


# Estadística y modelación de sistemas socioecológicos en R



Laboratorio  
Nacional  
de Ciencias  
de la Sostenibilidad

**Dra. Yosune Miquelajauregui Graf**

# Plan del día

1. Manipulación de datos : match values
  2. Markdowns en R
  3. Introducción a la estadística
  4. Estadística descriptiva
- 

# Manipulación de datos : match

En ocasiones necesitamos identificar los atributos que coinciden entre dos o más hojas de datos.

```
head(HD1)
```

```
ID V1 V2
```

```
1  1 0.31561892
```

```
2 11 0.64778713
```

```
3 21 0.35609124
```

```
4 31 0.82515905
```

```
5 41 0.49330436
```

```
6 51 0.8434979
```

```
length(HD1[,1])
```

```
1000
```

```
head(HD2)
```

```
ID    D1    D2
```

```
1  3.867982  52
```

```
2  7.241679  43
```

```
3  6.434363  34
```

```
4  9.706802  85
```

```
5  7.612441  66
```

```
6  6.010280   8
```

```
length(HD2[,1])
```

```
72
```

# Manipulación de datos : match

Encontrar los elementos de HD1 que coincidan con los elementos de HD2 con base en una variable en común, en este caso la variable ID.

```
HDMatch <- HD1[HD1$ID %in% HD2$ID, ]
```

```
head(HDMatch)
```

ID	V1	V2
1	1	0.31561892
2	11	0.64778713
3	21	0.35609124
4	31	0.82515905
5	41	0.49330436
6	51	0.8434979

```
length(HDMatch[,1])
```

```
72
```

# Manipulación de datos : match

```
Todo <- merge(HDMatch,HD2, by="ID")
```

```
head(Todo)
```

ID	V1	V2	D1	D2
1	1	0.3156189	3.867982	5
2	11	0.6477871	7.241679	43
3	21	0.3560912	6.434363	34
4	31	0.8251590	9.706802	85
5	41	0.4933043	7.612441	66
6	51	0.8434979	6.010280	8

```
length(Todo[,1])
```

```
72
```

# Markdowns en R

Los Markdowns en R son un formato para escribir reportes reproducibles y dinámicos en R. Se pueden utilizar para introducir código escrito en R, resultados de análisis y gráficas en diapositivas, pdfs, documentos html, documentos word y más. Permite además llevar un historial de los análisis realizados.

# Markdowns en R

**rmarkdown-cheatsheet**

# R Markdown Cheat Sheet

learn more at [rmarkdown.rstudio.com](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf)  
rmarkdown 0.2.50 Updated: 8/14

i. **Open** - Open a file that uses the .Rmd extension.

ii. **Write** - Write content with the easy to use R Markdown syntax

iii. **Embed** - Embed R code that creates output to include in the report

iv. **Render** - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.

```
.Rmd → A report. A plot: → A report. A plot: {r} hist(co2) → A report. A plot: {r} hist(co2) = A report. A plot: [PDF icon] [Word icon] [HTML icon]
```

## 2. Open File

Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ► New File ► R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK

## 3. Markdown

Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

syntax	becomes
Plain text End a line with two spaces to start a new paragraph. <i>*italics*</i> and <i>_italics_</i> <b>**bold**</b> and <b>__bold__</b> <sup>superscript^2^</sup> <del>--strikethrough--</del> <a>[link] (www.rstudio.com)</a>	Plain text End a line with two spaces to start a new paragraph. <i>italics</i> and <i>italics</i> <b>bold</b> and <b>bold</b> <sup>superscript</sup> <del>strikethrough</del> <a href="#">link</a>
# Header 1 ## Header 2 ### Header 3 #### Header 4 ##### Header 5 ##### Header 6	Header 1 Header 2 Header 3 Header 4 Header 5 Header 6
endash: -- emdash: --- ellipses: ... inline equation: \$A = \pi r^2\$ image:	endash: – emdash: — ellipses: … inline equation: $A = \pi * r^2$ image:
horizontal rule (or slide break): ***	horizontal rule (or slide break): ---
> block quote	block quote
* unordered list + item 2 + sub-item 1 + sub-item 2	• unordered list • item 2 • sub-item 1 • sub-item 2
1. ordered list 2. item 2 + sub-item 1 + sub-item 2	1. ordered list 2. item 2 • sub-item 1 • sub-item 2
Table Header   Second Header ----- Table Cell   Cell 2	Table Header   Second Header ----- Table Cell   Cell 2

## 4. Choose Output

Write a YAML header that explains what type of document to build from your R Markdown file.

**YAML**

A YAML header is a set of key:value pairs at the start of your file. Begin and end the header with a line of three dashes (---)

```
---
title: "Untitled"
author: "Anonymous"
output: html_document
---
```

This is the start of my report. The above is metadata saved in a YAML header.

The RStudio template writes the YAML header for you

The output value determines which type of file R will build from your .Rmd file (in Step 6)

output: **html\_document** ..... html file (web page)

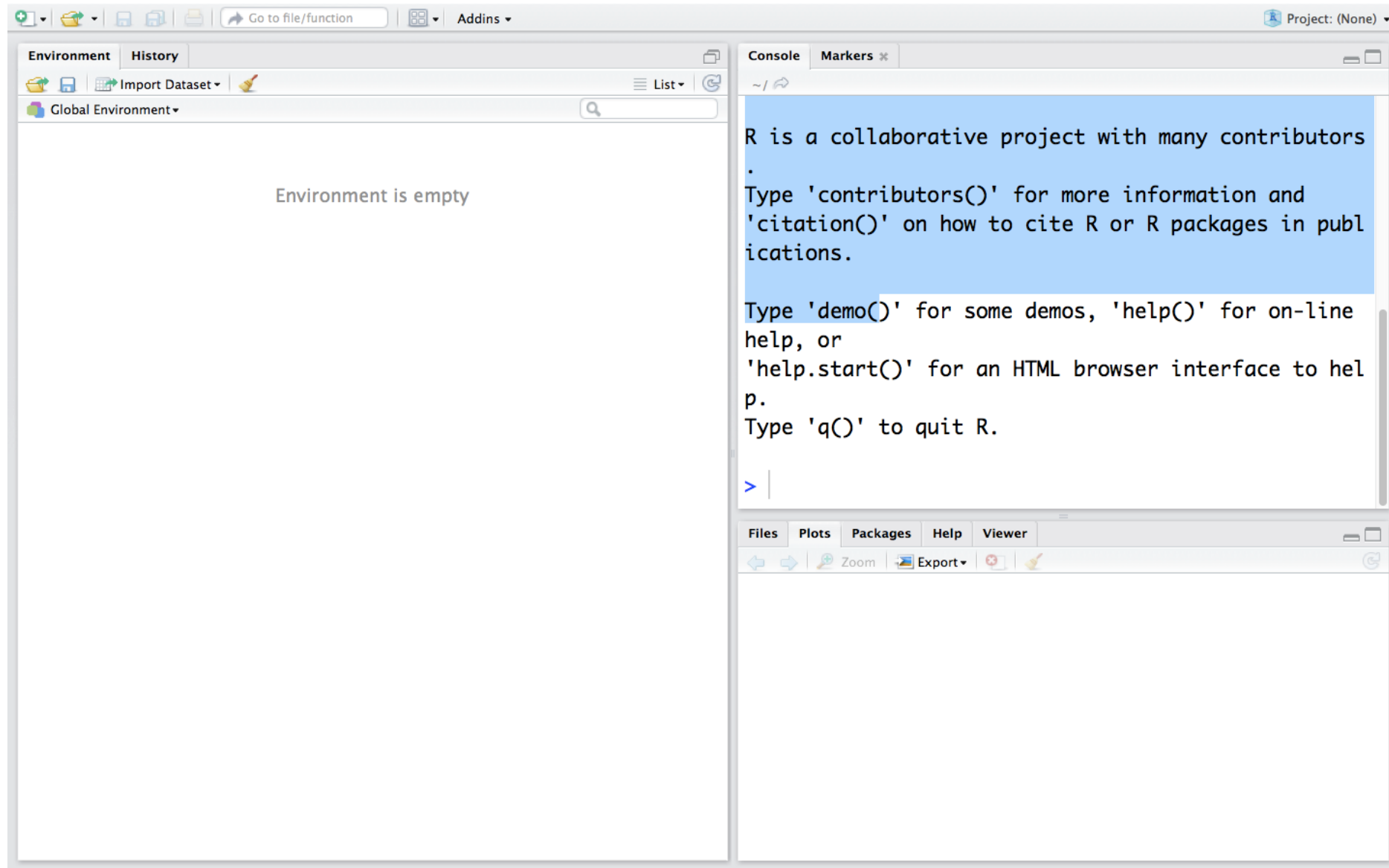
output: **pdf\_document** ..... pdf document

output: **word\_document** ..... Microsoft Word .docx

output: **beamer\_presentation** ..... beamer slideshow (pdf)

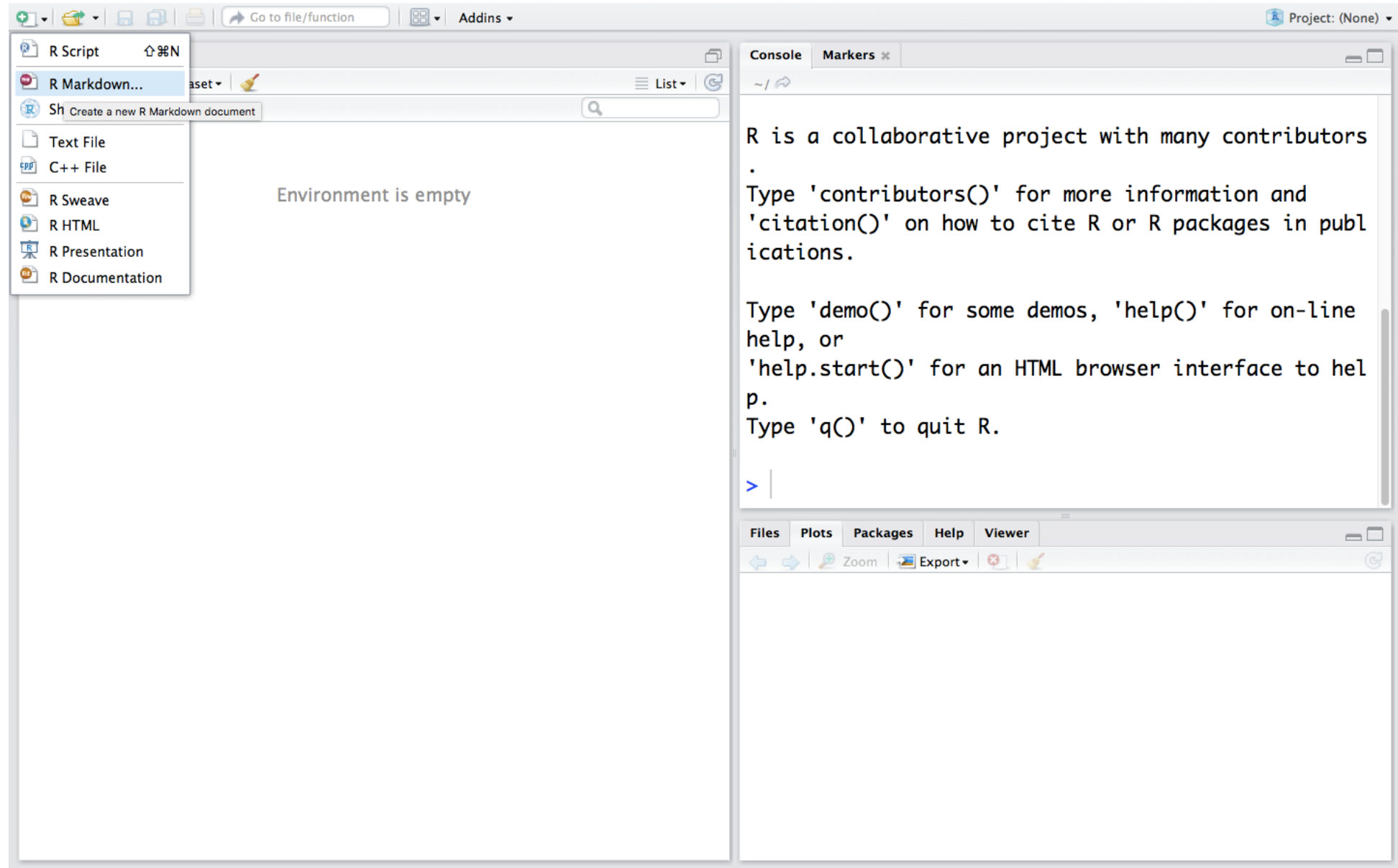
output: **ioslides\_presentation** ..... ioslides slideshow (html)

# Markdowns en R

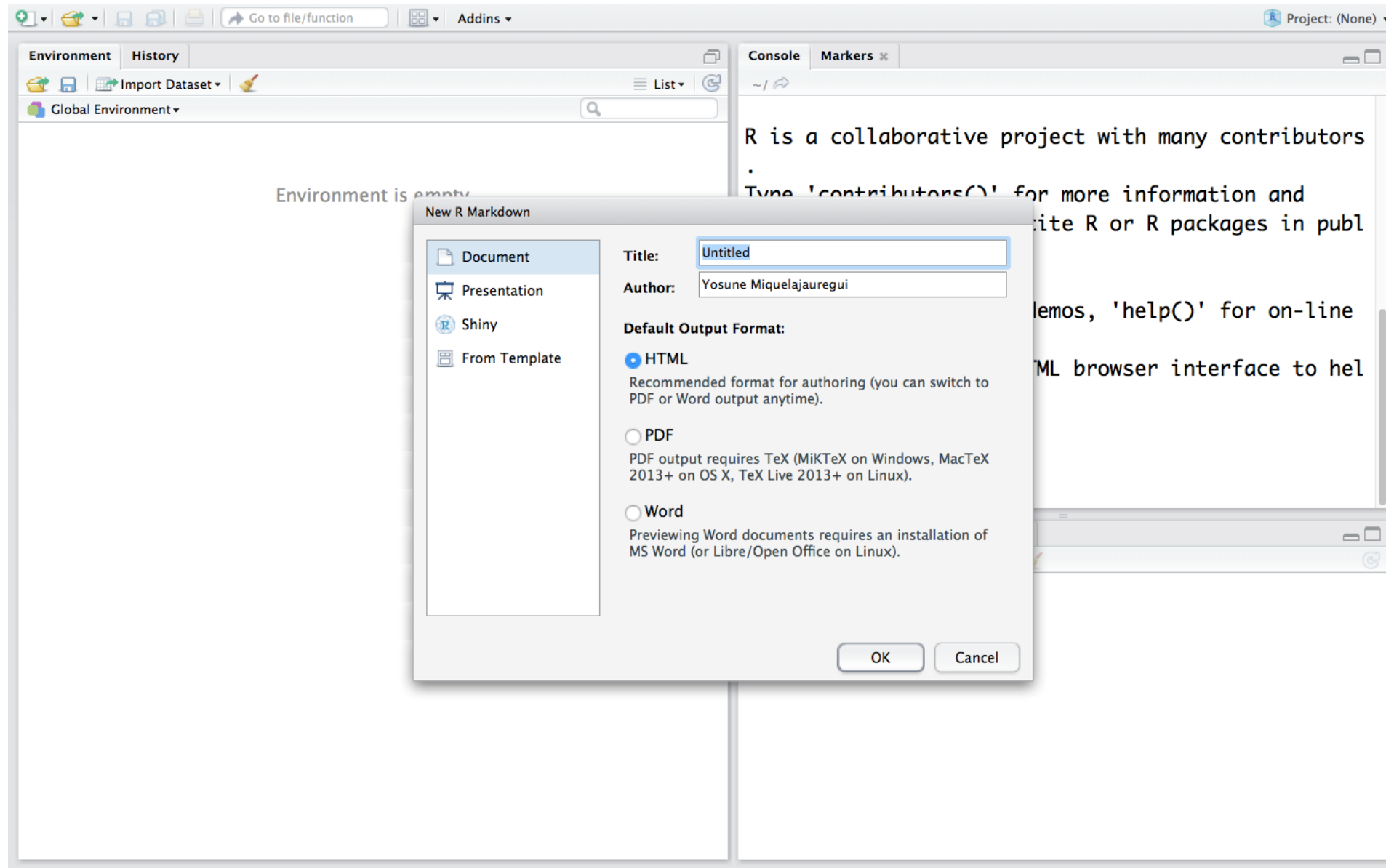




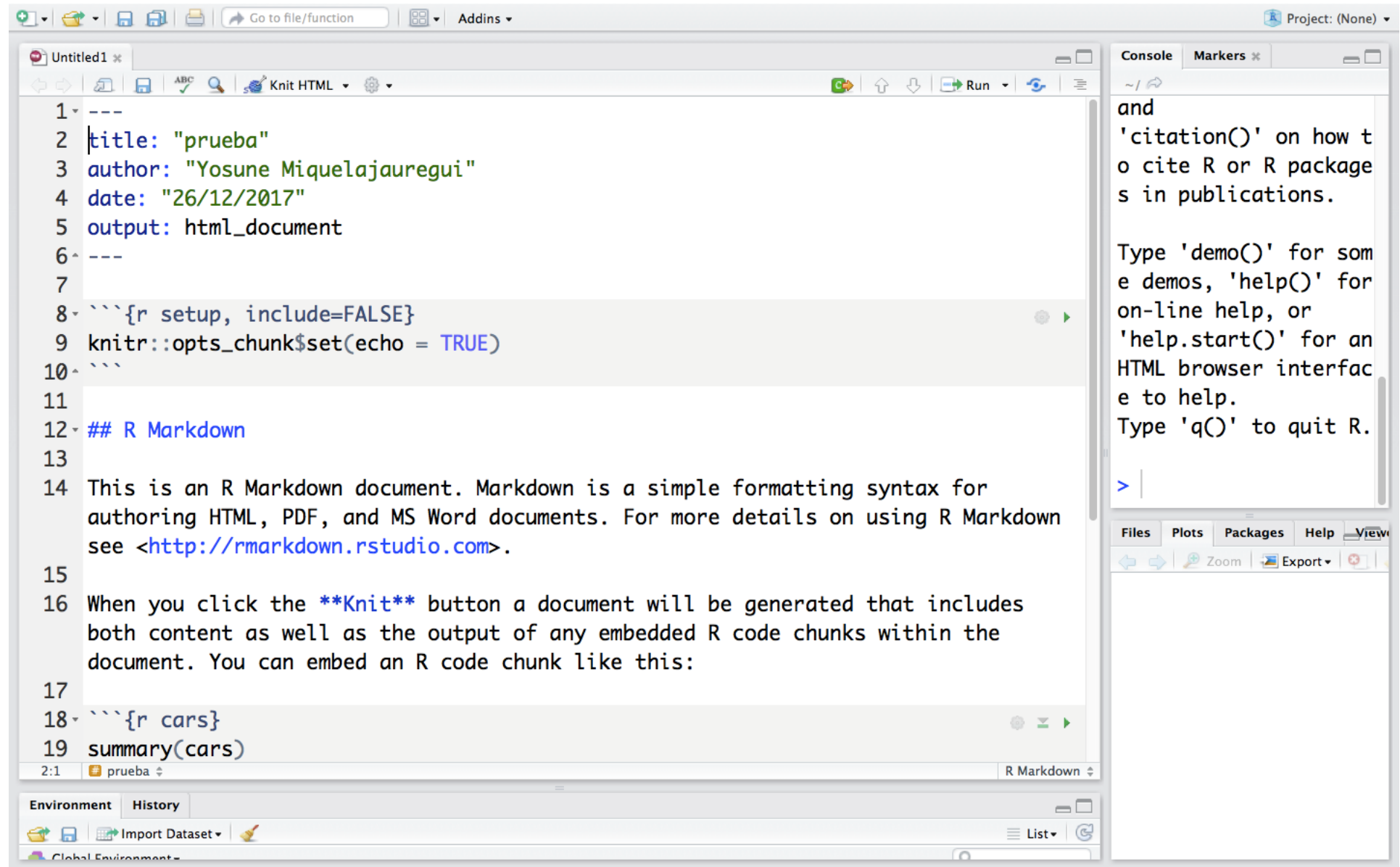
# Markdowns en R



# Markdowns en R



# Markdowns en R



The screenshot shows the RStudio IDE with an R Markdown document titled 'Untitled1'. The document content is as follows:

```
1 ---
2 title: "prueba"
3 author: "Yosune Miquelajauregui"
4 date: "26/12/2017"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for
15 authoring HTML, PDF, and MS Word documents. For more details on using R Markdown
16 see <http://rmarkdown.rstudio.com>.
17
18 When you click the Knit button a document will be generated that includes
19 both content as well as the output of any embedded R code chunks within the
20 document. You can embed an R code chunk like this:
21
22 ```{r cars}
23 summary(cars)
24 ```
```

The right sidebar shows the Console with the following text:

```
and
'citation()' on how to
cite R or R packages
in publications.

Type 'demo()' for some
demos, 'help()' for
on-line help, or
'help.start()' for an
HTML browser interface
to help.
Type 'q()' to quit R.

> |
```

The bottom of the RStudio window shows the Environment and History panels, and the status bar indicates the current line is 2:1 in the 'prueba' file.

# Markdowns en R

The screenshot displays the RStudio interface with an R Markdown script open in the editor. The script is titled 'Script3.Rmd' and contains the following code:

```
5 output: html_document
6 ---
7
8 # Script de prueba
9 ##Este script tiene el propósito de mostrar la manera de utilizar los markdowns
   para generar documentos html, pdf y word.
10
11 1. Importar datos
12 ```{r importar datos eval=T, echo=T}
13 carros <- mtcars[c(1,2,6,10)]
14 names(carros)<- c("eficiencia","cilindros","peso","velocidades")
15 str(carros)
16
17 ```
18 2. Hacer gráficas
19 ```{r importar datos eval=T, echo=T}
20 conteo <- table(carros$velocidades)
21 barplot(conteo,names.arg=c("3 vel","4 vel","5 vel"),xlab="Número de
   velocidades",ylab="Número de carros") # barras verticales
22 barplot(conteo, ylab="Número de velocidades",xlab="Número de carros",
   horiz=TRUE,names.arg=c("3 vel","4 vel","5 vel")) # barras horizontales
23 ```
24
```

The rendered HTML output is shown in the right-hand pane. It includes the title 'Script de prueba', a subtitle 'Este script tiene el propósito de mostrar la manera de utilizar los markdowns para generar documentos html, pdf y word.', and two sections: '1. Importar datos' and '2. Hacer gráficas'. The first section shows the R code for importing data from the 'mtcars' dataset. The second section shows the R code for creating two bar plots: one vertical and one horizontal, both showing the frequency of cars for different velocity categories (3, 4, and 5). The rendered HTML output is as follows:

and  
'citation()' on how to  
cite R or R package  
s in publications.


Type 'demo()' for some  
demos, 'help()' for on-  
line help, or  
'help.start()' for an  
HTML browser interface  
to help.  
Type 'q()' to quit R.

> |

Files Plots Packages Help View

R: Search Results Find in Topic

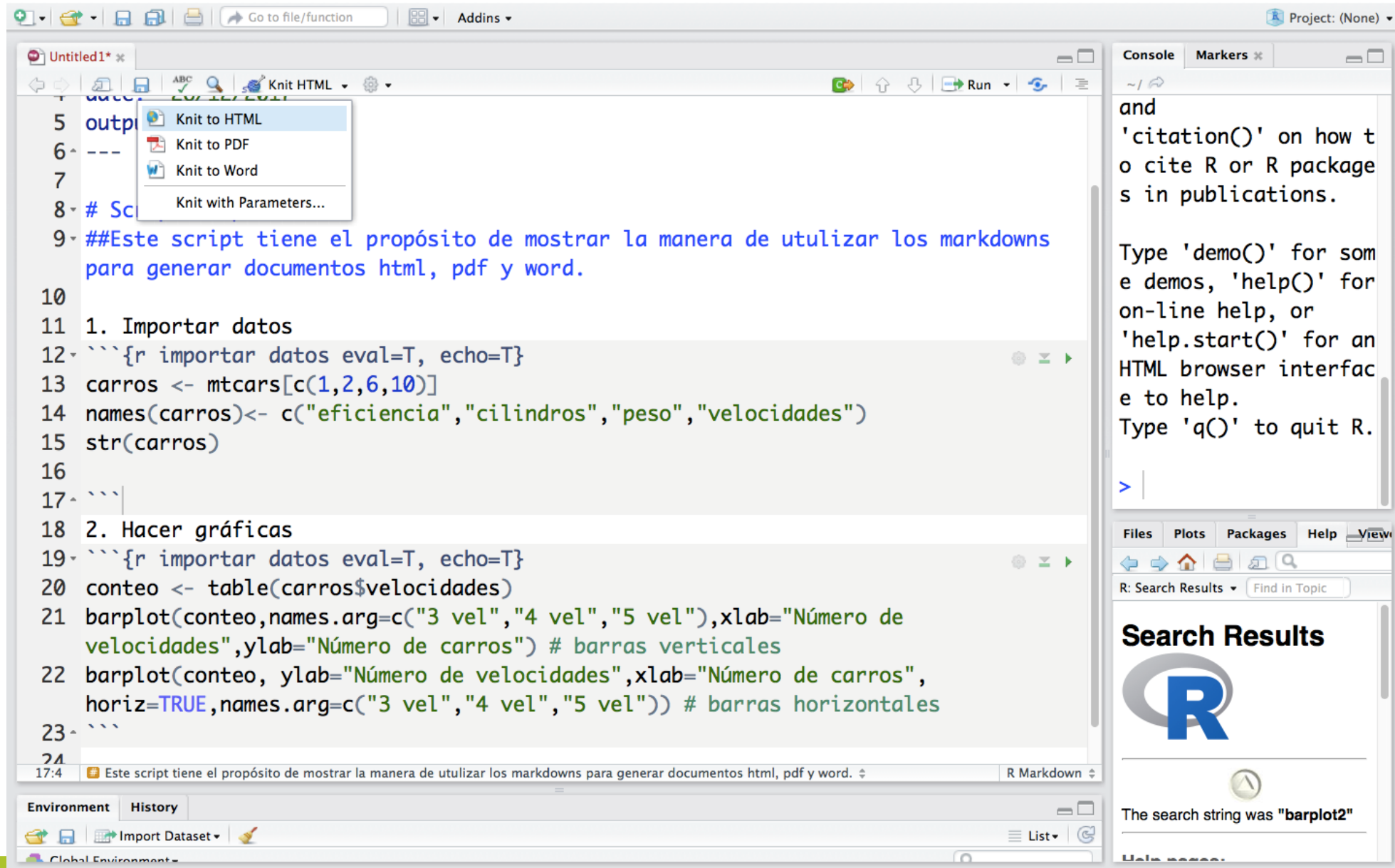
### Search Results



The search string was "barplot2"

Help pages

# Markdowns en R



The screenshot displays the RStudio interface. The main editor window shows a script with R code and Markdown comments. A context menu is open over the script, offering options to knit the document to HTML, PDF, Word, or with parameters. The right-hand pane is split into a Console and a Search Results panel. The Console shows the execution of R code, including data import and plotting. The Search Results panel displays the R logo and the search string "barplot2".

**Knit Menu Options:**

- Knit to HTML
- Knit to PDF
- Knit to Word
- Knit with Parameters...

**Script Content:**

```
5 output
6 ---
7
8 # Script
9 ##Este script tiene el propósito de mostrar la manera de utilizar los markdowns
   para generar documentos html, pdf y word.
10
11 1. Importar datos
12 ```{r importar datos eval=T, echo=T}
13 carros <- mtcars[c(1,2,6,10)]
14 names(carros)<- c("eficiencia","cilindros","peso","velocidades")
15 str(carros)
16
17 ```
18 2. Hacer gráficas
19 ```{r importar datos eval=T, echo=T}
20 conteo <- table(carros$velocidades)
21 barplot(conteo,names.arg=c("3 vel","4 vel","5 vel"),xlab="Número de
   velocidades",ylab="Número de carros") # barras verticales
22 barplot(conteo, ylab="Número de velocidades",xlab="Número de carros",
   horiz=TRUE,names.arg=c("3 vel","4 vel","5 vel")) # barras horizontales
23 ```
24
```

**Console Output:**

```
and
'citation()' on how to
o cite R or R package
s in publications.

Type 'demo()' for some
e demos, 'help()' for
on-line help, or
'help.start()' for an
HTML browser interfac
e to help.
Type 'q()' to quit R.


> |
```

**Search Results:**

Files Plots Packages Help View

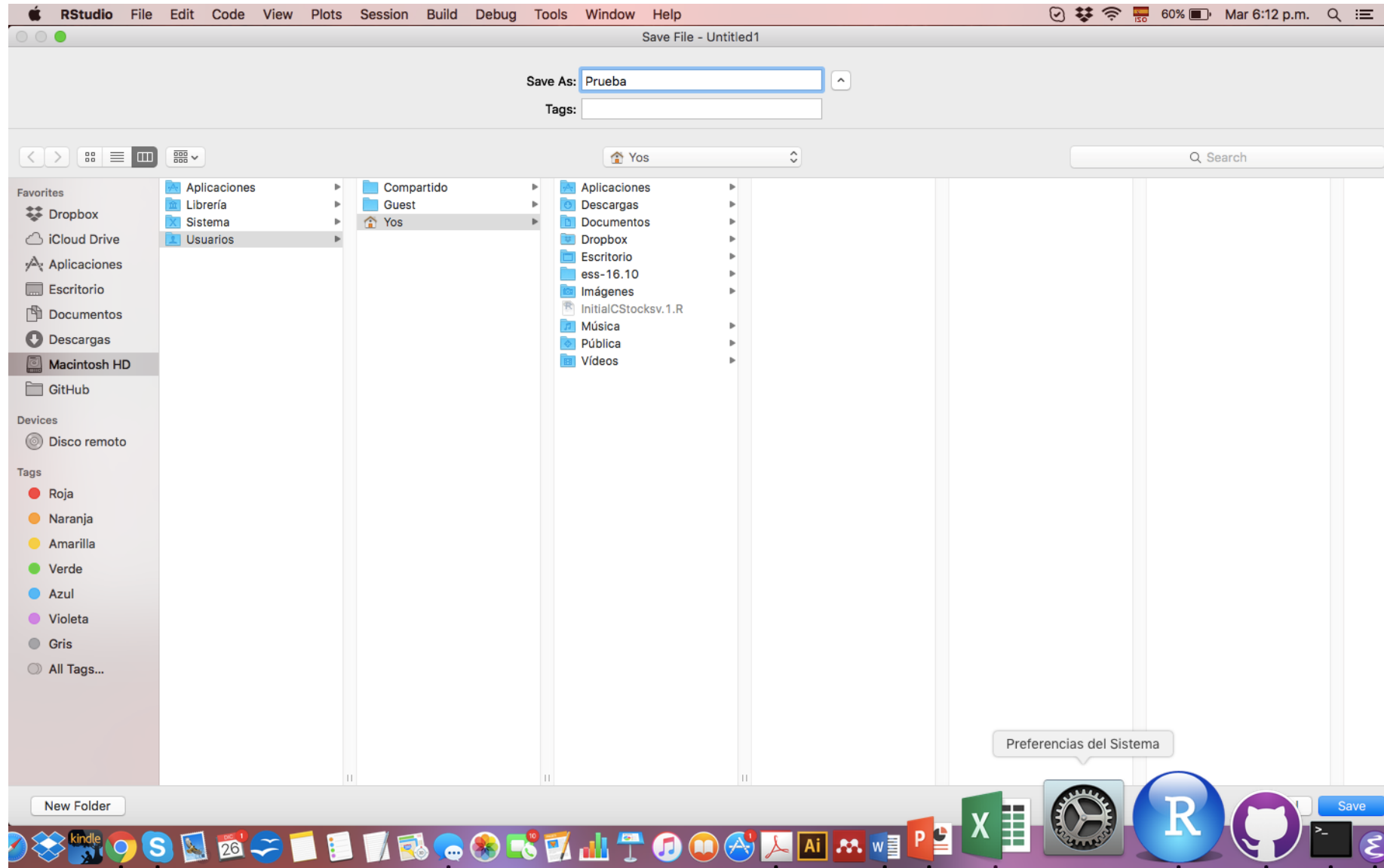
R: Search Results Find in Topic

**Search Results**



The search string was "barplot2"

# Markdowns en R



# Markdowns en R

file:///localhost/Users/Yos/Desktop/ClaseEstadisticaModelacion/Prueba.html



**prueba**  
Yosune Miquelajauregui  
26/12/2017

## Script de prueba

Este script tiene el propósito de mostrar la manera de utilizar los markdowns para generar documentos html, pdf y word.

1. Importar datos

```
carros <- mtcars[c(1,2,6,10)]
names(carros) <- c("eficiencia", "cilindros", "peso", "velocidades")
str(carros)
```

```
## 'data.frame':  32 obs. of  4 variables:
## $ eficiencia : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cilindros  : num  6 6 4 6 8 6 8 4 4 6 ...
## $ peso       : num  2.62 2.88 2.32 3.21 3.44 ...
## $ velocidades: num  4 4 4 3 3 3 3 4 4 4 ...
```

2. Hacer gráficas

```
conteo <- table(carros$velocidades)
barplot(conteo, names.arg=c("3 vel", "4 vel", "5 vel"), xlab="Número de velocidades", ylab="Número de carros") # barras verticales
```



Velocidad	Número de carros
3 vel	14
4 vel	12

# Introducción a la estadística

## ¿Para qué hacer estadística?

### 1. Estimar parámetros

*¿Cuál es la densidad del ocelote en el Área Natural Protegida, Reserva de la Biósfera Janos, en Chihuahua?*

### 2. Hacer pruebas de hipótesis

*¿Las diferencias de crecimiento entre los niños que viven en zonas urbanas y los que viven en zonas rurales se debe al azar o es resultado de un efecto del tipo de alimentación?*

### 3. Hacer inferencias

*Cerca del 20% de la población mexicana votará por Morena en las próximas elecciones del 2018.*



# Introducción a la estadística

## **Población y muestra**

La población estadística es el conjunto de elementos sobre el cual basamos nuestras conclusiones. Desconocido.

(hombres de 20 a 35 años en la UNAM)

Sin embargo no conocemos los parámetros que caracterizan a la población (e.x. la altura)

Estrategia 1 : Medir la altura de todos los hombres de la UNAM (poco práctico logísticamente)

Estrategia 2 : Utilizar una muestra de 50 hombres de 20 a 35 años seleccionados aleatoriamente.

# Introducción a la estadística

## Población y muestra

Se puede inferir sobre la población a partir de la muestra:

Si  $\bar{x} = 1.7$  m; podríamos decir que la media de la muestra  $\bar{x}$  es un estimador de la media de la población y que 1.7 es un estimado de ese valor (el estimador produce un estimado).

Característica de la población : PARÁMETRO (e.x. la media poblacional  $\mu$  )


Característica de la muestra : ESTIMADOR O ESTADÍSTICO (e.x. la media muestral  $\bar{x}$  )

# Estadística descriptiva

## Medidas de tendencia central


Distintos estadísticos permiten caracterizar una muestra y de estimar los parámetros de la población:

Medidas de tendencia central (posición):

1. Media - mean ()
  2. Mediana - median ()
  3. Moda - mode ()
- 

# Estadística descriptiva

## Medidas de dispersión

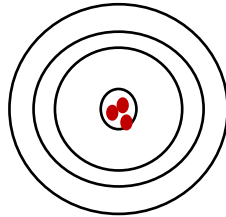
1. Varianza - `var ()`
  2. Desviación estándar - `sd ()`
  3. Rango - `range ()`
  4. Suma de cuadrados del error (SCE)
- 

# Estadística descriptiva

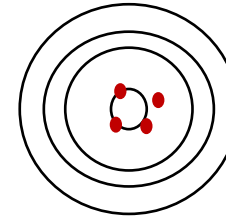
## Medidas de precisión

1. Error tipo o error estándar del estimador (SE) – una medida de la imprecisión de los valores estimados. El SE mide la variabilidad de las diferentes estimaciones, si el muestreo se repite un gran número de veces.

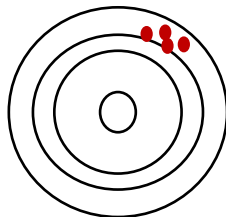
Preciso y exacto



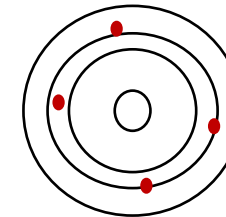
No preciso pero exacto



Preciso pero no exacto



Ni preciso ni exacto



# Estadística descriptiva

## Sesgo

1. Diferencia entre el valor esperado del estimador y el valor real del parámetro a estimar. Es deseable que un estimador sea insesgado.
2. El sesgo representa la tendencia de los estimadores de un parámetro a diferir sistemáticamente.

$$Sesgo = E(\hat{\theta}) - \theta$$

donde  $\hat{\theta}$  es el estimador de un parámetro,  $E(\hat{\theta})$  es el valor esperado del estimador del parámetro y  $\theta$  es el valor del parámetro.

# Estadística descriptiva

## Variables aleatorias

1. Una variable cuyos valores observados son resultado de un proceso aleatorio (experimento aleatorio).
2. Las variables aleatorias pueden ser discretas o continuas:

Discretas: Toma únicamente valores enteros:

- a) Binarias (e.x. presencia/ausencia, muerto/vivo)
- b) Categóricas y ordinales (e.x. pequeño, mediano, grande)
- c) Número de individuos (e.x. 0, 1, 23, 54)

Continuas: Toma un número infinito de valores dentro del intervalo dado (e.x. distancia, temperatura, largo).

# Estadística descriptiva

## Distribuciones estadísticas

La mayor parte de los análisis estadísticos dependen de una distribución estadística (análisis paramétricos tales como prueba T, ANOVA, regresión lineal y múltiple).

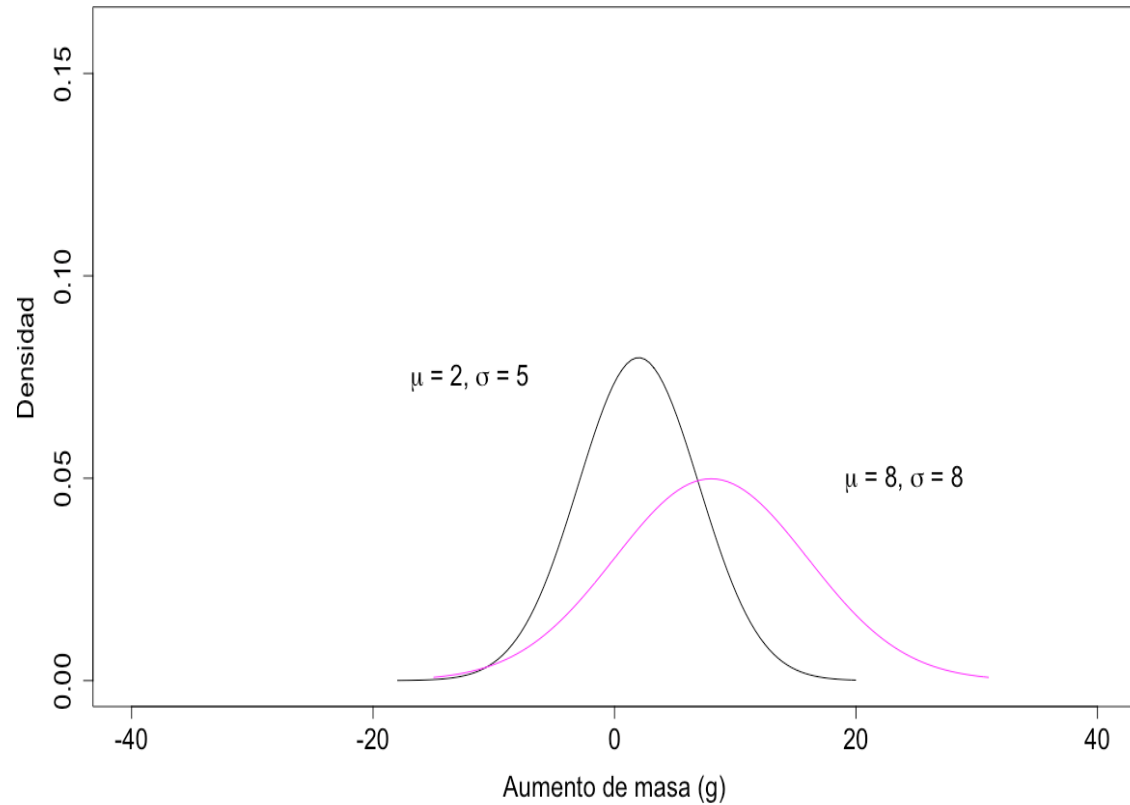
Los análisis paramétricos involucran una serie de supuestos asociados a los parámetros de la distribución.

Por ejemplo, la prueba de T para dos grupos independientes supone que las muestras son aleatorias y que provienen de poblaciones normales cuyas varianzas son iguales.

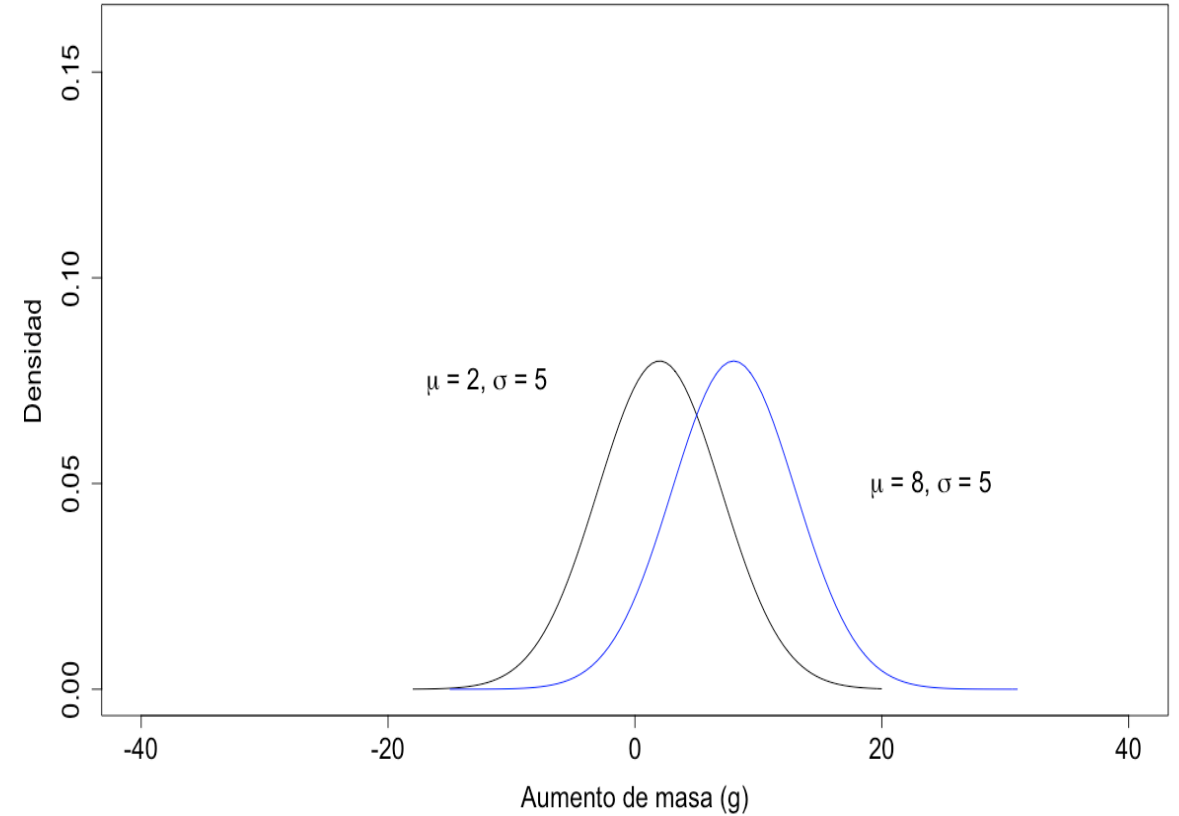


# Estadística descriptiva

Comparación entre dos grupos

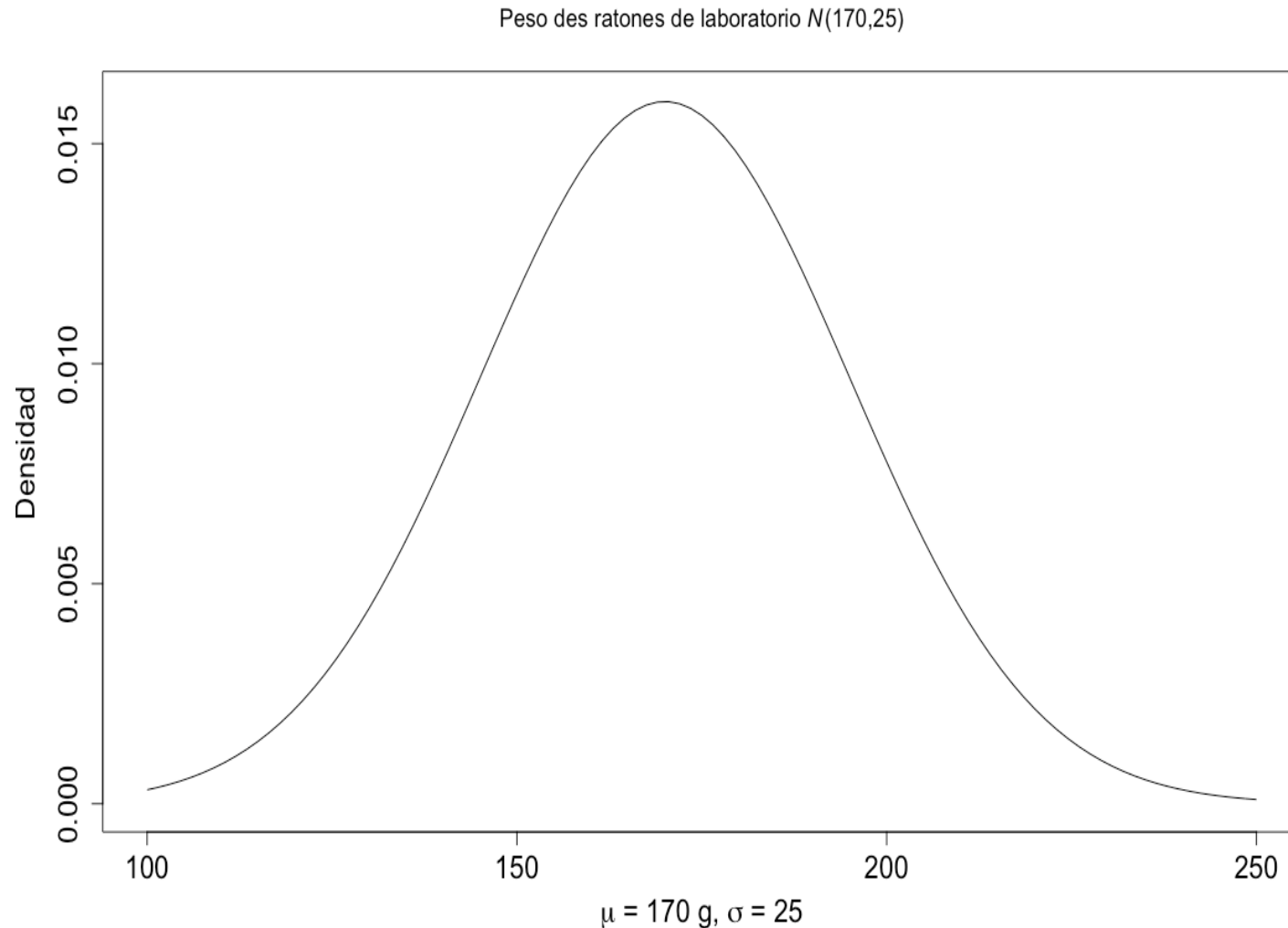


Comparación entre dos grupos



# Estadística descriptiva

## Distribución normal

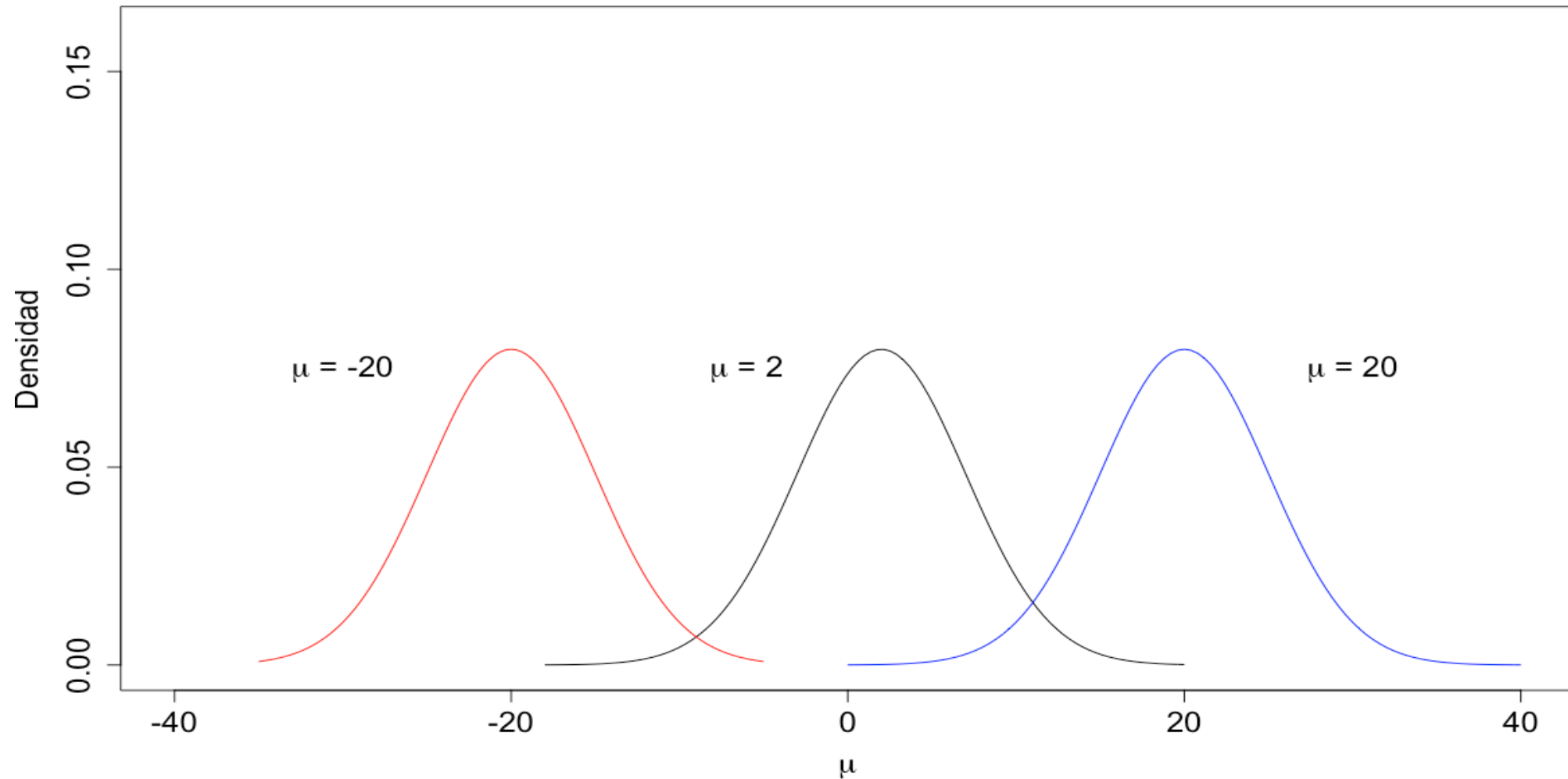


Características:

- Distribución continua
- La suma del área bajo la curva es 1
- Distribución simétrica
- 90% de las observaciones se encuentran a  $1.64\sigma$  de  $\mu$
- 95% de las observaciones se encuentran a  $1.96\sigma$  de  $\mu$
- 99% de las observaciones se encuentran a  $2.58\sigma$  de  $\mu$

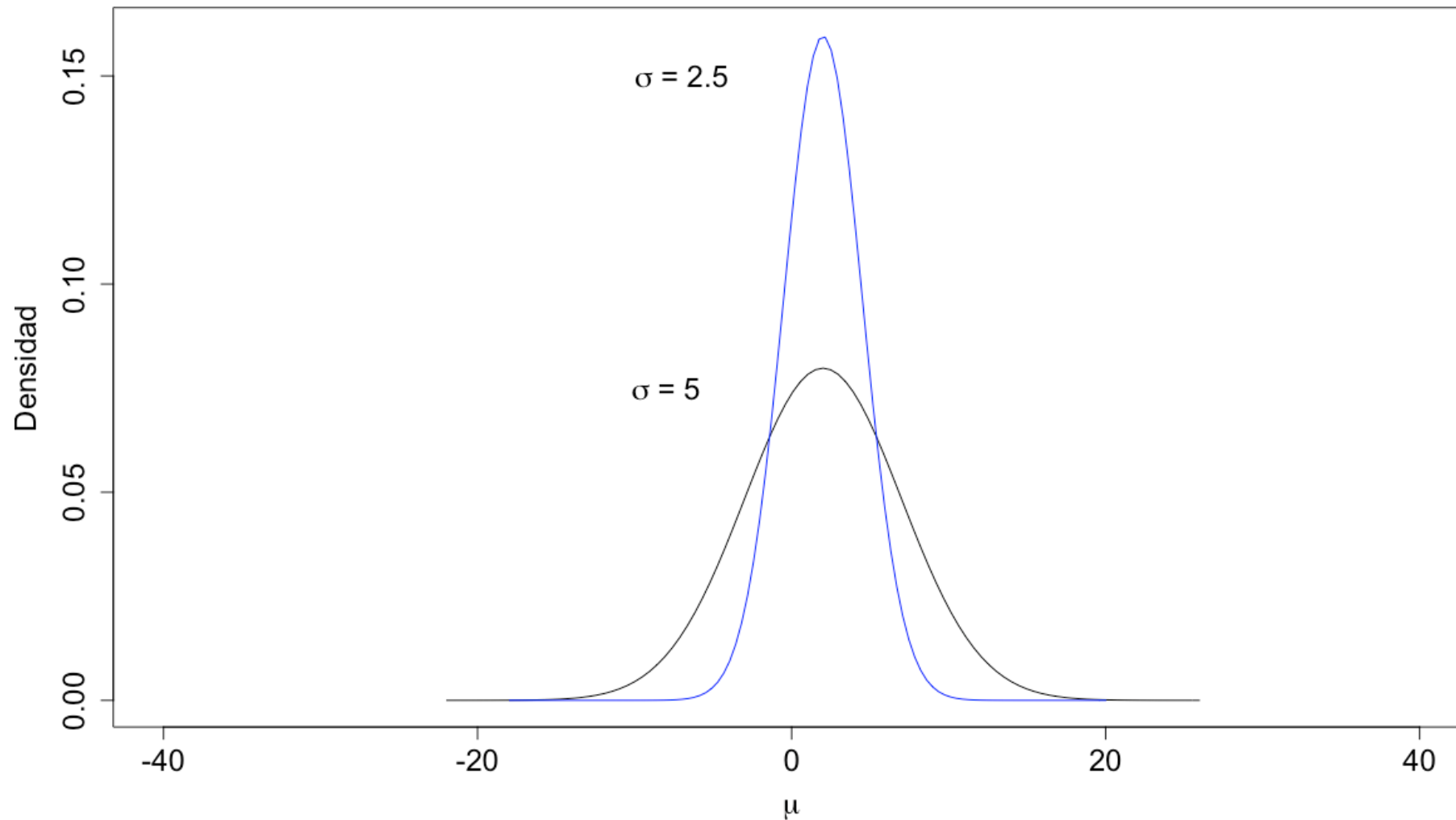
# Estadística descriptiva

**Distribución normal: la media determina la posición**



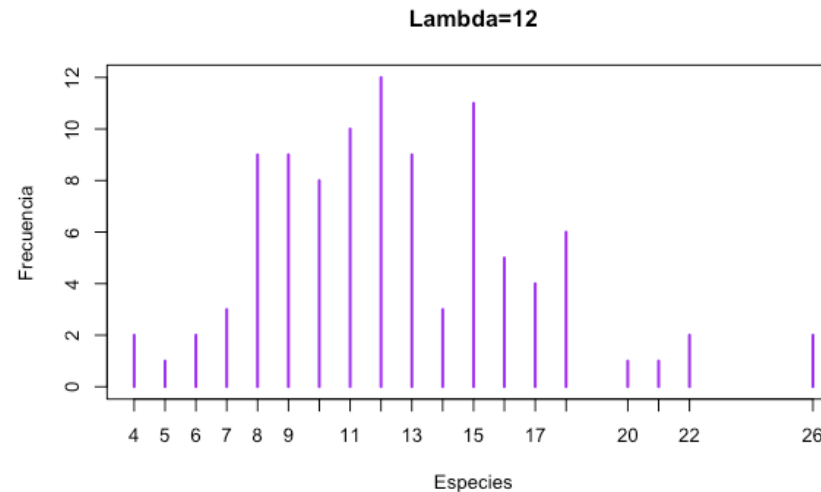
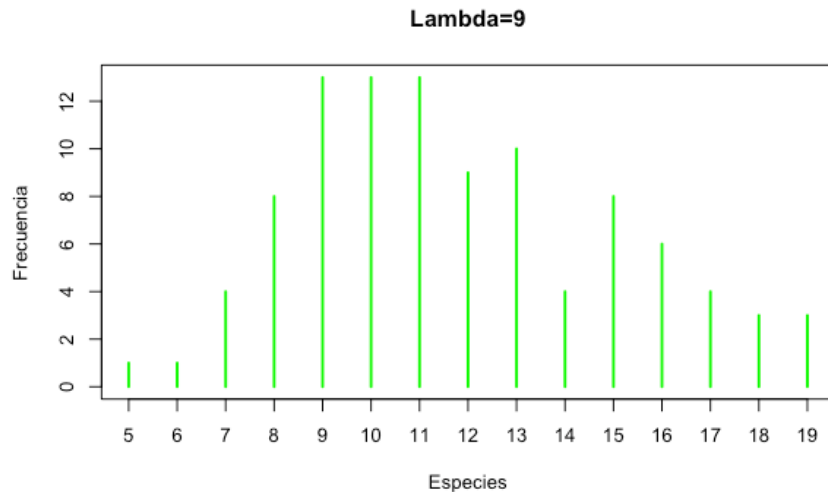
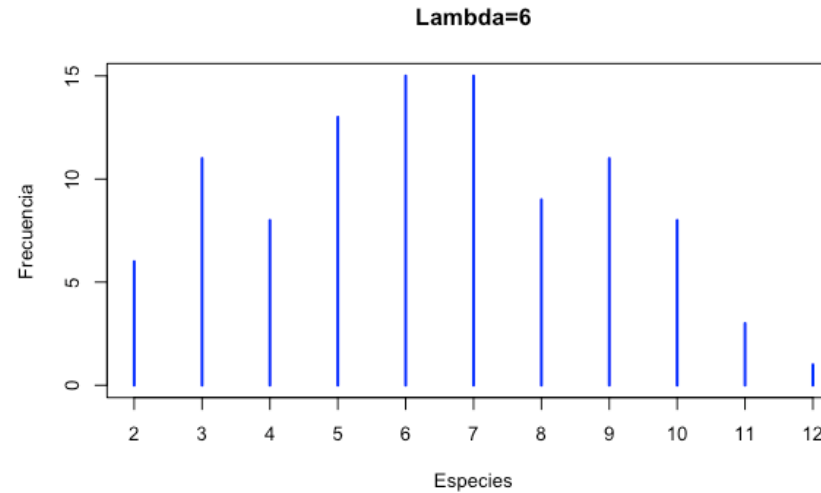
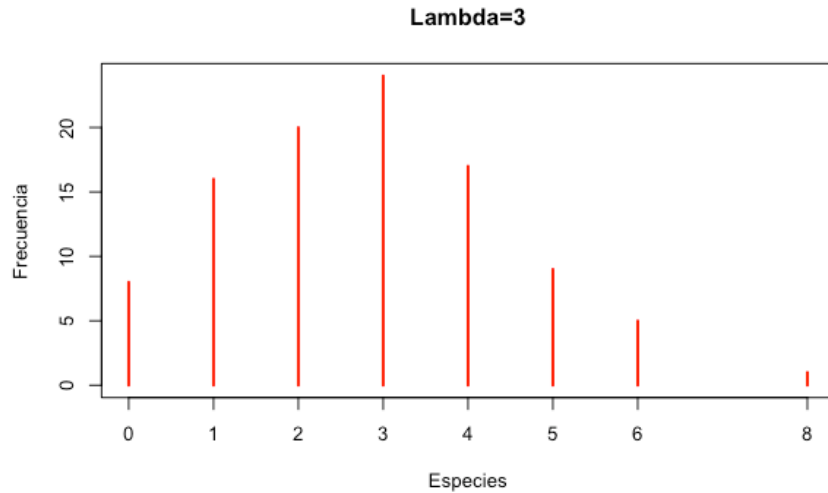
# Estadística descriptiva

**Distribución normal: la varianza determina la forma**



# Estadística descriptiva

## Distribución Poisson

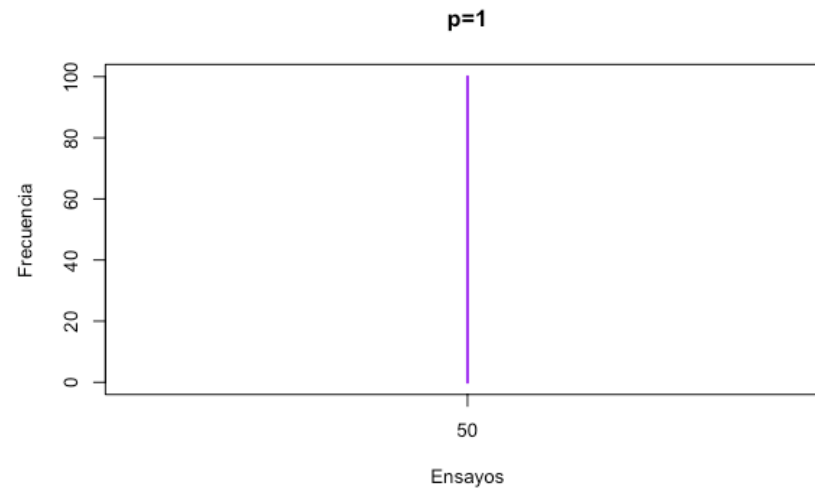
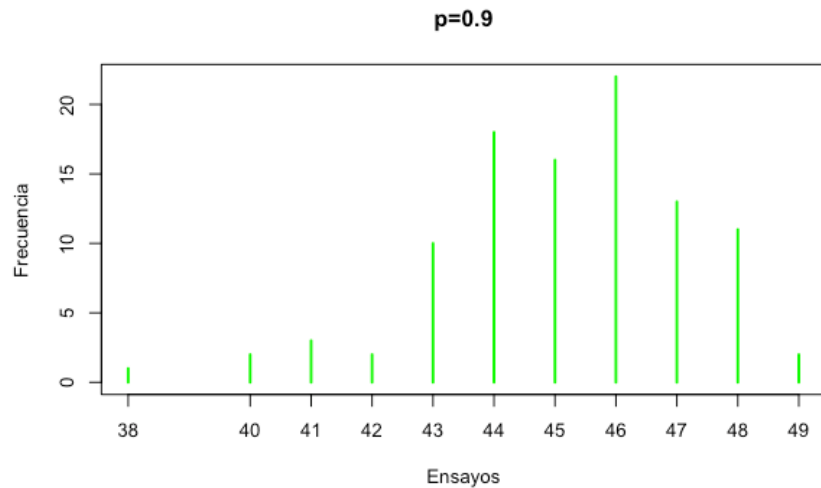
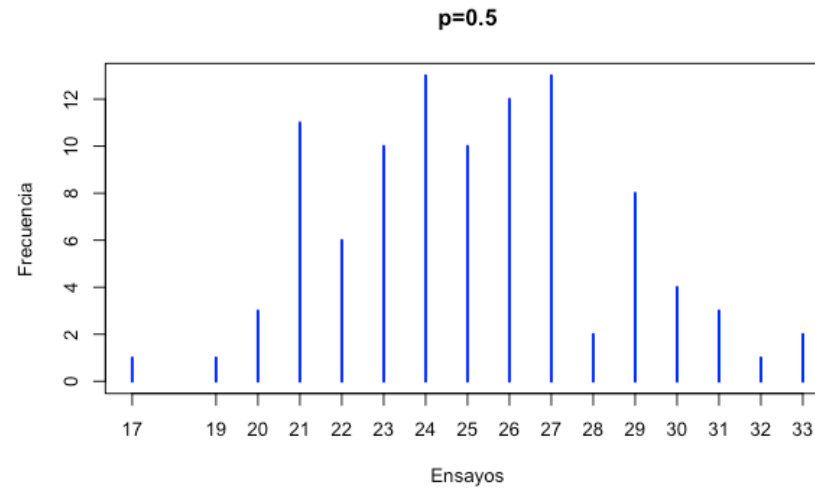
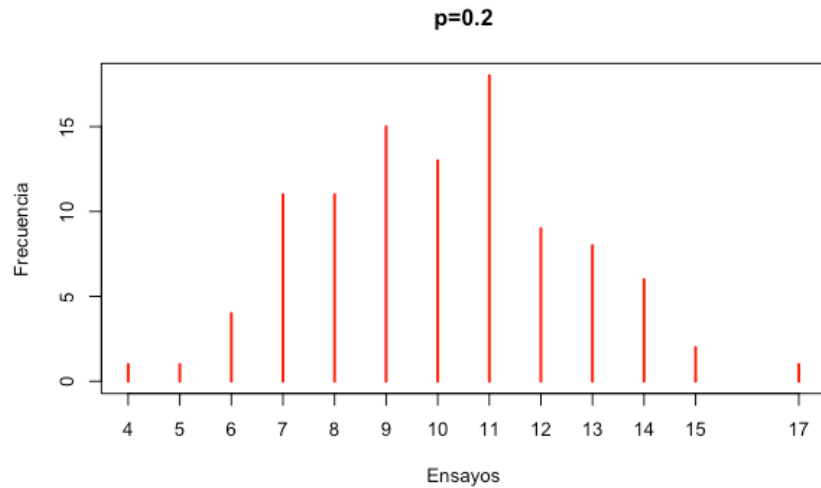


### Características:

- Se aplica a fenómenos discretos de la naturaleza
- Definida por el parámetro  $\lambda$ , que representa el número de veces que se espera que ocurra el fenómeno.

# Estadística descriptiva

## Distribución Binomial



### Características:

- Discreta que cuenta el número de éxitos  $x$  en una secuencia de  $n$  ensayos independientes entre sí, con una probabilidad fija  $p$  de ocurrencia del éxito entre los ensayos.

# Ejercicio

1. Generar un markdown en R para los ejercicios 1,2,3,4 y 5
2. Documentar apropiadamente cada paso
3. Generar un .pdf con las porciones de código evaluadas