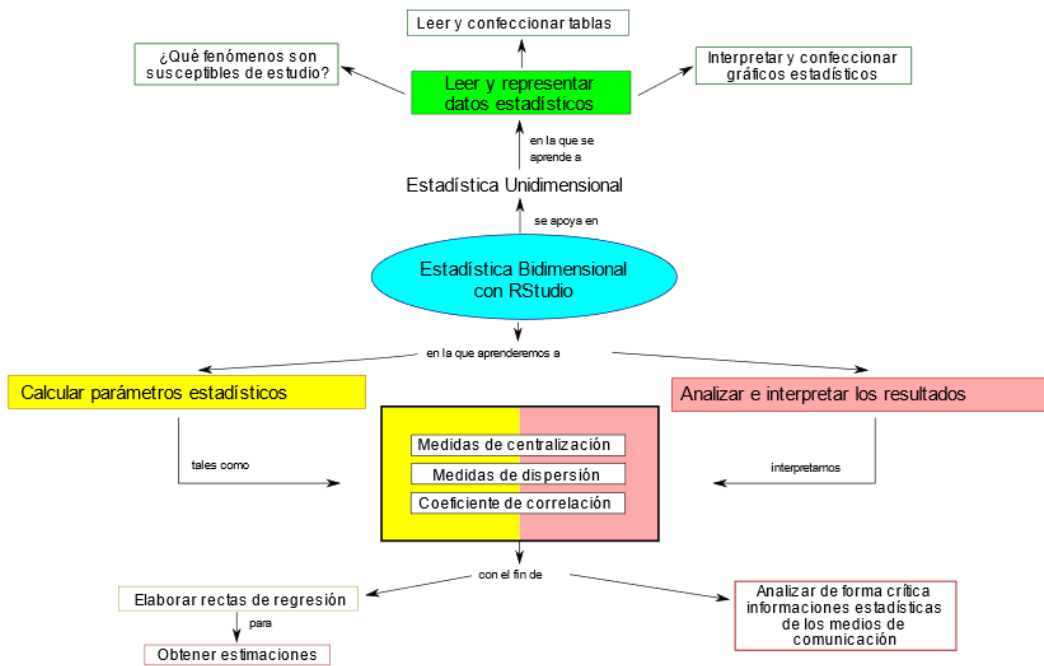


Figura 3: Mapa Conceptual.

Es decir, usaremos RStudio como soporte para alcanzar los objetivos de cada uno de los tres focos señalados.



Figura 4: Representación gráfica de los focos de objetivos.

5. Propuesta de tareas

A continuación, expondremos una serie de tareas en las que se pondrá de manifiesto el uso y utilidad de RStudio en el aprendizaje de la estadística frente al método “tradicional”.

- Tareas 1 y 2:

Las dos primeras tareas son de carácter puramente instructivo. En ellas nuestros alumnos tendrán una primera toma de contacto con RStudio y aprenderán a realizar sus primeras nubes de puntos y gráficos estadísticos. Dichas tareas, tituladas “Práctica 1” y “Práctica 2”, son las siguientes:

Práctica 1: Nociones básicas de RStudio

Cargar los datos

Una de las primeras cosas que debemos aprender es a crear vectores con los datos que vayamos a utilizar. Veamos un ejemplo:

```
Datos <- c(1,2,3,4,5,6,7,8,9,10)
```

La flecha indica que los datos se guardarán en el vector de nombre “Datos”, para crear vectores tendremos que escribir “ c() ”y dentro del paréntesis escribir los datos separados por comas.

Para ver los datos que hemos guardado basta con escribir el nombre del vector como sigue:

```
Datos
## [1]  1  2  3  4  5  6  7  8  9 10
```

Supongamos ahora que tenemos que vamos a trabajar con datos bidimensiones, al igual que estamos haciendo en clase. Para ello tendremos que crear dos vectores, uno con la primera coordenada de los datos y otro con la segunda.

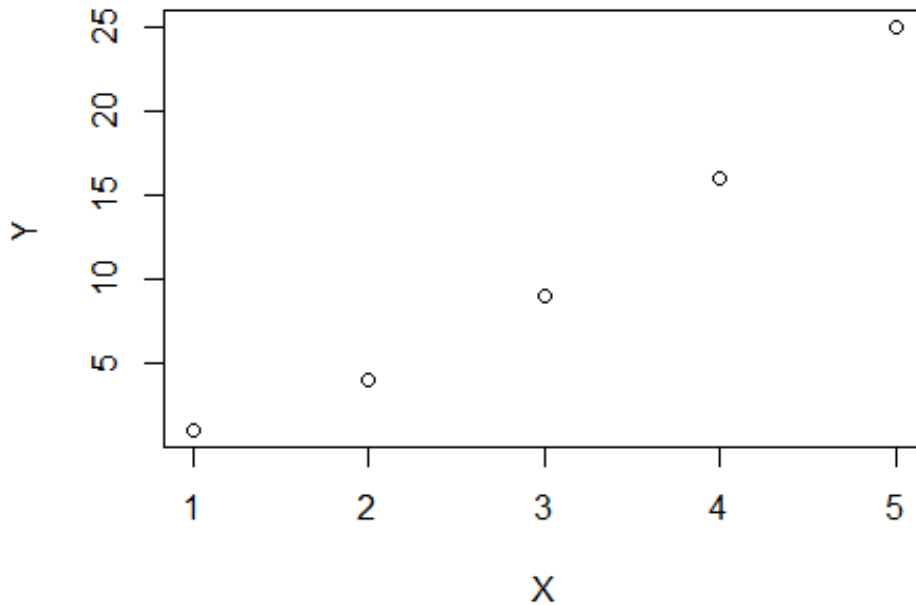
Para verlo más claro veamos un ejemplo de como guardaríamos los datos (1,1),(2,4),(3,9),(4,16),(5,25):

```
X= c(1,2,3,4,5)
Y= c(1,4,9,16,25)
```

Graficar los datos

Una vez guardados podremos usarlos para distintos propósitos, por ejemplo, graficarlos. Para ello usaremos la función plot:

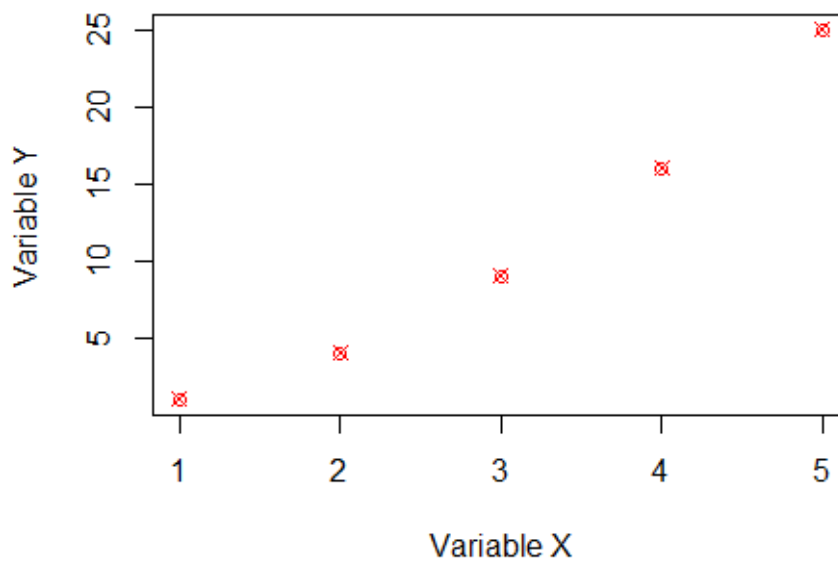
```
plot(X,Y)
```



Esta función tiene numerosas opciones, entre las que se encuentra cambiar el color y el tipo de puntos, incluir una leyenda, cambiar el título de los ejes... Para aprender cómo utilizar todas estas opciones basta con buscar “plot” en la pestaña de ayuda o bien escribiendo en la terminal “? plot”, ahí encontrareis con detalle la sintaxis de la función. Veamos un ejemplo de lo que podemos llegar a hacer con la función plot:

```
plot(X,Y,col="red",pch=13,main="Gráfico de ejemplo",
     xlab="Variable X", ylab="Variable Y")
```

Gráfico de ejemplo



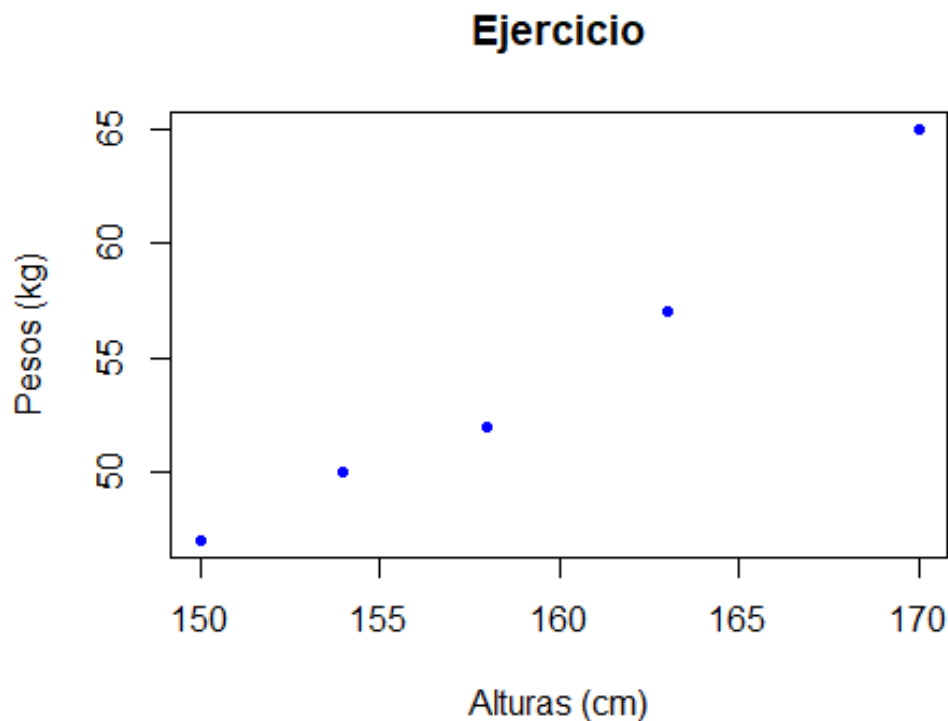
Ejercicio práctico

Crea dos vectores con las siguientes pesos y alturas y graficalos:

Pesos (kg)	47	50	52	57	65
Alturas (cm)	150	154	158	163	170

Solución

```
Pesos <- c(47,50,52,57,65)
Alturas <- c(150,154,158,163,170)
plot(Alturas, Pesos, col="blue", pch=20,
main="Ejercicio",xlab="Alturas (cm)",ylab="Pesos (kg)")
```



Práctica 2: Estadística descriptiva con RStudio

Estadísticos muestrales

RStudio es uno de los principales software estadísticos en la actualidad. Uno de sus puntos fuertes es que gracias a él podemos calcular fácilmente múltiples estadísticos muestrales como la media, varianza, desviación típica... En primer lugar cargaremos los datos del ejercicio de la práctica anterior, los cuales usaremos para aprender a calcular dichos parámetros:

```
Pesos <- c(47,50,52,57,65)
Alturas <- c(150,154,158,163,170)
```

Veamos ahora como se calculan en un conjunto de datos (haremos el ejemplo con los pesos):

```
mean(Pesos) # Calcula La media
## [1] 54.2

var(Pesos) # Calcula La cuasivarianza
## [1] 49.7

sd(Pesos) # Calcula La desviación típica
## [1] 7.049823
```

Además de estos parámetros podemos obtener el mínimo, el máximo, media y cuartiles de los datos haciendo uso de la función summary:

```
summary(Pesos)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      47.0   50.0   52.0   54.2   57.0   65.0
```

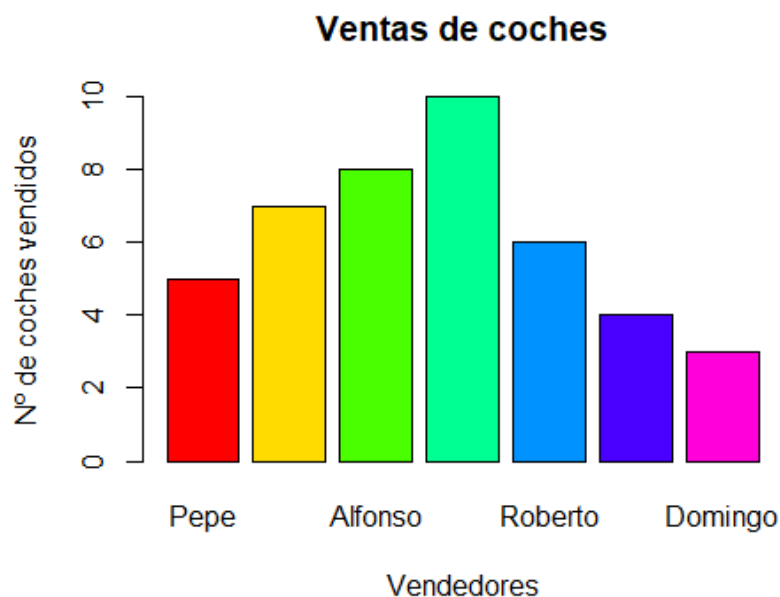
Gráficos estadísticos

Otro de los principales puntos fuertes de RStudio es el apartado gráfico, gracias a él podremos crear de forma muy sencilla gráficos estadísticos tales como diagramas de barras, gráficos de sectores, gráficos de caja y bigotes...

Veamos una serie de ejemplos:

- Gráfico de barras

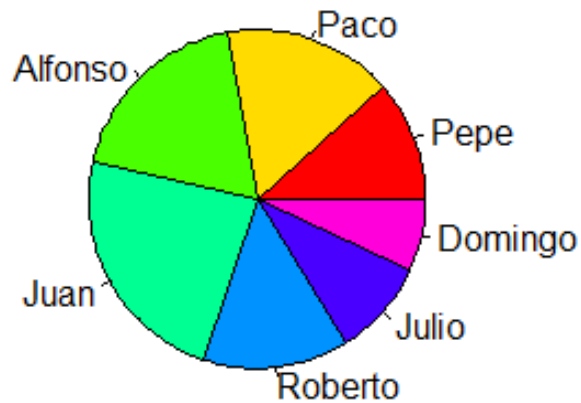
```
Datos <- c(5,7,8,10,6,4,3)
Nombres <- c("Pepe", "Paco", "Alfonso", "Juan",
"Roberto", "Julio", "Domingo")
barplot(Datos, names.arg = Nombres, main="Ventas de coches", xlab =
"Vendedores", ylab= "Nº de coches vendidos", col=rainbow(7))
```



- Gráfico de sectores

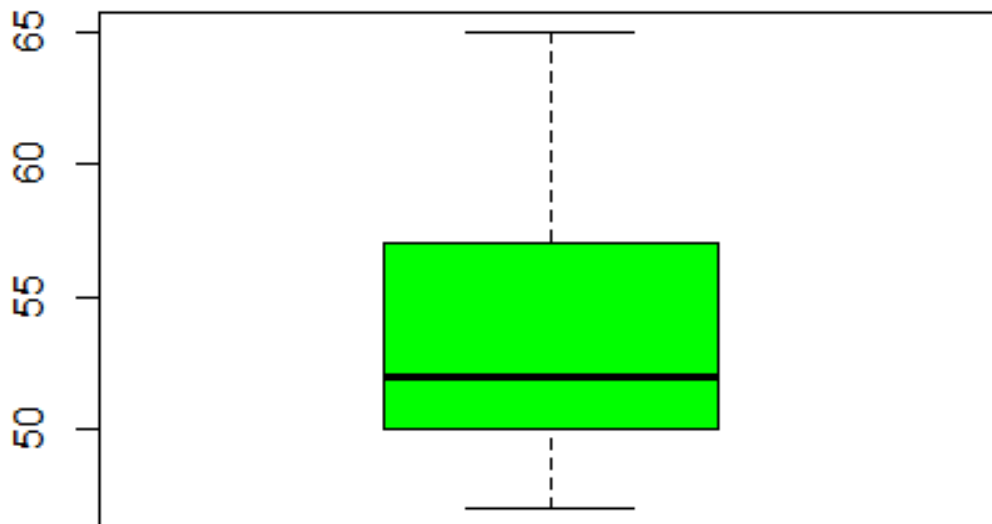
```
Datos <- c(5,7,8,10,6,4,3)
Nombres <- c("Pepe", "Paco", "Alfonso", "Juan",
"Roberto", "Julio", "Domingo")
pie(Datos, labels=Nombres, main="Ventas de coches", col=rainbow(7))
```

Ventas de coches



- Caja y bigotes

```
Pesos <- c(47,50,52,57,65)
boxplot(Pesos, col="green")
```



Todos estos gráficos tienen muchas más opciones, las cuales pueden ser consultadas haciendo uso de la pestaña de ayuda de Rstudio.

- Tarea 3: Análisis crítico de gráficos estadísticos

El enunciado de la tercera tarea que proponemos es el siguiente:

Los siguientes gráficos contienen errores o transmiten sensaciones incorrectas. Identifica dichos fallos y propón una solución para ellos. Finalmente, elige tres de los gráficos erróneos anterior y corrígelos haciendo uso de RStudio.



Gráfico 1



Gráfico 2

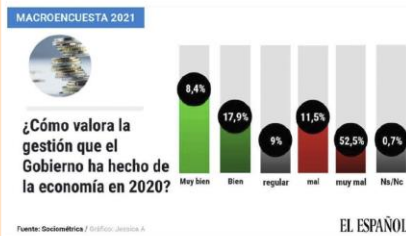


Gráfico 3

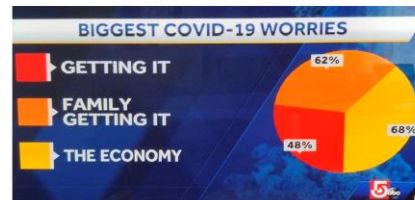


Gráfico 4

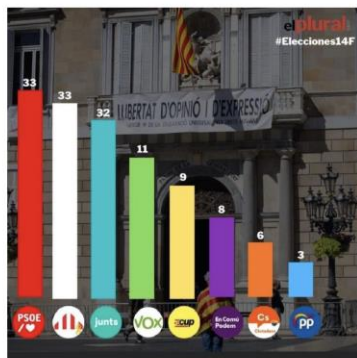


Gráfico 5

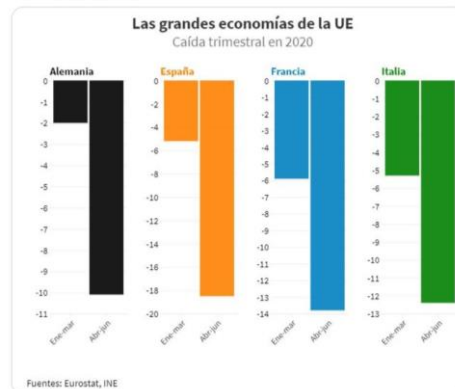


Gráfico 6

Figura 5: Gráficos aportados para la realización de la actividad.

Nótese que con esta tarea nuestros alumnos van a aprender a detectar errores en los medios de comunicación, corregirlos y rehacerlos haciendo uso de RStudio. Por tanto, con esta tarea hemos trabajado los objetivos O3, O13 y O14.

- Tarea 4: Ejercicio práctico de estadística descriptiva unidimensional

Esta tarea también tiene como objetivo seguir repasando la estadística descriptiva unidimensional. En esta ocasión el uso de RStudio será complementario ya que su uso se restringirá a los apartados 5 y 6. El enunciado es el siguiente:

El siguiente gráfico muestra las calificaciones de un grupo de 30 personas. Responda razonadamente a las siguientes preguntas:

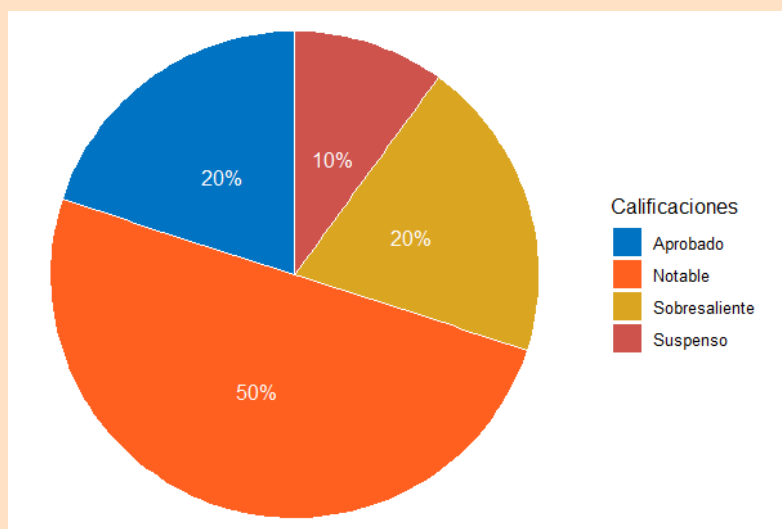


Figura 6: Gráfico, generado en RStudio, que se les proporciona a los alumnos.

- 1) ¿Qué tipo de gráfico es el anterior?
- 2) Elabora una tabla de frecuencias absolutas y frecuencias absolutas acumuladas.
- 3) ¿Cuántos alumnos tienen notable o sobresaliente?
- 4) ¿Es posible calcular la nota media de los alumnos a partir del gráfico dado? En caso contrario justifica tu respuesta.
- 5) Elabora, haciendo uso de RStudio, un gráfico de barras a partir de los datos proporcionados en el gráfico.
- 6) Supongamos que todos los *suspensos* sacaron un 3, todos los *aprobados* sacaron un 6, todos los *notables* un 8 y todos los *sobresalientes* un 10. Calcula la media, mediana, desviación típica y varianza de las notas de los alumnos; comprueba los resultados haciendo uso de RStudio.

Con esta tarea seguimos familiarizando a nuestro alumnado con la creación de gráficos estadísticos y les enseñamos a calcular los diferentes estadísticos muestrales. Por tanto, con esta tarea hemos trabajado los objetivos *O2*, *O3*, *O4*, *O5* y *O14*.

- Tarea 5: Regresión y correlación con RStudio

Esta tarea podrá ser realizada a la vez que se vea en clase todo el contenido teórico acerca de la correlación y la regresión de las distribuciones estadísticas bidimensionales. El objetivo de esta tarea es que el alumnado aprenda a realizar cálculos hechos haciendo uso del lenguaje de programación R y así poder comprobar que aquellos cálculos realizados a mano están hechos correctamente o simplemente agilizar el monótono proceso de calcular los distintos estadísticos muestrales. Adicionalmente permite al alumnado manipular rápidamente los datos del problema y así poder analizar las consecuencias de dichos cambios. Esta tarea es, al igual que la Tarea 1 y la Tarea 2, puramente instruccional. Esta tarea, titulada “Práctica 3”, es la siguiente:

Práctica 3: Regresión y correlación en R

En esta práctica aprenderemos a calcular las rectas de regresión y los coeficientes de determinación de una distribución bidimensional.

Para ejemplificar el proceso tomaremos los siguientes datos de ejemplo:

```
X <- c(1,2,3,5,7,9,10)
Y <- c(0,3,5,6,8,9,14)
```

A continuación calcularemos la recta de regresión de Y sobre X (para realizar la de X sobre Y basta con intercambiar los papeles de ambas variables en los sucesivos pasos). Para ello haremos uso de la función *lm* :

```
regresion= lm (formula = Y~X)
```

Mostraremos en pantalla ahora los datos de dicha recta:

```
summary(regresion)

##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      1      2      3      4      5      6      7
## -1.16732  0.60506  1.37743 -0.07782 -0.53307 -1.98833  1.78405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06031    1.07557  -0.056  0.957455
## X            1.22763    0.17351   7.075  0.000873 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.487 on 5 degrees of freedom
## Multiple R-squared:  0.9092, Adjusted R-squared:  0.891
## F-statistic: 50.06 on 1 and 5 DF, p-value: 0.0008726
```

Los datos que nos interesan de los anteriores son los que se encuentran en la sección *Coefficients* y el coeficiente de determinación (*Multiple R-squared*).

En nuestro caso:

(Intercept)	-0.06031
X	1.22763

y

Multiple R-squared: 0.9092

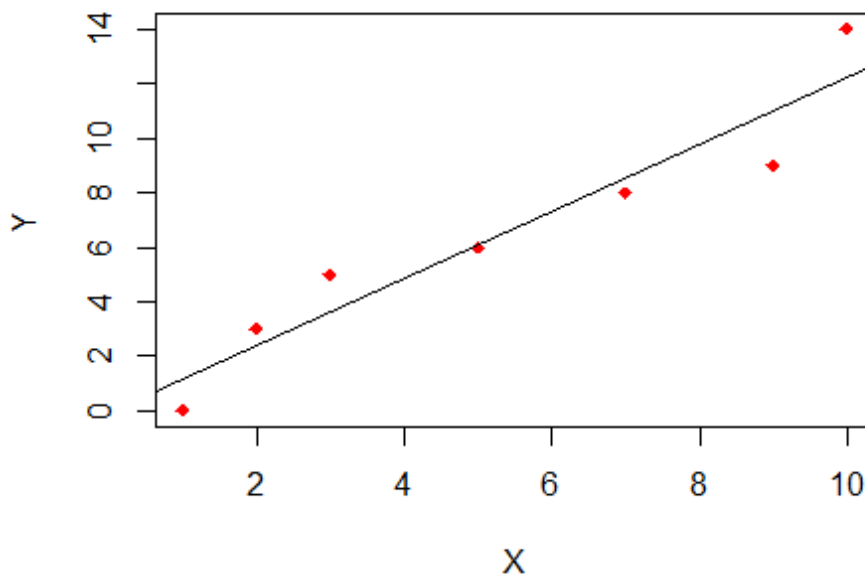
Estos datos nos dicen que la recta de regresión de Y sobre X es:

$$y = -0.06031 + 1.22763x$$

y que el coeficiente de determinación de la recta es 0.9092.

Finalmente pintaremos nuestros datos y la recta de regresión:

```
plot(X,Y,pch=18, col="red")  
abline(regresion)
```



Si quisiéramos calcular el coeficiente de correlación una posible forma de hacerlo sería calculando la raíz cuadrada al valor del coeficiente de determinación y poniendo el resultado con signo positivo o negativo dependiendo de si la correlación es positiva o negativa (se ve claramente en el dibujo). En este caso:

```
CoefCorrelacion = sqrt(0.9092)  
CoefCorrelacion
```

```
## [1] 0.9535198
```

con signo positivo ya que la correlación es positiva.

- Tarea 6: Correlación y causalidad

Esta tarea es la continuación natural de la tarea anterior. En ella se pretende que el alumno haga uso de todo lo aprendida en la Práctica 3 y, además, aprenda que correlación y causalidad no son conceptos equivalentes, lo cual es un error típico. Para ello realizaremos la siguiente tarea:

A partir de la siguiente tabla de datos responda a las preguntas:

X	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Y	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7

Tabla 3: Datos de la tarea 6.

- Calcula la recta de regresión de Y sobre X y gráficala con RStudio.
- ¿Qué coeficiente de determinación tiene dicha recta?
- A partir de los observado, ¿podemos suponer que variable X causa la variable Y? ¿Y viceversa?

Cuando nuestros alumnos realizan la tarea obtendrán una recta de regresión con un coeficiente de determinación muy cercano a uno, es decir, existirá una correlación muy fuerte entre las variables X e Y.

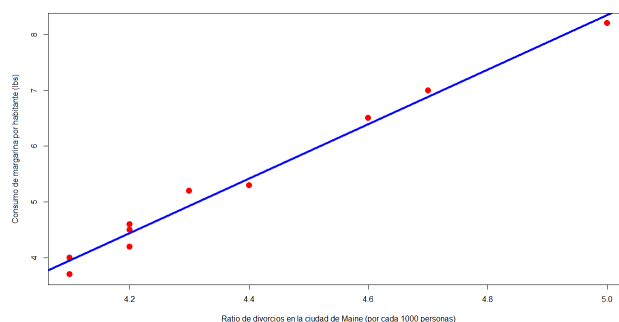


Figura 7: Nube de puntos y recta de regresión obtenida con RStudio.

Por tanto, algunos alumnos probablemente asuman que, en consecuencia, una variable causa la otra. Tras debatir esto con los alumnos, le diremos que la variable X corresponde al ratio de divorcios en la ciudad de Maine y la variable Y al consumo de margarina, en libras, por habitante. Ahí verán que, evidentemente, las variables no guardan ninguna relación. Finalmente, explicaremos que si dos variables X e Y están correlacionadas puede ser por los siguientes motivos:

- 1) X causa a Y.
- 2) Y causa a X.
- 3) Hay una tercera variable que causa a X e Y.
- 4) Pura coincidencia.

En esta tarea hemos, por tanto, trabajado los objetivos *O2*, *O6*, *O7*, *O11* y *O14*.

- Tarea 7: Realización de estimaciones

En la siguiente tarea usaremos de nuevo lo visto en la Práctica 3 y, además, haremos predicciones haciendo uso de la recta de regresión que calcularemos. El enunciado de la tarea es el siguiente:

Se midieron los valores de concentración, en microgramos por centímetro cúbico, de una sustancia A en suero fetal y los valores de su concentración en suero materno. Se obtuvieron los siguientes datos de una muestra de seis embarazadas al final de la gestión, los cuales vienen recogidos en la siguiente tabla:

Concentración suero madre (X)	5	4.7	4.6	4.4	4.3	4.1
Concentración suero feto (Y)	8.2	7	6.5	5.3	5.2	4

Tabla 4: Datos de la tarea 7

- Halla la recta de regresión que permita estimar la concentración de sustancia A en el suero fetal a partir de la correspondiente en suero materno.
- Halla el coeficiente de determinación e interprétalo para estudiar la bondad del ajuste.
- Determina la estimación de sustancia A en el suero fetal que se obtendrá si en el suero materno se han encontrado 10 microgramos por centímetro cúbico de dicha sustancia.

El cálculo de la recta de regresión y de los coeficientes de correlación y determinación es, haciendo uso de RStudio, bastante rápido y sencillo. Es por ello que el alumno, al realizar esta actividad, se centrará más en “saber qué está haciendo” que en calcular los correspondientes estadísticos muestrales a mano (no obstante, el alumno debe terminar la unidad sabiendo calcular a mano los distintos coeficientes, aunque finalmente opte por hacerlo de forma computacional).

Con esta tarea hemos trabajado los objetivos *O2*, *O6*, *O7*, *O11*, *O12* y *O14*.

- Tarea 8: ¿Cómo varía la media y la desviación típica si modificamos los datos de partida?

El objetivo de esta tarea es que el alumno compruebe, de forma computacional, las siguientes propiedades:

- Si la media de una muestra X es μ , entonces la media de $a \cdot X + b$ es $a \cdot \mu + b$.
- Si la varianza de una muestra X es σ^2 , entonces la varianza de $a \cdot X + b$ es $a^2 \cdot \sigma^2$.

Para ello les diremos que creen ellos mismos una muestra, que calculen su media y varianza, la media y varianza de $a \cdot X + b$ y que comprueben que, efectivamente, las propiedades anteriormente descritas se cumplen. Tras ello se demostrará, de forma teórica, dichas propiedades y se explicará su interpretación.

En esta tarea hemos, por tanto, trabajado los objetivos *O4*, *O5*, *O8*, *O9* y *O14*.

- Tarea 9: Importación de grandes conjuntos de datos

Hasta el momento hemos aprendido a graficar y hacer cálculos haciendo uso de datos que hemos introducido de forma manual. Al hacerlo de esta forma no hemos podido introducir grandes cantidades de datos y, en consecuencia, no hemos utilizado todo el potencial que nos ofrece RStudio. Esta tarea es la “Práctica 4, la cual procedemos a mostrar:

Práctica 4: Importación de grandes cantidades de datos con Studio

Hasta la fecha hemos trabajado con datos de poca envergadura ya que teníamos que ingresarlos de forma manual. En esta práctica aprenderemos a importar datos desde un archivo .csv, los cuales son archivos de texto en los que los datos están separados por comas (CSV = “Comma Separated Values”).

Una de las ventajas que tiene RStudio es que podemos, incluso, obtener los datos desde un link sin necesidad de descargarlo. Sin embargo, trabajaremos con un archivo previamente descargado para no necesitar conexión a internet. Tomaremos un archivo obtenido de la página el cual recoge las poblaciones de los distintos países desde el año 1800.

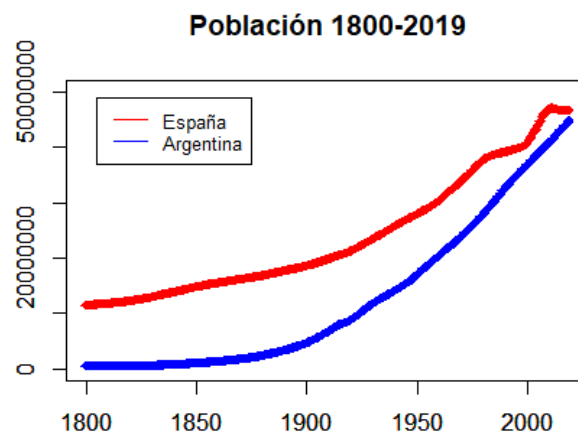
Para leer el archivo .csv haremos uso de la función read.csv, si los datos están separados por “punto y coma” en vez de comas será necesario usar la función read.csv2.

```
Datos <- read.csv(file="Poblacion.csv")
```

Tras ello tomaremos los datos de España y de Argentina haciendo uso de la función subset:

```
Spain <- subset (Datos, Entity== "Spain")  
Argentina <- subset (Datos, Entity== "Argentina")
```

Finalmente, haciendo uso de los conocimientos vistos en prácticas anteriores podremos graficar la población de España (en rojo) y de Argentina (en azul) a lo largo de los años:



Esto es simplemente un ejemplo de lo que se puede hacer, una vez que tenemos los datos importados podemos realizar hacer con ellos todo lo que hemos aprendido en las prácticas anteriores.

- Tarea 10: Creación de informes con RStudio haciendo uso del paquete RMarkdown.

Finalmente, se les pedirá a los alumnos que elijan un tema a estudio, recopilen los datos pertinentes, hagan un análisis completo de los mismos y, finalmente, den una serie de conclusiones. Para ello tendrán que hacer uso de todos los conocimientos vistos a lo largo de la unidad. Para la presentación de los resultados usarán el paquete RMarkdown. Dicho paquete es muy sencillo de usar y es con el que he generado las prácticas presentes en este artículo. Los resultados se pueden exportar en diversos formatos (pdf, word o html).

En esta tarea, por tanto, se trabajan todos y cada uno de los objetivos didácticos de la unidad, a excepción del *O13*.

6. Conclusiones

El uso del software R en el aula, en particular RStudio, no es un tema nuevo. Trabajos como los de Armero (2019), Briz y Serrano (2018), Revuelta (2017) y Alemany (2015) ya proponen su uso para la enseñanza de distintas ramas de las matemáticas: álgebra, análisis matemático o la propia estadística. En cierto modo, este trabajo surgió inspirado por el de Alemany (2015), que proponía su uso en la enseñanza de la estadística de 1º de Bachillerato de Ciencias Sociales, aunque, en su caso, a partir de proyectos.

La idea que perseguía este trabajo era explicar dicha unidad didáctica como se explicaría normalmente, pero introduciendo, de forma natural y complementaria, el software R tal y como se suele hacer, por ejemplo, con GeoGebra en las unidades de geometría.

Como se ha comentado en el trabajo, la implantación de RStudio tiene una serie de dificultades, como lo son la complejidad, en primera instancia, del lenguaje o el número de horas adicionales que habría que dedicar a dicha unidad didáctica. Sin embargo, tal y como se señaló, es una apuesta de futuro.

Si conseguimos que el alumnado consiga manejo con RStudio podremos usarlo en el tema de probabilidad, de forma similar a la que se hace, por ejemplo, en Revuelta (2017) ya que es muy útil para realizar simulaciones de experimentos probabilísticos. En consecuencia, sería muy interesante que el tema de estadística bidimensional no fuese simplemente un tema en el cual se pueda utilizar RStudio, sino el primero de una serie de temas de estadística y probabilidad que se traten con este software. Además de ello, mejoraremos el pensamiento computacional del alumno, lo cual le ayudará a la hora de razonar y, adicionalmente, a programar de forma más eficiente en caso de que lo necesiten en la universidad o en su futuro profesional.

7. Referencias

Alemaný, N. (2015). La estadística de 1º de Bachillerato a través de proyectos y el software R [*Trabajo de fin de máster*].

Armero, C. (2019). Algoritmos matemáticos en R: aplicaciones a la docencia y al cálculo científico [*Trabajo fin de grado*].

Briz, Á. & Serrano, Á. (2018). Aprendizaje de las matemáticas a través del lenguaje de programación R en Educación secundaria. *Educación matemática*, 30(1), 133-162.

Galindo, A. (2017). Didáctica con R. Menos cuentas y más pensamiento crítico. *Pensamiento Matemático*, 7(1), 53-73.

Gea, M.M., Batanero, C., Arteaga, P., Cañadas, G. R. & Contreras, J.M. (2015). Los recursos tecnológicos en la estadística bidimensional en los textos españoles de bachillerato. *Terné, Episteme y Didaxis: TED* (38).

PYPL Popularity of Programming Language (Junio de 2022). <https://pypl.github.io/PYPL.html>

Revuelta, A. (2017). R como herramienta de apoyo a la enseñanza de la Estadística en Bachillerato [*Trabajo fin de grado*].

Rico, L. & Moreno, A. (2016). Elementos de didáctica de la matemática para el profesor de secundaria. *Pirámide*.