Grupo MAT

ESI

UCLM

Sesión 1: Introducción a R

1

EUSEBIO ANGULO SÁNCHEZ-HERRERA

LABORATORIO DE ESTADÍSTICA



ESCUELA SUPERIOR
DE INFORMÁTICA
CIUDAD REAL

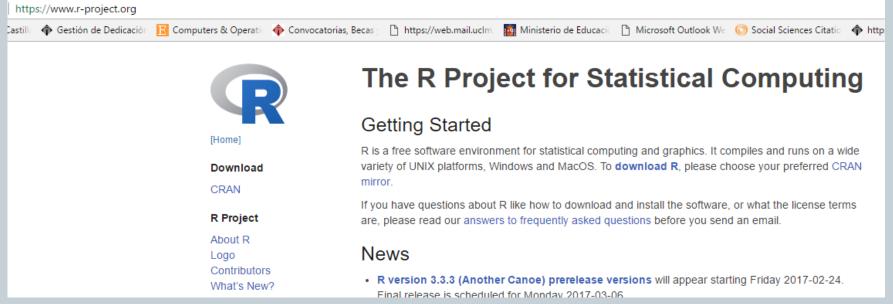




Instalación R







- R es software libre y está disponible para Windows, Linux y MacOS.
- Organizado en paquetes.
- Para cálculos estadísticos y desarrollo de gráficas.
- Recomendable instalarlo en portátil propio.



Instalación RStudio



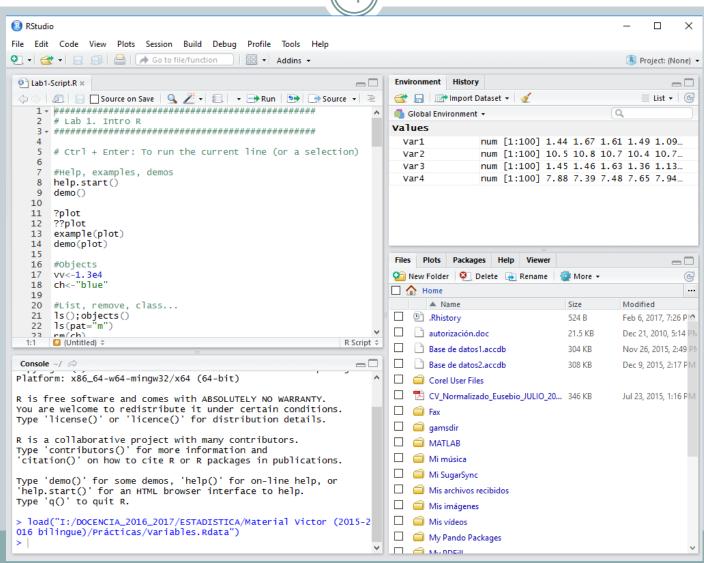






Entorno RStudio







Objetos: <-, =



5

• • •

• • •

Recomendación:

- <- para asignación
- = para especificar parámetros nombrados

https://renkun.me/2014/01/28/difference-between-assignment-operators-in-r/

- Nombres de objetos: Letras, Números, 'y'_'
- No pueden comenzar con un número



Asignación y Eliminación de Objetos



- El operador '<-' puede interpretarse como una abreviatura de:
 - > assign("x", c(10.4, 5.6, 3.1, 6.4, 21.7)).

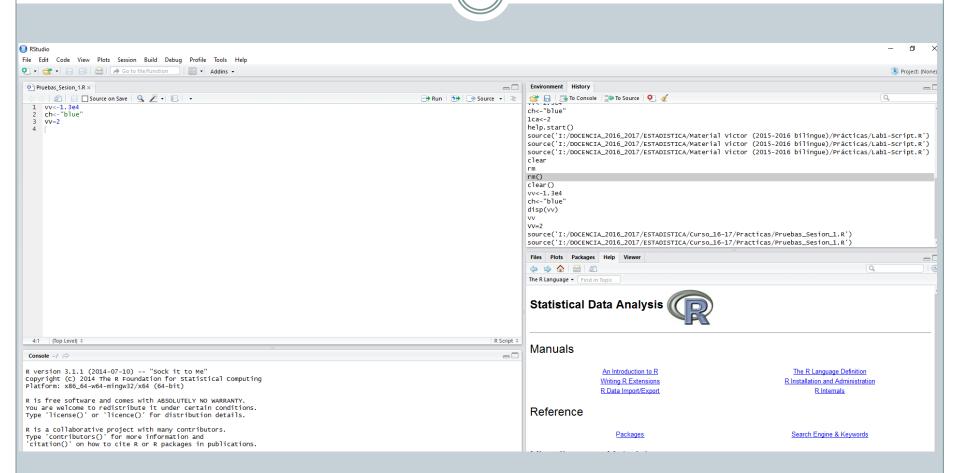
 Para eliminar objetos se puede utilizar la orden rm. Por ejemplo si queremos eliminar los dos objetos creados con anterioridad:

Rm(vv, ch)



Script/Console/History







Tipos de Objetos



• Vector:

- o Numeric 3, -24.5, 1.3e4...
- o Complex 3+2i
- o Character "blue" ...
- Logical TRUE, FALSE

• Data.frame: tablas de datos

```
> x <- data.frame(foo = 1:4, bar = c(T, T, F, F))
> x
    foo bar
1    1 TRUE
2    2 TRUE
3    3 FALSE
4    4 FALSE
> nrow(x)
[1] 4
> ncol(x)
[1] 2
```



Valores especiales



-
- NA (Valor perdido)
- NaN (No es un número o/o)
- NULL (Sin valor)
- Inf (Infinito)
- Sintáxis:
 - C≠c (distingue mayúsculas y minúsculas)
 - o # (para hacer comentarios en el código)
 - o; (ejecuta varios comandos en una línea)

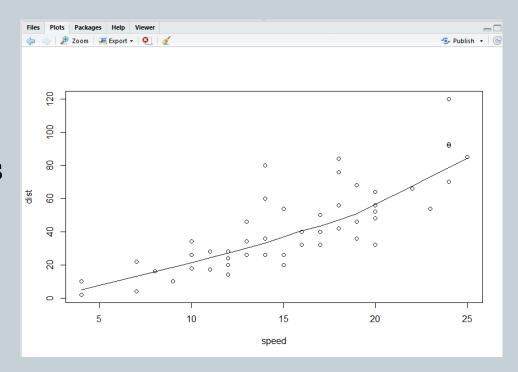


Errores y help()



10

- >sqrt 144
 - Error: unexpected numeric constant in "sqrt 144"
- >sqrt(144)[1] 12
- >?plot #general
- >??plot #lista funciones
- o >example(plot)
- o >demo(plot)





Operadores





- Relacionales: >, >=, <, <=, ==, !=
- Lógicos: !, &, |○
- >x<-144
- >x==12 (no es asignar)[1] FALSE
- >sqrt(x)==12[1] TRUE



Operadores



(12)

- R permite las operaciones aritméticas básicas: +, -, *, /
- Otros operadores son: ^, %%, %/%...
- Funciones matemáticas: abs(x), sqrt(x), floor(x), ceiling(x),...
 exp(x), log(x),..., sin(x), cos(x), tan(x),...

```
> help("%%")
> 128%%6
[1] 2
> 10%%6
[1] 4
> 10%/%6
[1] 1
> 128%/%6
[1] 21
> 2^10
[1] 1024
> |
```



Input, vectores



(13)

- o >incid<-c(123,89,142,118,109)
- >names(incid)<-c(1:5)</p>
- >names(incid)<-seq(1,5)</p>
- o >names(incid)<-c("Mon","Tue","Wed","Thu","Fri")

> incid Mon Tue Wed Thu Fri 123 89 142 118 109

#Número de incidentes tratados por un # equipo de trabajo



Funciones para objetos



14

>sum(incid); length(incid) #no considera nombres

Mean:
$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

o >sum(incid)/length(incid) or >mean(incid) #mean obvia NA

[1] 116.2

[1]116.2

Sample Variance: $var(x) = \frac{\sum_{i=1}^{n} (\bar{x} - x_i)^2}{n-1}$

- > >xbar<-mean(incid) or >var(incid)
- o >n<-length(incid) [1] 376.7
- >sum((xbar-incid)^2)/(n-1)
 [1] 376.7



Funciones para objetos



>min(incid); max(incid); range(incid)

[1] 89

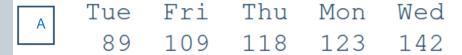
[1] 142

[1] 89 142 #no es la diferencia

• Ejercicio:

B > cumsum (incid)

Tue Wed Thu Fri
$$-34 53 -24 -9$$





Trabajar con objetos





o Incidentes que el equipo ha sido capaz de resolver

Monday	Tuesday	Wednesday	Thursday	Friday
65	54	67	59	53

- >resueltos<-c(65,54,67,59,53)</p>
- o ¿Cómo se recuperan los que se quedan sin resolver?
- >sin_resolver<-incid-resueltos</p>
- o ¿Qué porcentaje del total se ha resuelto de forma positiva?
- >porcentaje_resueltos<-(resueltos/incid)*100</p>



Acceder mediante indice



(17)

- o >incid[1]
- o >incid[length(incid)]
- o >incid[1:3]
- o >incid[c(1,3,5)]
- o >incid[-1]
- o >incid[-(1:3)]
- o >incid["Tue"]

- # inician en 1, accede al primer elemento
- # accede al último elemento
- # accede del 1 al 3 elemento
- # muestra elementos 1, 3 y 5
- # muestra todos menos el primero
- # muestra todos menos los 3
- primeros
- # muestra el del "Tue"



Ejercicio



TRUE...

18

```
• Ejercicio:
```

- A >incid>110 valores mayores que 110
- B >which(incid>110)
- C >incid[incid>110]

cuales mayor que 110 y el contenido

posiciones mayores que 110 se utiliza para la moda

- B Names and indices Mon Wed Thu
 - A TRUE/FALSE List Mon Tue Wed...

TRUE

FALSE

C Names and values Mon Wed Thu

- Sums...
- >sum(incid>110) ... sumatorio que lo cumplen[1] 3
- >sum(incid[incid>110]) ... sumatorio valores[1] 383



Data input: scan()





o Para añadir incidentes producidos en la próxima semana:

Monday	Tuesday	Wednesday	Thursday	Friday
143	178	98	121	101

o >incid2<-scan() # para meter datos doble intro para cerrar

```
1: 143 178 98 121 101
6:
Read 5 items
```

- >names(incid2)<-names(incid) #intercambiar nombres</p>
- >incidT<-c(incid,incid2) # combinar datos, pero misma fila</p>



Data input: scan()





>matrix1<-rbind(incid,incid2) # r de row (fila)</p>

>matrix2<-cbind(incid,incid2) # c de column (columna)</p>

```
> matrix2
incid incid2
Mon 123 143
Tue 89 178
Wed 142 98
Thu 118 121
Fri 109 101
```

- Un factor es un vector utilizado para especificar la clasificación discreta de los elementos de otro vector.
 - Ejemplo tenemos 20 personas de diferentes regiones de España, de tal modo que su región se especifica mediante un vector de caracteres con las abreviaturas de los mismos:
 - >regiones<- c("clm", "and", "mad", "cat", "mad", "cat", "mad", "clm", "and", "cat", clm", "and", "and", "cat", "mad", "cat", "mad", "cat", "and", "and", "cat", "mad", "cat", "mad", "cat", "and")
 - o Un factor se crearía:
 - >FactorRegiones <- factor(regiones)

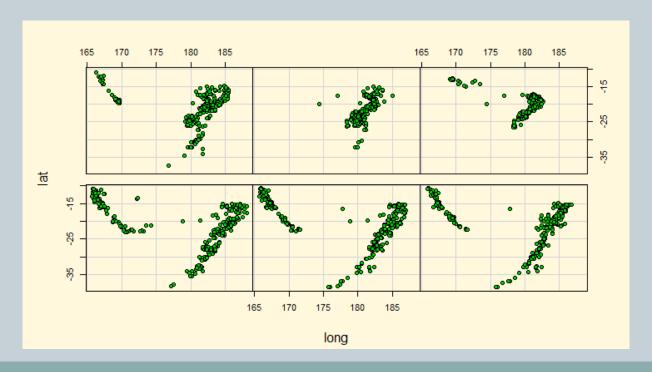
- o levels es para ver los niveles del factor:
 - > levels(FactorRegiones)
 - [1] "and" "cat" "clm" "mad"
- Otro vector se corresponde con los ingresos medios de las mismas personas al mes:
 - >ingresos<- c(800, 950, 1500, 1600, 1900, 2100, 1900, 950, 1000, 2100, 1950, 750, 800, 2500, 2100, 1950, 1900, 1800, 1100, 1400)
- Para calcular la media muestral para cada región podemos usar tapply():
- > MediaIngresos <- tapply(ingresos, FactorRegiones, mean)
 #devuelve el vector de medias con las componentes



demo(graphics)



- o Los gráficos disponibles en R son de gran calidad y de una versatilidad impresionante.
- Para hacernos una idea, podemos ejecutar la demo del programa mediante: >demo("graphics")

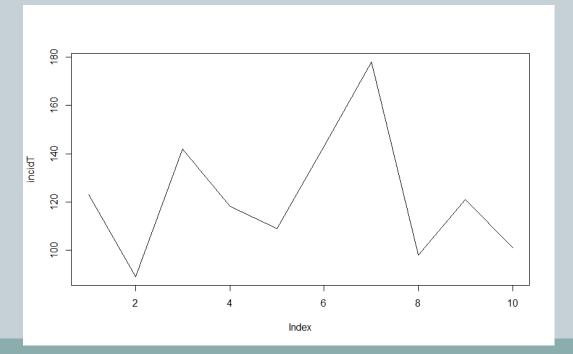




Comandos para plot()



- 24
- o Comandos para hacer una gráfica nueva.
 - >plot(); hist(); barplot(); boxplot();...
- o Para añadir elementos a un plot existente.
 - >text(); lines(); points();...

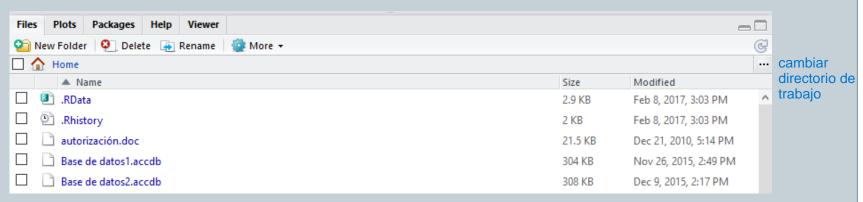




Saving and loading data



- o Directorio de trabajo
 - >getwd() # Devuelve el actual directorio de trabajo
 - >setwd("...") # Modifica el directorio de trabajo



- - >write.table(...)
 - >incid<-scan("incid.txt")</pre>
 - >read.table(...)



Saving and loading data



(26)

- R es compatible con todos los formatos de datos: .csv, .xls, .sav, sas...
- o .Rdata >save(incid,file="incid.Rdata")
 - >load("incid.Rdata")
- o .csv >write.csv(...)
 - >read.csv(...)



Packages



- R esta organizado en *packages* (módulos adicionales)
 - >install.packages("...") # lugar de descarga
 - >library(...)
 - >library()

- # todas las sesiones
- # paquetes disponibles en HDD

File	s Plots Packages Help View	rer		-0
0	Install Q Update		Q	G
	Name	Description	Version	
Sys	em Library			^
	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18	⊗
	class	Functions for Classification	7.3-14	⊗
	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.5	8
	codetools	Code Analysis Tools for R	0.2-15	8
	compiler	The R Compiler Package	3.3.2	8
	datasets	The R Datasets Package	3.3.2	8
	foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase,	0.8-67	8
	graphics	The R Graphics Package	3.3.2	8
	grDevices	The R Graphics Devices and Support for Colours and Fonts	3.3.2	8
	grid	The Grid Graphics Package	3.3.2	8
	KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-15	8
	lattice	Trellis Graphics for R	0.20-34	8
	MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-45	8
	Matrix	Sparse and Dense Matrix Classes and Methods	1.2-7.1	8
	methods	Formal Methods and Classes	3.3.2	8
	mgcv	Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation	1.8-15	8
	nlme	Linear and Nonlinear Mixed Effects Models	3.1-128	8
	nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models	7.3-12	⊗
	parallel	Support for Parallel computation in R	3.3.2	⊗
	rpart	Recursive Partitioning and Regression Trees	4.1-10	⊗
	spatial	Functions for Kriging and Point Pattern Analysis	7.3-11	8
	splines	Regression Spline Functions and Classes	3.3.2	⊗
	stats	The R Stats Package	3.3.2	⊗ ∨



Ejercicio



28)

o Los siguientes datos contienen los resultados de una encuesta en la que la pregunta era "¿Eres un fumador?"

- × 1) Obtenga las frecuencias relativas para cada respuesta.
- × 2) Representar gráficos de barras y sectores mostrando las frecuencias de cada respuesta.

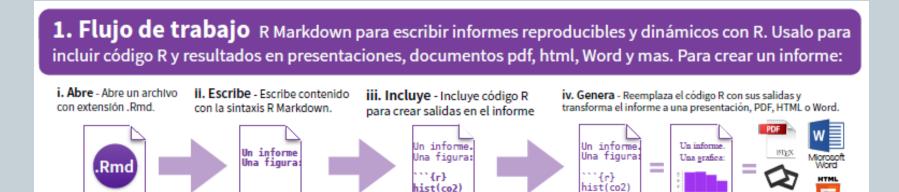






https://www.rstudio.com/wpcontent/uploads/2015/02/rmarkdown-cheatsheet.pdf

http://rmarkdown.rstudio.com



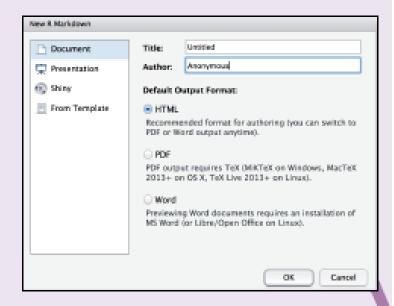






2. Abre archivo Empieza guardando un archivo de text con la extensión .Rmd o abre RStudio.

- En el menu haz clic en:
 File ➤ New File ➤ R Markdown...
- Una ventana abre donde puedes seleccionar el tipo de formato que quieres tener como resultado de tu archivo .Rmd.
- Selecciona el tipo de formato que quieres como salida con los botones (puedes cambiarlo después)
- Haz clic en OK









3. Markdown Ahora, escribe tu informe en texto plano. usa la sintaxis de markdown para describir el formato de tu texto en el informe final.

sintaxis

```
Texto plano
Termina linea con dos espacios para nuevo párrafo.
*cursivo* y _cursivo_
**negrita** y __negrita__
superindice^2^
~~tachado~~
[eslabón](www.rstudio.com)
# Encabezado 1
## Encabezado 2
### Encabezado 3
#### Encabezado 4
##### Encabezado 5
###### Encabezado 6
raya em: --
raya em: ---
elipsis: ...
ecuación en línea: $A = \pi*r^{2}$
imagen: ![](camino/a/smallorb.png)
regla horizontal (o nueva diapositiva):
***
```

resulta en

Texto plano

Termina linea con dos espacios para nuevo párrafo.

cursivo y cursivo

negrita y negrita

superindice2

tachado

eslabón

Encabezado 1

Encabezado 2

Encabezado 3

Encabezado 4

Encabezado 5

Encabezado 6

raya em: -

raya em: -

elipsis: ...

ecuación en línea: $A = \pi * r^2$

imagen:



regla horizontal (o nueva diapositiva):





32

3. Markdown Ahora, escribe tu informe en texto plano. usa la sintaxis de markdown para describir el formato de tu texto en el informe final.

sintaxis

- > cita en bloque
- * lista sin orden
- * elemento 2
 - + sub-elemento 1
 - + sub-elemento 2
- 1. lista ordenada
- 2. elemento 2
 - + sub-elemento 1
 - + sub-elemento 2

Encabezado Tabla	Segundo Encabezado
Celda de tabla	Celda 2
Celda 3	Celda 4

resulta en

cita en bloque

- · lista sin orden
- elemento 2
 - sub-elemento 1
 - sub-elemento 2
- lista ordenada
- elemento 2
 - sub-elemento 1
 - sub-elemento 2

Encabezado Tabla	Segundo Encabezado	
Celda de tabla	Celda 2	
Celda 3	Celda 4	





33

4. Escoge salida Escribe un encabezado en YAML que explica que tipo de documento quieres construir con tu archivo R Markdown.

YAML

Un encabezado YAML es un conjunto de pares llave:valor al comienzo de tu archivo. Comienza y termina el encabezado con tres rayas (---) title: "Sin título" author: "Anónimo" output: html_document

Este es el comienzo de mi informe. Arriba están los metadatos en un encabezado YAML El generador de planillas de RStudio te genera el encabezado YAML

El valor de salido determina que tipo de archivo R construirá con base en tu archivo .RmD (en Paso 6)

output: html_document ----- archivo html (página web)

output: pdf_document ----- archivo pdf

output: word_document - - - - Microsoft Word .docx

output: beamer_presentation----- presentación beamer (pdf)

output: ioslides_presentation----- presentación ioslides (html)

















5. Incluye código Usa sintaxis de knitr para incluir código R en tu informe.
R correrá el código e incluirá los resultados cuando generas el documento.

código incrustado

Surround code with back ticks and r. R replaces inline code with its results.

Dos mas dos equivale `r 2 + 2`.

Dos mas dos equivale 4.

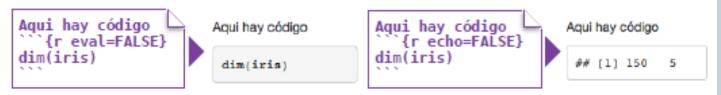
pedazos de código

comienza un trozo (chunk) con ```{r}.
Termina un trozo con ```



opciones para mostrar

Usa las opciones de *knitr* para cambiar el formato de un trozo. Pon las opciones entre llaves encima del trozo correspondiente.









5. Incluye código Usa sintaxis de knitr para incluir código R en tu informe.
R correrá el código e incluirá los resultados cuando generas el documento.

opción	defecto	efecto
eval	TRUE	Indica si se va a evaluar el código e incluir los resultados
echo	TRUE	Indica si se muestra el código a la par de los resultados
warning	TRUE	Indica si se muestran advertencias
error	FALSE	Indica si se muestran errores
message	TRUE	Indica si se muestran mensajes
tidy	FALSE	Indica si se muestra código de forma organizada
results	"markup"	Opciones: "markup", "asis", "hold", o "hide"
cache	FALSE	Indica si se guardan resultados en cache
comment	"##"	Carácter de comentario para anteponer a resultados
fig.width	7	Ancho en pulgadas para figuras generadas en el trozo
fig.height	7	Alto en pulgadas para figuras generadas en el trozo







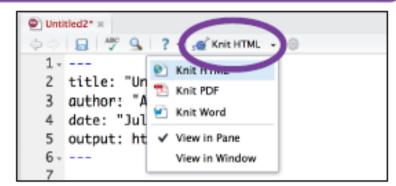
6. Genera usa to archivo .Rmd como plantilla para generar un informe terminado.

Genera tu informe en dos maneras

- Corre rmarkdown::render("<camino/a/ archivo>")
- Haz clic en el botón knit HTML en la parte de arriba de la ventana de RStudio scripts

Cuando generas un informe, R

- ejecutará cada trozo de código incrustado en el documento e incluirá los resultados
- construirá una nueva version de tu informe en el formato que haz indicado
- abre una prevista del archivo de salida en la ventana viewer
- guarda el archivo de salida en tu carpeta de trabajo









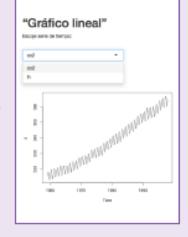
7. Documentos interactivos Convierte tu informe en un documento interactivo Shiny en 3 pasos

Añade runtime: shiny a encabezado YAML

title: "Gráfico lineal"
output: html_document
runtime: shiny

En los trozos de código, añade funciones de Shiny **input** para incrustar widgets. Añade funciones Shiny **render** para salidas reactivas

Render with rmarkdown::run or click Run Document in RStudio



* Nota: tu informe será un app de Shiny, esto significa que tienes que seleccionar un formato de salida html, como **html_document** (para informe interactivo) o **ioslides_presentation** (para presentación interactiva).



Sumario





- o www.r-project.org
- o www.rstudio.com
- Objetos y sintaxis en R
- Manejar entrada y salida de datos en R
- o Representación de gráficas e imágenes
- Paquetes
- o R Mark Down