

Módulo 5: Documentación

RMarkdown y knitr

Leticia Debera, Darío Padula y Daniel Alessandrini

Intendencia de Montevideo

2023/12/07

Generación de reportes automáticos utilizando `Rmarkdown` + `knitr`

Repaso: visto hasta ahora

- Introducción / conceptos básicos: aprendimos
 - Qué es `R`, interfaz gráfica `RStudio`
 - Elementos: objetos, estructura, accesibilidad, operadores
 - Objetos: `vector`es, `data.frame`, `listas`,...
- Manipulación de datos:
 - Importación y Exportación de datos
 - tuberías o *pipes* `%>%`
 - `dplyr`: verbos (`mutate`, `filter`, `group_by`, `summarise`...), funciones (`mutate_at/if`, `summarise_at/if`, `across`...)
 - `tidyr`: `joins` para juntar datos, `pivot` para cambiar formas...
- Visualización: `ggplot` *grammar of graphics* (gráfico en varias etapas)

¿Qué veremos hoy?: reportes con RMarkdown y knitr

¿Qué es RMarkdown? Origen en *Programación Literaria* (D.Knuth 1984): texto ordenado como quiere quien lo escribe (y no la máquina), buscando mezclar narrativa y código informático

RMarkdown: Markdown enriquecido, pensado para documentos técnicos (tablas, notas al pie, ecuaciones, referencias bibliográficas, citas, etc.)

- Markdown

- Lenguaje de marcado: se agregan "marcas" (o etiquetas) al texto para enriquecerlo
- Uso muy extendido: ej. WhatsApp

- knitr

- Paquete de R que lo vincula con Markdown
- Transforma código R en lenguajes de documentación, permitiendo reproducibilidad

- Pandoc

- Pandoc: conversor de documentos Markdown a otros formatos (PDF, HTML, DOC,)
- Mirar ventana Render...

Rmd >>knitr>> Markdown >>Pandoc>> compilación: HTML, PDF, DOC, etc

Ejemplo de principio a fin

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R Script Ctrl+Shift+N

Quarto Document...

Quarto Presentation...

R Notebook

R Markdown...

Shiny Web App...

Plumber API...

Text File

C++ File

Python Script

SQL Script

Stan File

D3 Script

R Sweave

R HTML

R Documentation...

datP x Soluc_Ejerc_Mod3_C03.Rmd x 0_carglimp.R x 9_clase.Rmd x

Save ABC Knit

Next Prev All Replace Replace All

Whole word Regex ☒ Wrap

preparamos todo pa trabajar -->
 chivos en Moodle, instrucciones CLARAS -->
 : 1 {eval = F}, 2 {eval = T,warning=F,message=FALSE,echo =
 de'} ADENTRO VA LO MISMO!! -->

to hasta ahora

cción / conceptos básicos: aprendimos a xxxx
 acción de datos:
 verbos ('mutate', 'filter', 'group_by', 'summarise...'), funciones

R Markdown

Background Jobs

/RStudio/Clases/M5-Reportes/

R is free software and comes with ABSOLUTELY NO WARRANTY.
 You are welcome to redistribute it under certain conditions.
 Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
 Type 'contributors()' for more information 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.

> |

Environ

R

Files

Install

System L

abir

asky

bac

bas

bas

bit

bit6

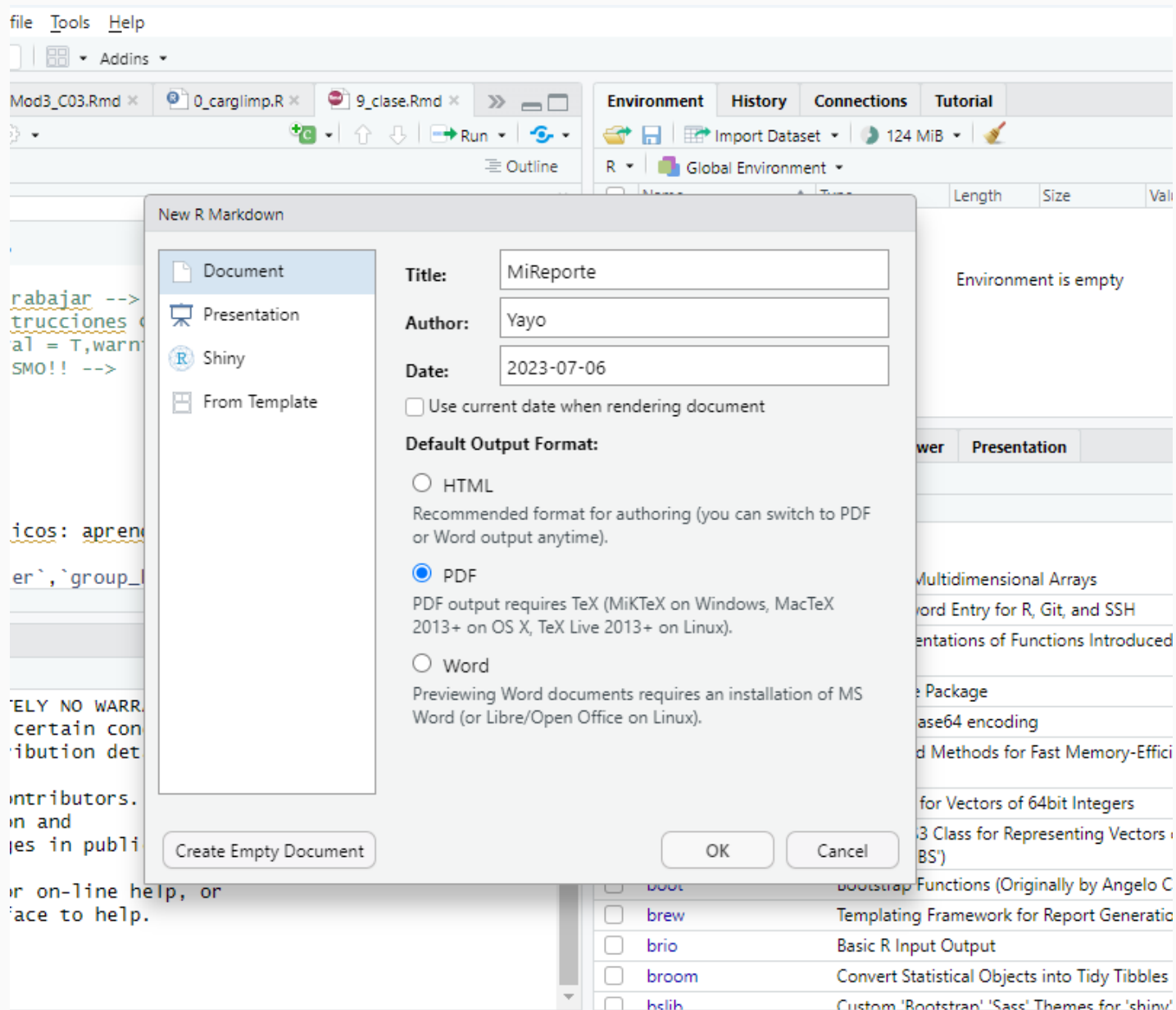
blot

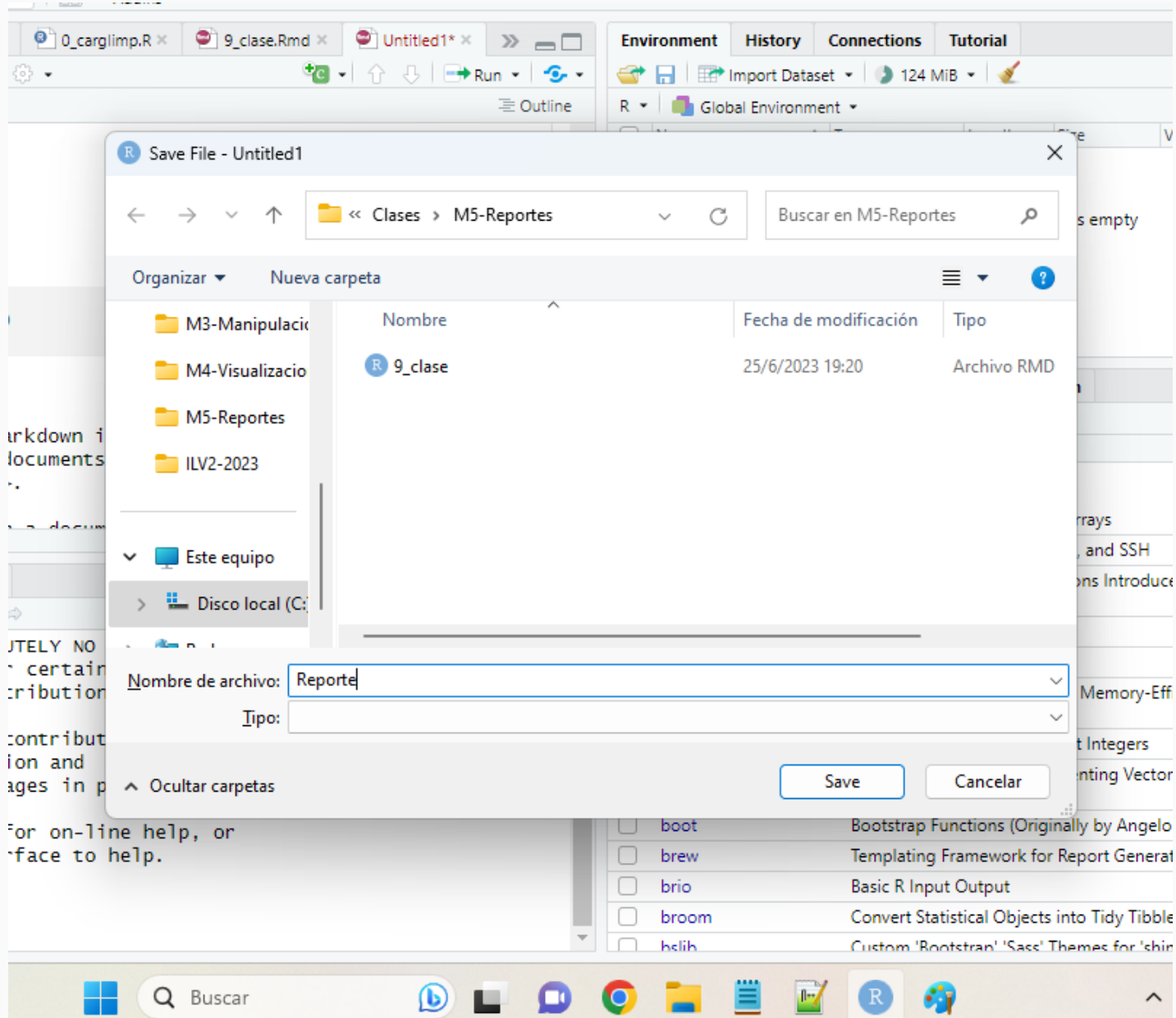
boo

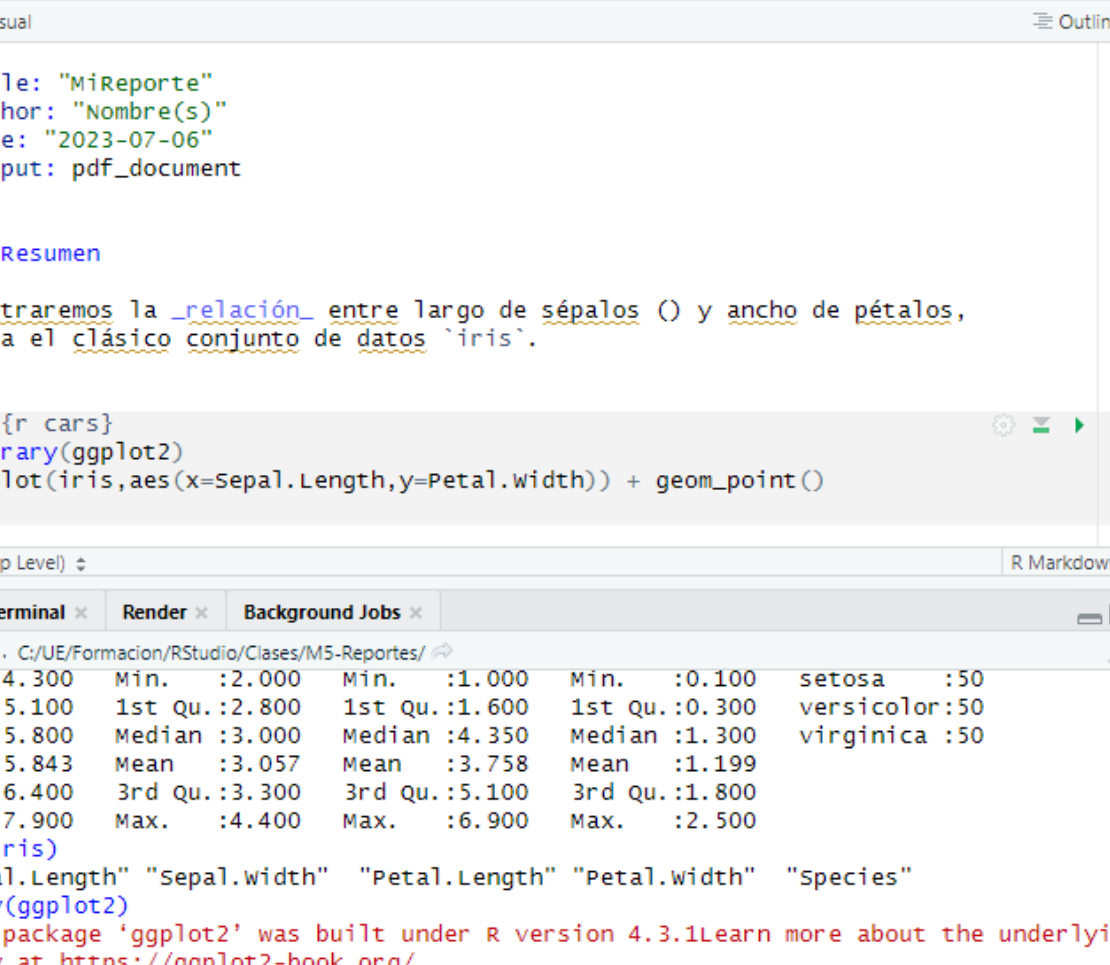
bre

bric

bro







The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for a report and a plot.


```

1 ---
2 title: "MiReporte"
3 author: "Nombre(s)"
4 date: "2023-07-06"
5 output: pdf_document
6 ---
7 |
8 ## Resumen
9
10 Mostraremos la relación entre largo de sépalos () y ancho de pétalos,
11 para el clásico conjunto de datos `iris`.
12
13
14 ```{r cars}
15 library(ggplot2)
16 ggplot(iris,aes(x=Sepal.Length,y=Petal.width)) + geom_point()
17
18
      
```
- Console:** Displays the output of the R code.


```

R 4.3.0 - C:/UE/Formacion/RStudio/Clases/M5-Reportes/
Min.      :4.300   Min.      :2.000   Min.      :1.000   Min.      :0.100   setosa     :50
1st Qu.   :5.100   1st Qu.   :2.800   1st Qu.   :1.600   1st Qu.   :0.300   versicolor:50
Median    :5.800   Median    :3.000   Median    :4.350   Median    :1.300   virginica  :50
Mean      :5.843   Mean      :3.057   Mean      :3.758   Mean      :1.199
3rd Qu.   :6.400   3rd Qu.   :3.300   3rd Qu.   :5.100   3rd Qu.   :1.800
Max.      :7.900   Max.      :4.400   Max.      :6.900   Max.      :2.500

> names(iris)
[1] "Sepal.Length" "Sepal.width"  "Petal.Length" "Petal.width"  "species"
> library(ggplot2)
warning: package 'ggplot2' was built under R version 4.3.1 Learn more about the underlying theory at https://ggplot2-book.org/
> ggplot(iris,aes(x=Sepal.Length,y=Petal.width)) + geom_point()
>
      
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

03.Rmd datP Soluc_Ejerc_Mod3_C03.Rmd 0_carglimp.R 9_clase.Rmd Reporte.Rmd

Knit on Save Knit Run

Source Visual

- Knit to HTML
- Knit to PDF
- Knit to Word
- Knit with Parameters...
- Knit Directory
- Clear Knitr Cache...

```

1 ---
2 title: "Mi Reporte"
3 author: "Nombre"
4 date: "2023-01-01"
5 output: pdf_document
6 ---
7
8 ## Resumen
9
10 Mostraremos la relación entre el largo de sépalos () y ancho de pétalos,
11 para el clásico conjunto de datos 'iris'.
12
13
14 ```{r cars}
15 library(ggplot2)
16 ggplot(iris,aes(x=Sepal.Length,y=Petal.width)) + geom_point()
17 ```
18

```

7:1 (Top Level) R Markdown

Console Terminal Render Background Jobs

R 4.3.0 - C:/UE/Formacion/RStudio/Clases/M5-Reportes/

```

Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50
1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50
Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50
Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199
3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
> names(iris)
[1] "Sepal.Length" "Sepal.width" "Petal.Length" "Petal.width" "species"
> library(ggplot2)
warning: package 'ggplot2' was built under R version 4.3.1 Learn more about the underlying theory at https://ggplot2-book.org/
> ggplot(iris,aes(x=Sepal.Length,y=Petal.width)) + geom_point()
>

```

Files

Install

System L

abir

askj

bac

bas

bas

bit

bit6

blot

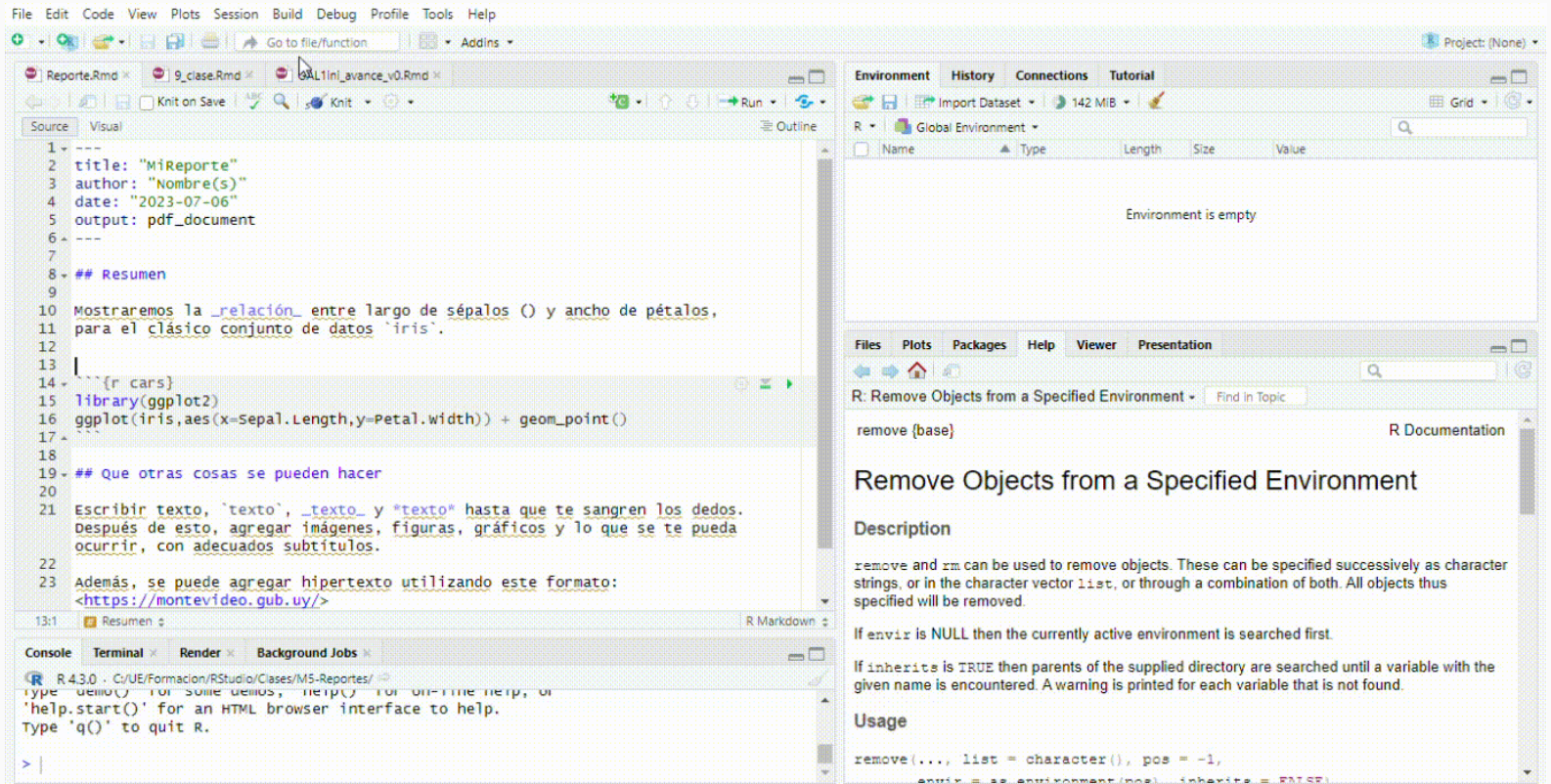
boo

bre

bric

bro

...



ADVERTENCIA: si al compilar en PDF da error, realizar estos pasos en la *consola* de R:

1. `install.packages("tinytex")`
2. `tinytex::install_tinytex()` (demora unos minutos en instalarse)



MiReporte

Nombre(s)

2023-07-06

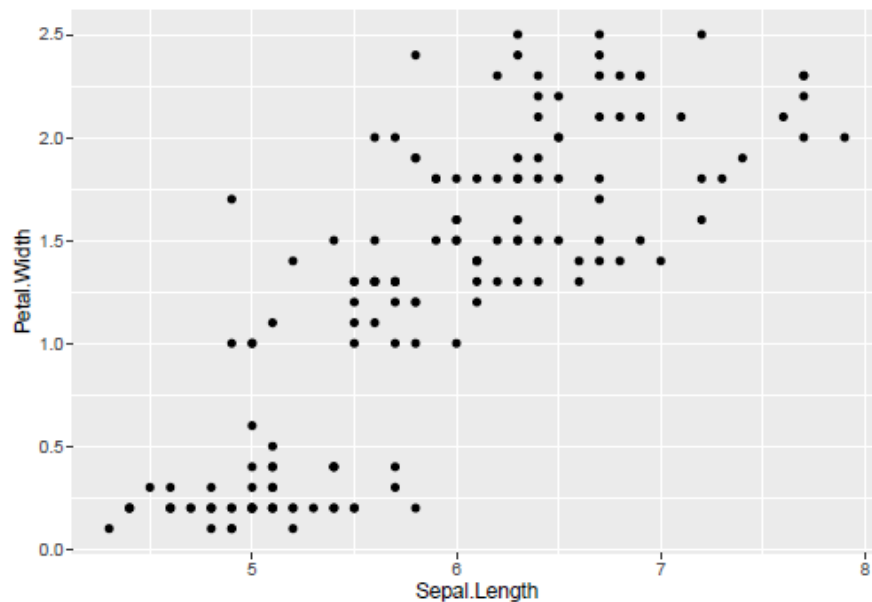
Resumen

Mostraremos la *relación* entre largo de sépalos () y ancho de pétalos, para el clásico conjunto de datos `iris`.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
ggplot(iris,aes(x=Sepal.Length,y=Petal.Width)) + geom_point()
```



Anatomía de un RMarkdown

```
1 ---
2 title: "MiReporte"
3 author: "Nombre(s)"
4 date: "2023-07-06"
5 output: pdf_document
6 ---
7
8 ## Resumen
9
10 Mostraremos la relación entre largo de sépalos () y ancho de pétalos,
11 para el clásico conjunto de datos `iris`.
12
13
14 ```{r cars}
15 library(ggplot2)
16 ggplot(iris,aes(x=Sepal.Length,y=Petal.width)) + geom_point()
17 ```
18
19 ## Que otras cosas se pueden hacer
20
21 Escribir texto, `texto`, texto y *texto* hasta que te sangren los dedos. Después
22 de esto, agregar imágenes, figuras, gráficos y lo que se te pueda ocurrir, con
23 adecuados subtítulos.
24
25 Además, se puede agregar hipertexto utilizando este formato:
26 <https://montevideo.gub.uy/>
27
28 Y muchas, muchas cosas más...
```

Preámbulo

Metadatos del documento: YAML

Formato general del documento: título, autor/a, tipo de documento (ej: html_document, pdf_document, word_document...), adornos

- Por defecto

- Parámetros pedidos al iniciar archivo
- Se pueden agregar más

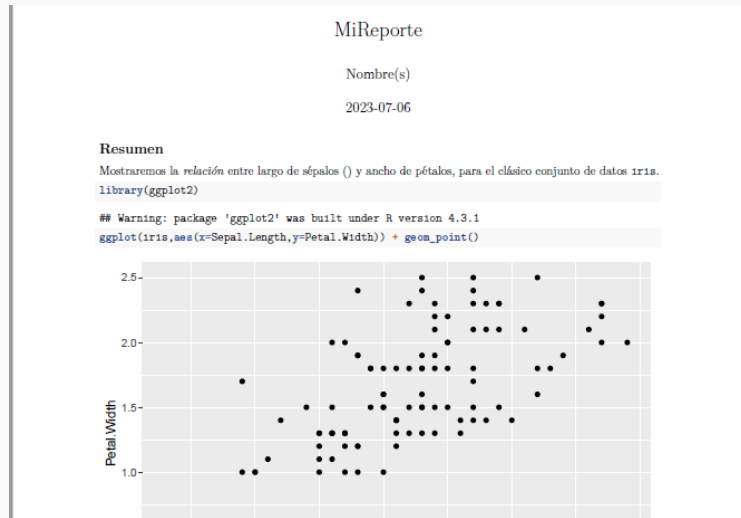
```
1 ---
2 title: "MiReporte"
3 author: "Nombre(s)"
4 date: "2023-07-06"
5 output: pdf_document
6 ---
```

- Agregamos como ejemplo:

- Fecha con código **R** (solo mes y año)
- Ajustamos tamaño de página en documento
- Agregamos logo IM

```
1 ---
2 title: "MiReporte"
3 author: "Nombre(s)"
4 date: "`r format(Sys.Date(),format='%B %Y')`"
5 output: pdf_document
6 header-includes:
7   - \usepackage{titling}
8   - \pretitle{\begin{center}}
9     \includegraphics[width=2cm,height=2cm]{logo_IM.png}\LARGE\\
10   - \posttitle{\end{center}}
11 geometry: left=1.5cm,right=1.5cm,top=2cm,bottom=2.5cm
12 ---
13
```

Antes:



Después:



Encabezado **YAML** + **Latex**

```
1 ---
2 title: "Solución: Ejercicio 1"
3 subtitle: "Importar datos y Funciones de dplyr"
4 author:
5   - Unidad de Estadística
6   - Servicio de Gestión Estratégica
7   - Departamento de Desarrollo Sostenible e Inteligente
8 date: "`r format(Sys.time(), '%d %B, %Y')`"
9 output:
10  pdf_document:
11    toc: true
12    toc_depth: 2
13  header-includes:
14    - \usepackage{booktabs}
15    - \usepackage[spanish]{babel}
16    - \usepackage{caption}
17 ---
18
19
20 \captionsetup[figure]{name=Gráfico}
21 \captionsetup[table]{name=Tabla}
22
23 \listoftables
24 \listoffigures
```

Agrega INDICE con profundidad 2

Agrega paquetes de latex

Código Latex para renombrar el prefijo de tablas y figuras

Código Latex para agregar Índice de figuras e Índice de tablas

Solución: Ejercicio 1	
Importar datos y Funciones de dplyr	
Unidad de Estadística Servicio de Gestión Estratégica	
Departamento de Desarrollo Sostenible e Inteligente	
03 Julio, 2023	
Índice	
Introducción	2
Librerías y conflictos	2
Arreglamos conflictos	2
Importar datos	2
Crear nuevas variables	2
Resumen de datos	3
Uso de lista y filtrado	4
Creamos carpeta y guardamos en RData	4
Tabla usando la funcion kbl	4
Grafico	5
Índice de cuadros	
1. salsfjs	5
Índice de figuras	
1. salsdfgsd	5

- Agregando **toc: yes** y **toc_depth: 2** le estamos agregando un índice con profundidad 2 al documento
- Podemos incluir paquetes de latex para configurar algunas cosas:
 - `\usepackage[spanish]{babel}` para configurar el idioma.
 - `\usepackage{caption}` para cambiar el prefijo de tablas y figuras.

Configuración inicial

Chunks **R** al inicio del documento (se puede modificar en cada chunk)

```
```${r} setup, include=FALSE}  
knitr::opts_chunk$set(echo=FALSE,message=FALSE,eval=TRUE)
```
```

Texto

Markdown

Formatos

`*italics*` and `_italics_`

`**bold**` and `__bold__`

`superscript^2^`

`~~strikethrough~~`

`[link](www.rstudio.com)`

`# Header 1`

`## Header 2`

`### Header 3`

`#### Header 4`

`##### Header 5`

`##### Header 6`

italics and *italics*

bold and **bold**

superscript²

~~strikethrough~~

[link](#)

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

Listas

Sin orden: * - +

```
74 ## Listas sin orden
```

```
75
```

```
76 Con asteriscos
```

```
77
```

```
78 * Pepe
```

```
79 * Juan
```

```
80 * Ana
```

```
81
```

```
82 Con guión
```

```
83
```

```
84 - Pepe
```

```
85 - Juan
```

```
86 - Ana
```

```
87
```

```
88 Con más
```

```
89
```

```
90 + Pepe
```

```
91 + Juan
```

```
92 + Ana
```

```
93
```

3.1. Listas sin orden

Con asteriscos

- Pepe

- Juan

- Ana

Con guión

- Pepe

- Juan

- Ana

Con más

- Pepe

- Juan

- Ana

Listas

Ordenadas: 1,2,3

```
94 ▾ ## Listas con orden
95
96 Numero consecutivo
97
98 1. Primero
99 2. Segundo
100 3. Tercero
101
102 Comienza ordenando por el mayor
103
104 3. Primero
105 1. Segundo
106 2. Tercero
107
108 Número constante
109
110 5. Primero
111 5. Segundo
112 5. Tercero
```

3.2. Listas con orden

Numero consecutivo

1. Primero
2. Segundo
3. Tercero

Comienza ordenando por el mayor

3. Primero
4. Segundo
5. Tercero

Número constante

5. Primero
6. Segundo
7. Tercero

Listas

Anidadas con y sin orden

```
114 ▾ ## Lista anidada
115
116   Con orden
117
118   1. Primero
119   1. Segundo
120   1. Tercero
121     1.Uno
122     1.Dos
123
124   Sin orden
125
126   + Pepe
127   + Juan
128   + Ana
129     - algo
130     - Otro
```

3.3. Lista anidada

Con orden

1. Primero
2. Segundo
3. Tercero 1.Uno 1.Dos

Sin orden

- Pepe
- Juan
- Ana
 - algo
 - Otro

Hipervínculos y Notas al pie

- Los Hipervínculos se ponen entre signos de mayor y menor (`<url>`), si no queremos que se muestre la url se le agrega un nombre entre parentesis rectos y la url entre parentesis curvos `[nombre](url)`. Si queremos que las url salgan con un color distinto dentro del encabezado YAMAL se agrega, por ejemplo, `urlcolor: blue`

En este párrafo vamos a poner un hipervínculo a la página de la IM donde aparece la url `<https://montevideo.gub.uy/>` y aquí pondremos que aparezca el nombre `[linkIM](https://montevideo.gub.uy/)` en vez de la url.

En este párrafo vamos a poner un hipervínculo a la página de la IM donde aparece la url `https://montevideo.gub.uy/` y aquí pondremos que aparezca el nombre `linkIM` en vez de la url.

Nota a pie: texto^[Nota_al_pie]

136 En este párrafo vamos a agregar una nota al pie justo aquí^[Hola, soy una nota al pie] bla bla bla bla....

En este párrafo vamos a agregar una nota al pie justo aquí¹ bla bla bla bla....

¹Hola, soy una nota al pie

<!-- -->

<!--

-->

Código R dentro de texto

Expresión:

El data.frame iris tiene 150 filas y 5 columnas

```
`r codigo_R`
```

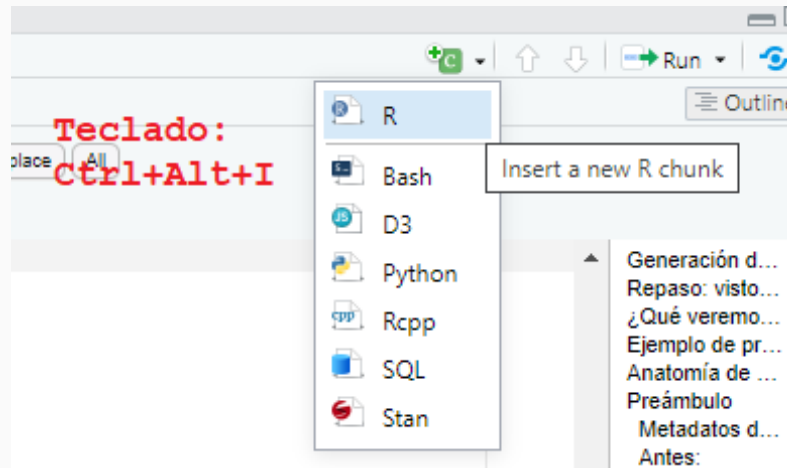
El data.frame iris tiene `r nrow(iris)` filas y `r ncol(iris)` columnas

Chunks de código R

Los chunks son trozos de código que pueden ser ejecutados una vez compilado el documento. Con éstos se puede

- Configurar documento
- Insertar imágenes, gráficos, etc.
- Mostrar resultados de código R

Chunks de código R



--

| Parámetro | Valor por defecto | Qué hace |
|-----------|-------------------|-----------------------------------|
| eval | TRUE | evalúa código, muestra resultados |
| echo | TRUE | muestra código R en documento |
| warning | TRUE | muestra advertencias en documento |
| error | TRUE | muestra los errores |
| message | TRUE | muestra todos los mensajes |

Tablas e imágenes

Tablas con `kable`

Generador de tablas simples

- Función dentro de paquete `knitr`
- Permite generar tablas provenientes de *datos rectangulares* (matrices, `data.frames`)
- Código interno generado en *L^AT_EX*

```
iris <- iris[1:4,]  
  
#Insertamos tabla simple con knitr::kable()  
kable(iris,  
      caption="Extracto de datos Iris")
```

Extracto de datos Iris

| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|--------------|-------------|--------------|-------------|---------|
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | setosa |

Tablas con `kable`

Mejoras a tablas

- Posibilita algunas mejoras en apariencia
- Parámetros: alineación, formato, cantidad decimales,...

```
iris <- iris[1:4,]  
  
#Insertamos tabla simple con knitr::kable(); agregamos mejoras  
kable(iris,  
      caption="Extracto de datos Iris",  
      align="c", #alineacion central  
      booktabs=TRUE, #elimina cuadriculado  
      col.names=c("Largo Sepalo", "Ancho Sepalo", "Largo Petalo", "Ancho Petalo", "Especie"),  
      format="latex") #genera tabla con codigo Latex
```

| Extracto de datos Iris | | | | |
|------------------------|--------------|--------------|--------------|---------|
| Largo Sepalo | Ancho Sepalo | Largo Petalo | Ancho Petalo | Especie |
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | setosa |

Tablas mejoradas con `kableExtra`

Extensión de funcionalidades de `knitr::kable()`

- Paquete que mejora cuestiones estéticas de tablas kable
- Ventaja: funcionalidad aplica para salidas HTML y PDF por igual

```
library(kableExtra)
#Insertamos tabla simple con knitr::kable()
kable(iris,
      caption="Extracto de datos Iris",
      align="c", #alineacion central
      booktabs=TRUE, #elimina cuadriculado
      col.names=c("Largo Sepalo", "Ancho Sepalo", "Largo Petalo", "Ancho Petalo", "Espe-
                    format="latex") %>%
      kable_styling(latex_options="striped")
```

| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species |
|--------------|-------------|--------------|-------------|---------|
| 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | setosa |

Tablas mejoradas con `kableExtra`

Más funcionalidades

- Fijar posición de tabla en documento:
- Fijar estética por filas/columnas
 - Fila: `kable_styling() %>% row_spec(2:5,height="4cm")`
 - Columna: `kable_styling() %>% column_spec(2,width="10cm")`
- Tamaño de fuente: `kable_styling(font_size=8)`
- Agregar nota al pie: `kable_styling() %>% footnote(...)`

Imágenes

Parámetros gráficos en chunk

Dentro del encabezado del chunk, estos parámetros comienzan con **fig.**

| Parámetro | Qué hace | Ejemplo |
|-----------------|----------------------------------|-----------------------|
| fig.height | altura (en pulgadas por defecto) | fig.height=5 |
| fig.width | ancho (en pulgadas por defecto) | fig.width=3 |
| fig.scale | escala/tamaño | fig.scale=0.8 |
| fig.align | alineación | fig.align='center' |
| fig.cap/sub.cap | Título/subtítulo | fig.subcap="Detalles" |
| fig.pos | posición en documento | fig.pos='H' |
| fig.path | Directorio donde está imagen | fig.path='C:/Carpeta' |

Imágenes

Incluir imágenes con `include.graphics()`

- Función del paquete `knitr` que permite agregar imágenes en chunk
- Resto de parámetros (altura, ancho, alineación, etc.) se controlan desde el preámbulo del chunk

Bibliografía adicional recomendada

- Cheatsheet (guía) RMarkdown
- "Libro de cocina" (cookbook) de RMarkdown
- Más guías para utilizar R y RStudio
- R para Ciencia de Datos (en español)
- Quarto: próxima generación de documentos RMarkdown