



Taller 1

Integrantes:

Isaac David Mendoza Correa
Ronny David Sierra Vence
Julian David Arrieta Barriosnuevo
Richard Jose Ospino

Universidad Nacional Sede la Paz
Prof.: Jose Francisco Ruiz Munoz

1 Resumen

R es un lenguaje de programación principalmente enfocado al análisis estadístico, eso quiere decir que la mayoría de sus funciones permitirán al usuario tratar con grandes cantidades de información para luego describir los datos, hacer un análisis y presentarlo. Para poder hacer uso de **R**, se tiene que tener acceso al programa, y al ser **R** un software libre, lo que quiere decir que es gratis y de libre uso, bastará con acceder a la página del programa para proceder a instalarlo y hacer uso de ello. **R** es uno de los lenguajes de programación más utilizados en el área de ciencia de datos debido a su accesibilidad y su gran capacidad para manipular de diferentes maneras grandes cantidades de información.

2 Abstract

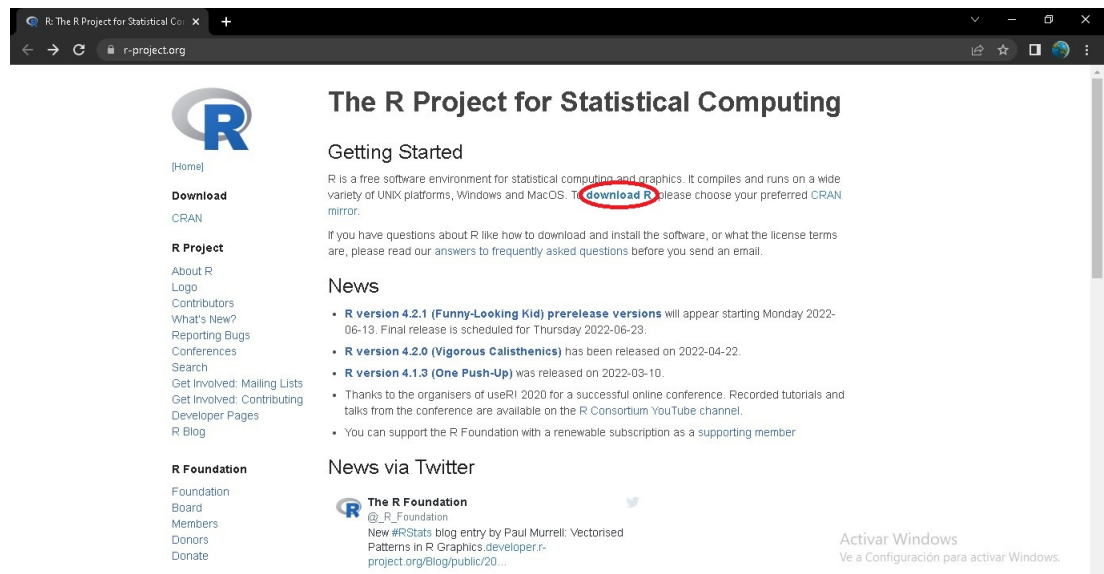
R is a programming language mainly focused on statistical analysis, that means that most of its functions will allow to the user to work with large amounts of information to later describe the data, make an analysis and present it. To be able to use R, you have to have access to the program, and being R a free software, which means that it is free and of free use, it would be enough to access the program page to proceed to install it and make use of it. R is one of the most widely used programming languages in the area of data science due to its accessibility and its great capacity to manipulate large amounts of information in different ways.

3 Pasos Para la instalación de R

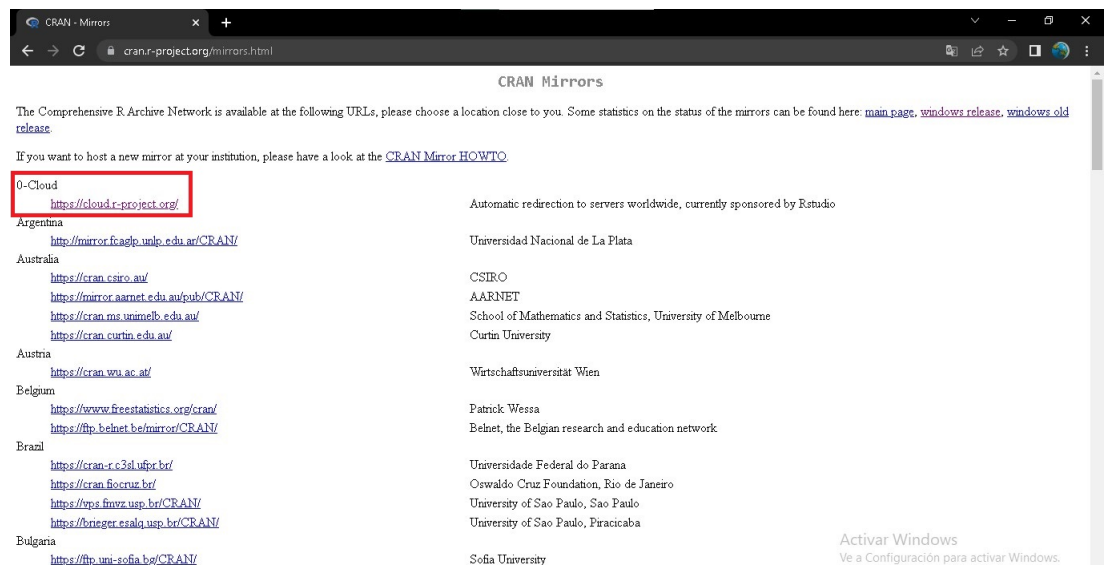
El primer paso para instalar R es acceder a su página web:

`https://www.r-project.org/`

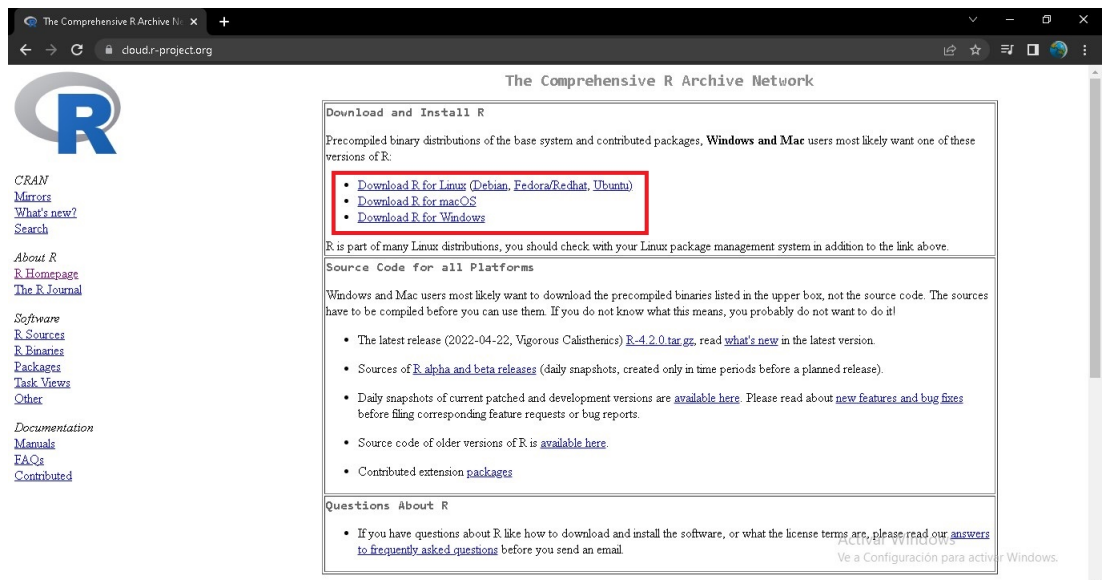
El enlace anterior conduce hacia la página principal de R donde se podrá conseguir distinta información sobre el programa y relacionados. Una vez en la página, dando click sobre la frase en azul "**download R**" (lo encerrado en rojo).



Esa acción conduce hacia otra página donde se podrá seleccionar el servidor de descarga (se recomienda seleccionar el encerrado en el cuadro rojo):

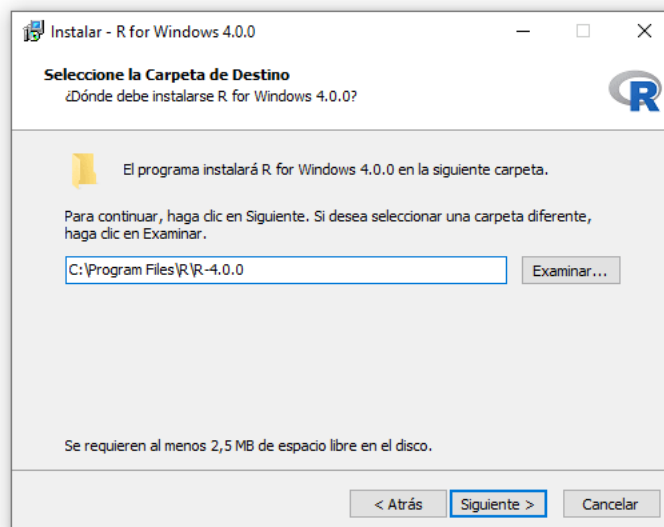


Una vez ahí seleccione para que sistema operativo quiere instalar el programa:

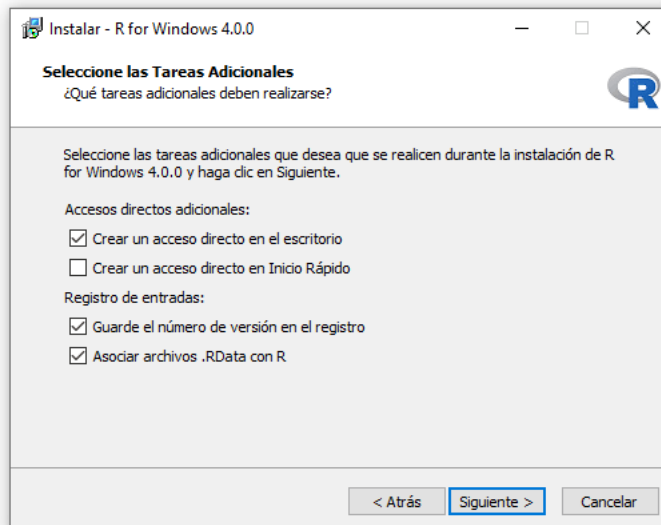


Una vez descargado, ejecuta el instalador y presiona siguiente en cada cuadro:

Seleccione donde quiere instalar el programa:



Por último puede elegir algunas preferencias respecto al programa:



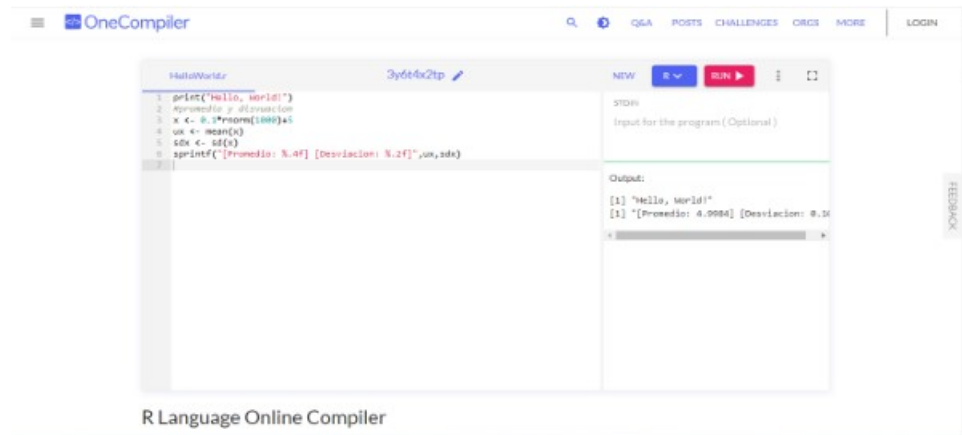
Una vez terminado eso ya se puede empezar a utilizar R

4 Ejecutar R en la web

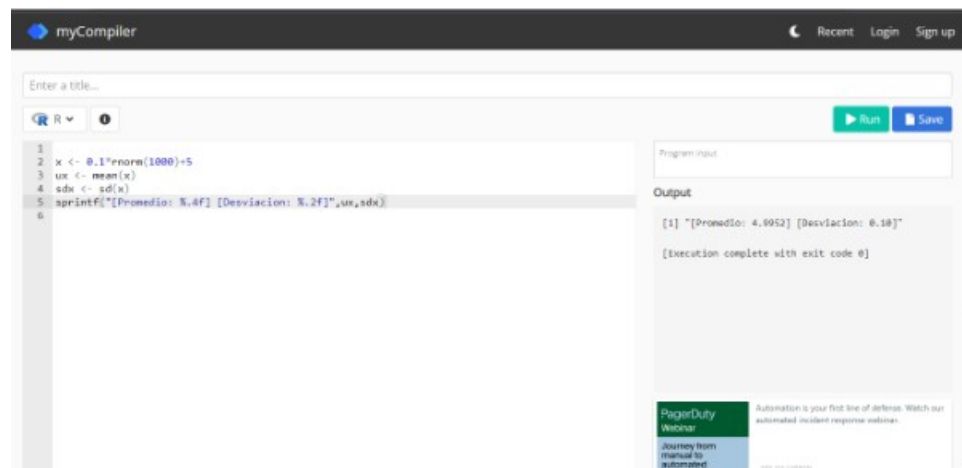
En la web existen muchas alternativas que nos permiten ejecutar códigos de **R**, en este tutorial se presentan algunas de ellas. Se trata de páginas web de acceso gratuito, fáciles de utilizar y que no es necesario ser experto en programación para poder manejarlas. Como primera opción presentamos Paiza, esta página web es de acceso gratuito y nos permite compilar códigos en más de 20 lenguajes de programación incluido **R**.



Otra alternativa útil es OneCompiler, es uno de los compiladores en línea más completos, ya que es rico en funciones para el lenguaje R y además permite ejecutar códigos con otros lenguajes de programación. Se puede escribir, ejecutar y compartir códigos en lenguaje R y cuenta con una guía informativa sobre la sintaxis del lenguaje de programación.



MyCompiler es un compilador en línea muy sencillo y fácil de usar. Al ingresar a la página se debe elegir el lenguaje con el cual se quiere compilar, se puede escribir, editar, ejecutar y guardar el código con el se está trabajando. No es necesario realizar el registro ni iniciar sesión en la página web.



Entre otras opciones de compiladores web para ejecutar códigos en lenguaje R se encuentran rdrv, uv.es, ideone.com, estadistica-dma.ulpgc.es. Todos estos son fáciles de utilizar, de acceso gratuito y sin requerimiento de registro e inicio de sesión.

5 Estructuras de datos básicas de R

R es un lenguaje de programación que posibilita operar con diversos tipos de datos, los cuales tienen la posibilidad de ver luego:

Tipo de dato	Descripción	Definición
Numeric	Números decimales	<code>numero <- 1.0</code>
Integer	Números enteros	<code>int <- 1</code>
Character	Cadenas de texto	<code>str <- "un texto"</code>
Complex	Números complejos	<code>comp <- 3+2i</code>
Logical	Verdadero (TRUE) o falso (FALSE). Es a menudo el resultado de operaciones lógicas.	<code>a <- 1; b <- 2; a < b</code>
Factor	Este no es estrictamente un tipo de dato, pero vale la pena describirlo aquí. Una variable factor es una variable categórica. Los vectores de caracteres a menudo se almacenan como factores para explotar funciones para tratar datos categóricos. Por ejemplo, en análisis de regresión.	Aplique <code>as.factor()</code> a un vector de caracteres.

Tal cual, se puede ver que R posibilita un desempeño simple en cualquier tipo de dato. Para este desempeño R usa, paralelamente, diferentes construcciones de datos.

5.1 vectores

La composición más sencilla es el vector, que es una recolección ordenada de números". (A cada vector, como cualquier otra operación en R, se le puede conceder una variable de forma que se conserve ahí toda la información que corresponde; en relación a los valores que dicho vector conserve, dichos tienen la posibilidad de ser en proporción de datos y los tipos tienen la posibilidad de ser cualquier persona de los que sean necesarios.

Por ejemplo:

Para producir un vector, que en esta situación se considerará como x, y que tenga 5 números como por ejemplo:4,5,2,34,12,5. Se utiliza la orden o comando:

```
x =c(4,5,2,34,12,5)
ejemplo 2:
```

x = c(4,15, "N", 9,8,"V") como se puede observar este vector tiene caracteres de tipo: float, int o str.

5.1.1 Operaciones entre los vectores

En R hay múltiples funciones para realizar operaciones en cualquier cantidad de número, estas funciones también se puede realizar en los vectores, las funciones que más se frecuentan son :

Función	Código en R
Suma	+
Resta	-
Multiplicación	*
División	/
Potenciación	^
Logarítmica	log
Exponencial	exp
Seno	sin
Coseno	cos
Tangente	tan
Raíz cuadrada	sqrt

Existen otros código los cuales son:

min(X)= este sirve para saber el número mínimo del vector

max(X)= este sirve para saber el numero máximo del vector

sum(X)= suma todos los datos que estén en el vector

prod(X)= encuentra el producto de todos los elementos

length(x)=Devuelve la longitud del vector, o la cantidad de elementos del vector

5.2 listas

En R, las listas son elementos formados por un conjunto ordenado de objetos. La naturaleza de una matriz se puede especificar mediante un vector numérico, un valor booleano, una matriz y una función. Para mostrar lo anterior, tenemos la siguiente función:

```
Lst = list(nombre="Arturo", esposa="Jimena", no.hijos=2,edad.hijos=c(3,8))
```

Lst se trata como una lista, y un elemento dentro de ella se puede ejecutar así:

```
Lst[[4]]
```

Este comando llama al elemento 4 de la lista; Ahora, dado que este es un elemento vectorial, dichos elementos vectoriales se pueden llamar individualmente de la siguiente manera:


```
Lst[[4]][1]
```

Por otro lado, las listas también se pueden enlazar usando la función `c()`, como se muestra a continuación:
`LisABC = c(LisA, LisB, LisC)`

5.3 data frame

Un marco de datos o mejor conocido en español es una hoja de datos que se puede pensar como un arreglo en el que las columnas contienen elementos con diferentes atributos. En el momento de la impresión, se puede considerar como una matriz que se diferencia de una matriz en la facilidad de visualización y organización de los datos con los que se está trabajando. La función de crear un Data Frame es:

```
Datos=.frame(  
Nombre=Alumnos  
Edad=Edades  
)
```

Dónde estudiante puede ser una variable de un vector con elementos de tipo `str` y edades es un vector con elementos de tipo `int` o `float`. Ahora, para llamar o indexar un Data Frame, esto se puede hacer con el signo `$` o de la forma en que se hace llamando a una columna o elemento en una matriz.

```
Datos[,1]  
Datos $ nombre
```

5.4 Attach y detach

Estas funciones se pueden implementar para acceder directamente a cada columna de un DataFrame, con `attach()`, el proceso comienza y termina con `detach()`.

```
attach(Datos)
```

```
Nombre
```

```
detach (Datos) # El acceso directo se realiza por completo, por lo que sí se  
llama a la variable de nombre, no aparecerá nada
```

5.5 Agregar y eliminar columnas y filas en un Data Frames

```
¿cbind ( ) #Esta agrega columnas
```

```
¿rbind ( ) #Es la que agrega filas
```

Para eliminar columnas en un marco de datos, puede usar el siguiente comando:

Primero, se crea otra variable que es un marco de datos con las columnas que desea conservar, y se puede ver como:

```
Datos1 = Datos[,c(1)]  
O se puede hacer de la siguiente manera:  
Datos1 = Datos[,c(Edad)]  
Con cualquiera de ellos, la columna 'Nombre' será eliminada.
```

5.6 Ordenar y Filtrar

Para ordenar datos en un bloque de datos, puede usar la función `order()`, así:

```
x = order( Datos$Edad) #Para ordenar los datos de menor a mayor  
x = order( - Datos$Edad) #Para ordenar los datos de mayor a menor
```

Para realizar el filtrado, cabe señalar que es el proceso de sustraer parte de la trama de datos si se cumple una condición necesaria.

Este procedimiento se puede realizar mediante el siguiente código:

```
subset(Datos, Edad=8)  
subset(Datos, Edad menor que 8)  
subset(Datos, Edad mayor que 8)
```

De esta forma se pueden obtener los datos requeridos del programador o del usuario.

6 Visualización de datos

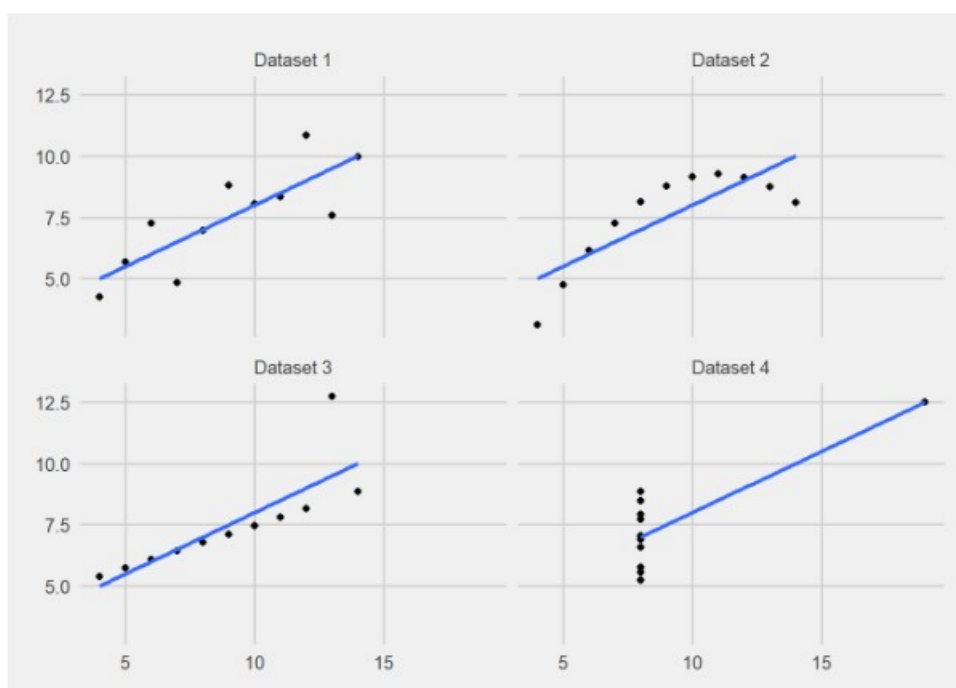
“Un simple gráfico ha brindado más información a la mente del analista de datos que cualquier otro dispositivo”-John Tukey

La visualización de datos ha presentado diversas posibilidades de un mejoramiento de análisis, sin embargo hay un punto que es en especial importante para la investigación de datos: las fronteras y coeficientes con los que solemos laborar no continuamente son tan básicas de interpretar como pensamos. Ejemplificando, ciertos autores recomiendan poderosamente graficar las predicciones del modelo frente a diversos valores (McElreath, 2016).

¿Qué es un data frame en R?: Son el objeto más distinguido para guardar datos en R. En esta clase de objeto, cada persona o fecha corresponde a una fila y cada columna corresponde a una variable. Los data frames son construcciones de datos bastante semejantes a las matrices, empero en la situación de los data

frames puedes tener diversos tipos de datos en las columnas.

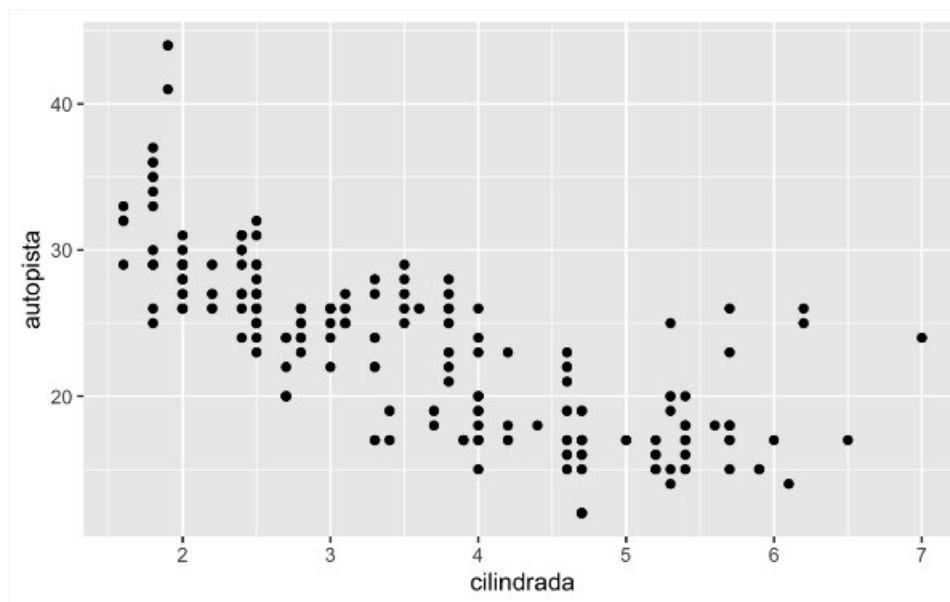
```
# Creamos vectores con los valores
nombre <- c("Juan", "Margarita", "Ruben", "Daniel")
apellido <- c("Sanchez", "Garcia", "Sancho", "Alfara")
fecha_nacimiento <- c("1976-06-14", "1974-05-07", "1958-12-25", "1983-09-19")
sexo <- c("HOMBRE", "MUJER", "HOMBRE", "HOMBRE")
nro_hijos <- c(1, 2, 3, 4)
##      nombre apellido fecha_nacimiento  sexo nro_hijos
## 1      Juan  Sanchez    1976-06-14 HOMBRE      1
## 2 Margarita   Garcia    1974-05-07  MUJER      2
## 3      Ruben   Sancho    1958-12-25 HOMBRE      3
## 4    Daniel   Alfara    1983-09-19 HOMBRE      4
```



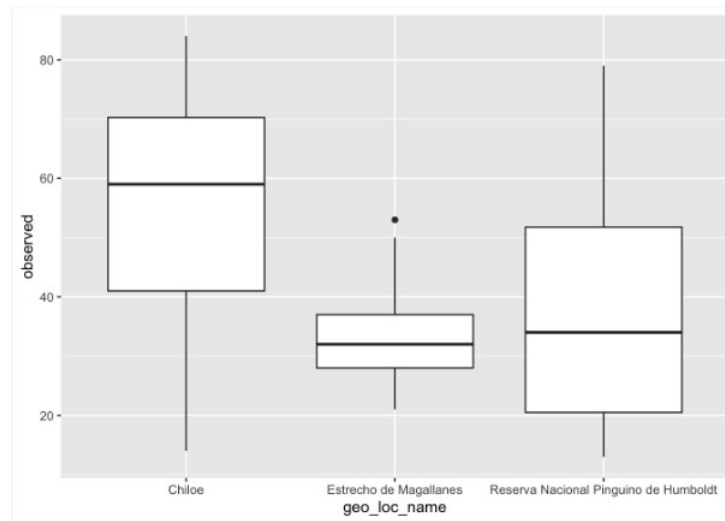
ggplot2: es uno de los paquetes principales que nos facilita crear gráficas con datos en un data frame y de los más elegantes y versátiles. Implementa un sistema coherente para describir y construir gráficos, conocido como la gramática de gráficos.

Ejemplos:

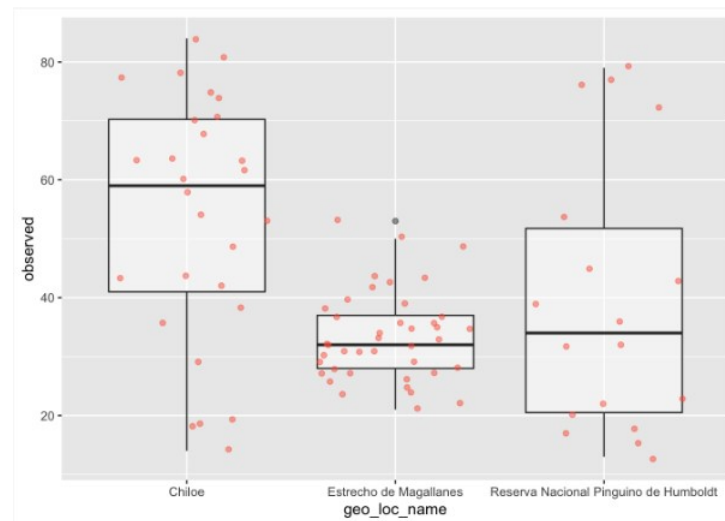
```
ggplot(data = millas) +  
  geom_point(mapping = aes(x = cilindrada, y = ←  
    autopista))
```



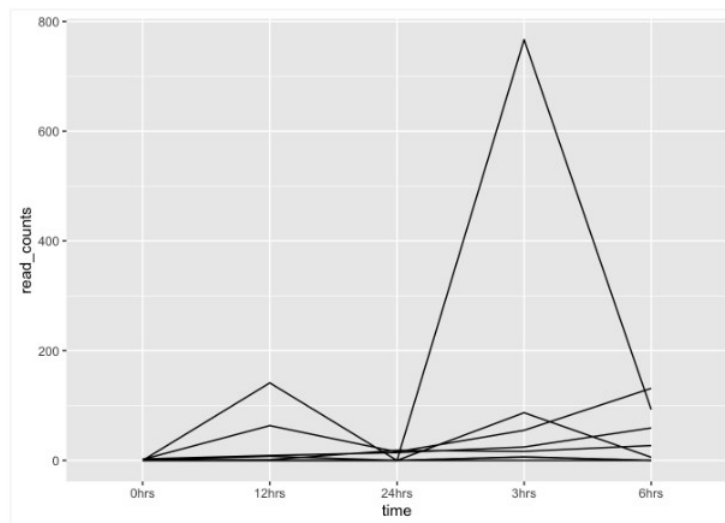
```
ggplot(data = data, aes(x = geo_loc_name, y = ←  
  observed)) +  
  geom_boxplot()
```



```
ggplot(data = data, aes(x = geo_loc_name, y = observed)) +
  geom_boxplot(alpha = 0.5) +
  geom_jitter(alpha = 0.5, color = "tomato")
```



```
ggplot(data = tr_plot, aes(x = time, y = read_counts, group = transcripts)) +
  geom_line()
```



7 Referencias

1. Data frames · ciencia-de-datos-con-r. (n.d.). Rubén Sánchez Sancho. Retrieved June 11, 2022, from https://rsanchezs.gitbooks.io/ciencia-de-datos-con-r/content/estructuras_datos/data_frames/data_frames.html
2. Visualización de datos usando ggplot2 — Diseño experimental y análisis de datos. (n.d.). Castro Lab. Retrieved June 11, 2022, from http://www.castrolab.org/teaching/data_analysis/visualizacion-de-datos-usando-ggplot2.html
3. Visualización de datos — _main. (n.d.). R para Ciencia de Datos. Retrieved June 11, 2022, from <https://es.r4ds.hadley.nz/visualizaci%C3%B3n-de-datos.html>
4. Visualizaciones de datos en R — Ciencia de datos para curiosos. (n.d.). Bookdown. Retrieved June 11, 2022, from <https://bookdown.org/martinmontaneb/CienciaDeDatosParaCuriosos/visualizaciones-de-datos-en-r.html>
5. ¿Cómo instalar R?. *R CODER*. <https://r-coder.com/instalar-r/>