Módulo 5: Documentación

RMarkdown y knitr

Leticia Debera, Darío Padula y Daniel Alessandrini Intendencia de Montevideo 2023/09/07 Generación de reportes automáticos utilizando Rmarkdown + knitr

Repaso: visto hasta ahora

- Introducción / conceptos básicos: aprendimos
 - O Qué es R, interfaz gráfica RStudio
 - o RMarkdown: estructura básica
 - Elementos: objetos, estructura, accesibilidad, operadores
 - Objetos: vectores, data.frame, listas,...
- Manipulación de datos:
 - Importación y Exportación de datos
 - o tuberías o pipes %>%
 - o dplyr: verbos (mutate, filter, group_by, summarise...), funciones (mutate_at/if, summarise_at/if, across...)
 - o 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
- O Uso muy extendido: ej. WhatsApp

• knitr

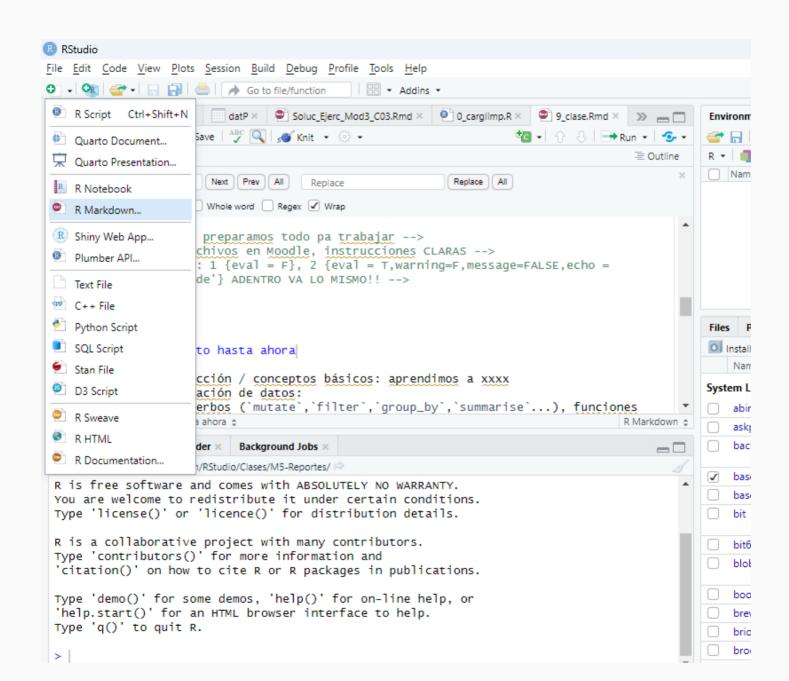
- Paquete de R que lo vincula con Markdown
- Transforma código R en lenguajes de documentación, permitiendo reproducibilidad

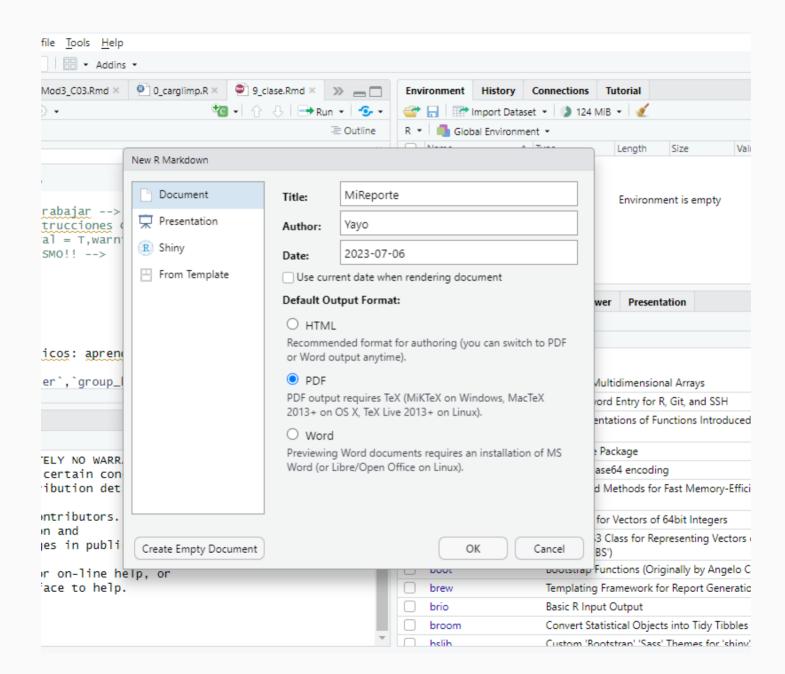
Pandoc

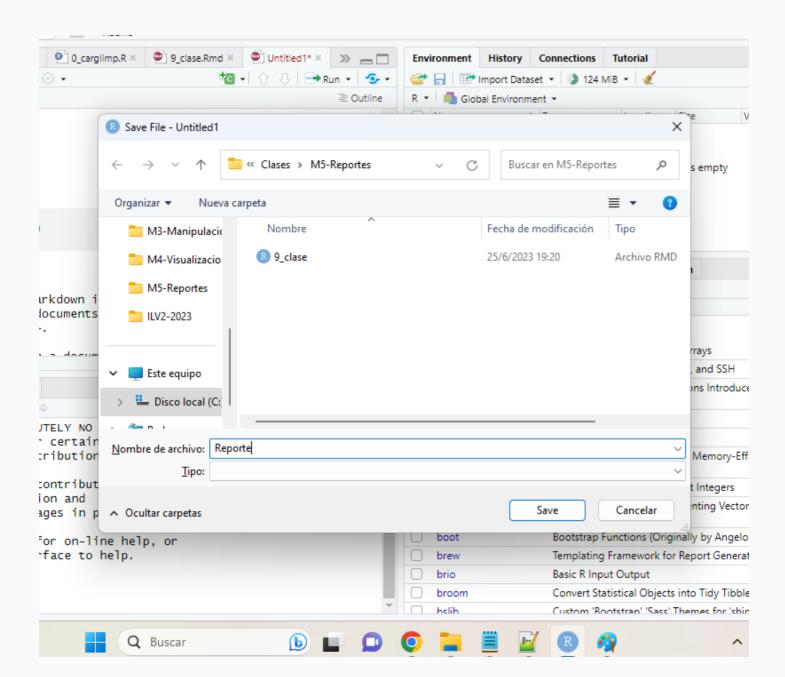
- Pandoc: conversor de documentos Markdown a otros formatos (PDF, HTML, DOC,)
- Mirar ventana Render...

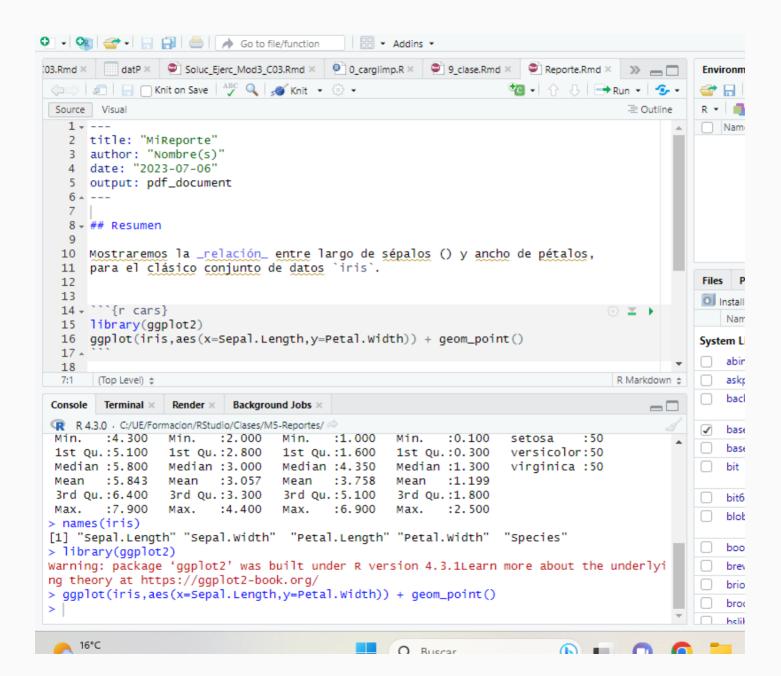
Rmd >>knitr>> Markdown >>Pandoc>> compilación: HTML, PDF, DOC, etc

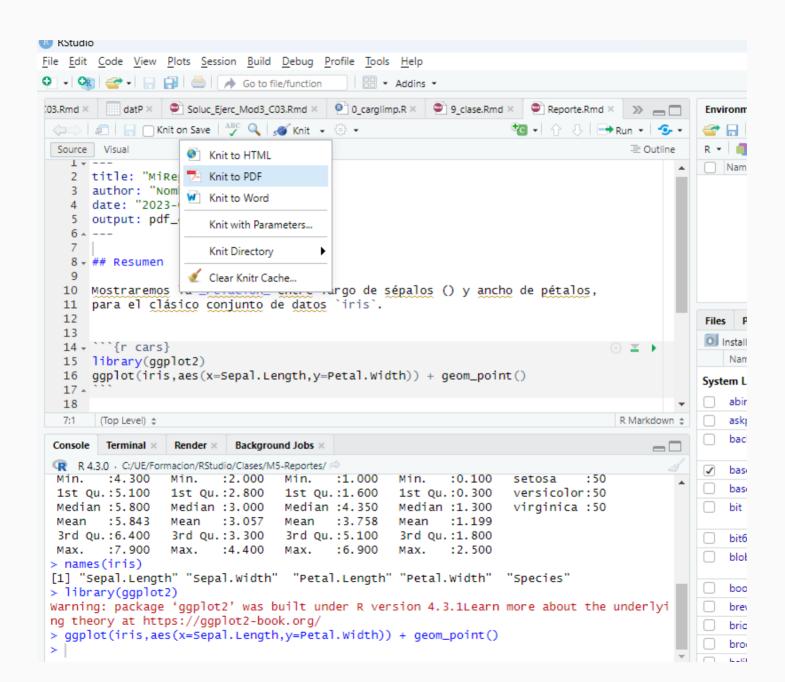
Ejemplo de principio a fin

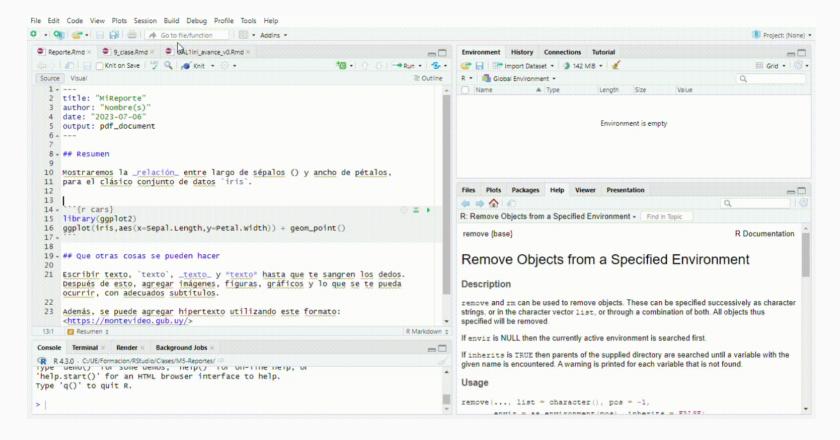






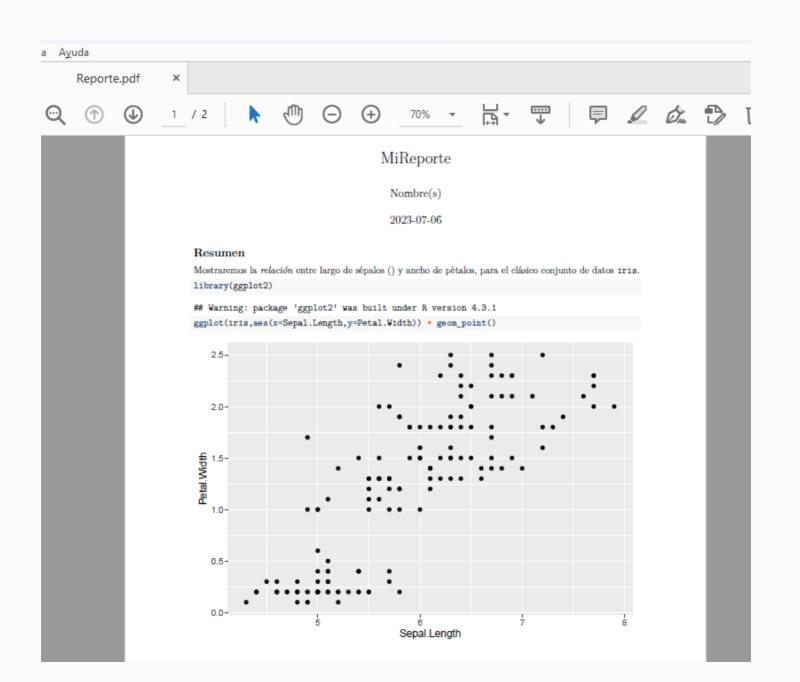






ADVERTENCIA: si al compilar en PDF da error, realizar estos pasos en la *consola* de R:

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



Anatomía de un RMarkdown

```
GAL1ini avance v0.Rmd × 9 clase,Rmd × PReporte,Rmd ×
আ • | ↑ ড | ➡ Run • | ★ •
Source Visual
                                                                               Outline
  2 title: "MiReporte"
                              preámbulo
  3 author: "Nombre(s)"
                                  YAML
  4 date: "2023-07-06"
  5 output: pdf_document
  8 - ## Resumen
 10 Mostraremos la _relación_ entre largo de sépalos () y ancho de pétalos,
     para el clásico conjunto de datos `iris`.
 12
 13
     ```{r cars}
 14
 chunk
 library(qqplot2)
 15
 ggplot(iris.aes(x=Sepal.Length.v=Petal.Width)) + geom_point()
 16
 R
 17
 18
 19 + ## Oue otras cosas se pueden hacer
 20
 21 Escribir texto, `texto`, _texto_ y *texto* hasta que te sangren los dedos. Después
 de esto, agregar imágenes, figuras, gráficos y lo que se te pueda ocurrir, con
 adecuados subtítulos.
 22
 Además, se puede agregar hipertexto utilizando este formato:
 <https://montevideo.gub.uy/>
 24
 25 Y muchas, muchas cosas más...
 26
 4:19
 # MiReporte $
 R Markdown $
```

# 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

- o 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 ---
```

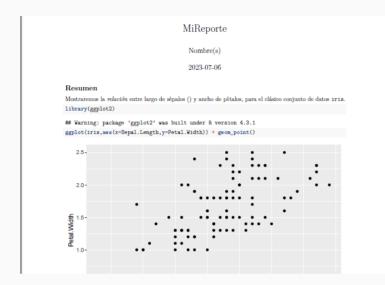
#### • Agregamos como ejemplo:

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

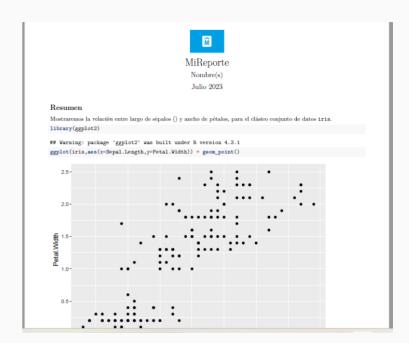
```
title: "MiReporte"
author: "Nombre(s)"
date: "`r format(Sys.Date(),format='%B %Y')`"
output: pdf_document
header-includes:
 - \usepackage{titling}
 - \pretitle{\begin{center}
 \includegraphics[width=2cm,height=2cm]{logo_IM.png}\LARGE\\}
 - \posttitle{\end{center}}
geometry: left=1.5cm,right=1.5cm,top=2cm,bottom=2.5cm

13
```

## Antes:



## Después:



## Encabezado YAML + Latex

```
title: "Solución: Ejercicio 1"
subtitle: "Importar datos y Funciones de dplyr"
author:
 - Unidad de Estadística
 - Servicio de Gestión Estratégica
 - Departamento de Desarrollo Sostenible e Inteligente
date: "`r format(Sys.time(), '%d %B, %Y')
output:
pdf_document:
 toc_depth: 2
header-includes
-\usepackage{booktabs}
-\usepackage[spanish]{babe]}
-\usepackage{caption}
\captionsetup[figure]{name=Gráfico} | Código Latex para renombrar
\captionsetup[table]{name=Tabla}
 el prefijo de tablas y figuras
\listoftables
 \listoffiaures
```

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. safsffs	5
Índice de figuras	
1. safsdgfsd	5
1. Satisfied	9

- 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
                                3.1. Listas sin orden
75
76
    Con asteríscos
                                Con asteríscos
77
78
    * Pepe
                                  Pepe
79
    * Juan

    Juan

80
                                  Ana
    * Ana
81
82
    Con guión
                                Con guión
83
84
     - Pepe
                                  Pepe
85
     – Juan

    Juan

86
    - Ana
                                  Ana
87
88
    Con más
                                Con más
89
90
    + Pepe
                                  Pepe
91
    + Juan

    Juan

    + Ana

    Ana
```

Listas

Ordenadas: 1,2,3

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

Listas

Anidadas con y sin orden

```
114 - ## Lista anidada
                             3.3. Lista anidada
115
116
     Con orden
                             Con orden
117
118 1. Primero
                               1. Primero
119 1. Segundo
                               2. Segundo
120
     1. Tercero
                               3. Tercero 1.Uno 1.Dos
121
        1.Uno
122
        1.Dos
                             Sin orden
123
124
     Sin orden
125
                                Pepe
126
     + Pepe

    Juan

127
     + Juan
                                Ana
128
     + Ana

    algo

129
        - algo

    Otro

130
        - 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 parentesís 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 <a href="https://montevideo.gub.uy/">https://montevideo.gub.uy/</a> 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]

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....

1Hola, soy una nota al pie

Markdown

Expresiones matemáticas

Ecuaciones entre $: e^{i\eta}+1=0$ devuelve $e^{i\pi}+1=0$

Ecuaciones en bloque entre \$\$: la expresión

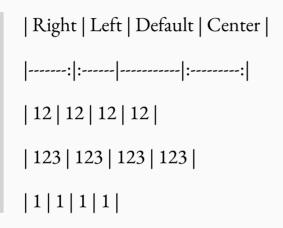
 $\label{left(sum{k=1}^n a_k b_k | right)^2 | leq | left(|sum{k=1}^n ak^2 | right) | left(|sum{k=1}^n ak^2 | right) | sum{k=1}^n b_k^2 \right) $$$

devuelve la desigualdad de Cauchy-Schwarz:

$$\left(\sum_{k=1}^n a_k b_k
ight)^2 \leq \left(\sum_{k=1}^n a_k^2
ight) \left(\sum_{k=1}^n b_k^2
ight)$$

Markdown

Tablas



Right	Left	Default	Center
12	12	12	12
123	123	123	123
1	1	1	1

Otras herramientas para introducir y mejorar texto

$L\!\!\!\!/T_E\!X$

Podemos crear un \textbf{texto} con agregados en \LaTeX usando expresiones entre \$:

Podemos crear un **texto** con agregados en LATEX usando expresiones entre \$

CSS

Necesario agregar opción css al preámbulo YAML

•••

output: html_document

css: "style.css"

A [green] {.my-color} word.

devuelve

A green word.

Código R y chunks

Código R dentro de texto

Expresión:

El data.frame iris tiene 150 filas y 5 columnas

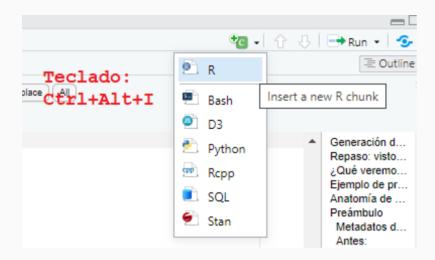
```
`r codigo_R`
El data.frame iris tiene `r nrow(iris) filas y `r
ncol(iris) `columnas
```

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

RMarkdown avanzado

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 LATEX

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,...

Extracto de datos Iris				
Largo Sepalo	Ancho Sepalo	Largo Petalo	Ancho Peta	loEspecie
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

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

```
o Fila: kable_styling() %>% row_spec(2:5,height="4cm")
```

- o 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

Referenciar documentos

Referencias internas

- Formas de citar un documento como referencia o fuente de otro
- Permite unir documentos en mismo texto para facilitar su ubicación
- Importante: instalar paquete bookdown e indicar en preámbulo YAML convenientemente:

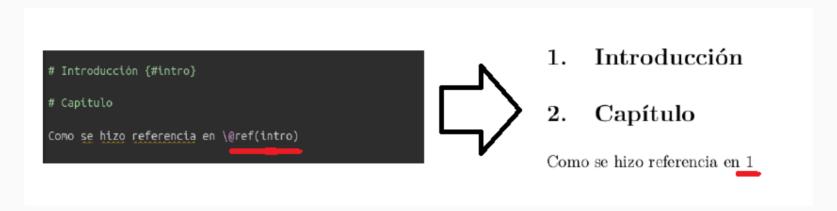
```
...
output:
bookdown::pdf_document2
```

Referencias cruzadas

- o Sección
- o Tabla
- Imagen

Referencias cruzadas para secciones

Se referencia con # la sección de interés y se la llama en el documento utilizando \@ref(seccion)



Referencias cruzadas para tablas

- Necesario darle nombre al chunk donde se encuentra la tabla
- Debe hacerse entre enseguida de la primer r: {r XXX, ...}

Código en informe Rmd:

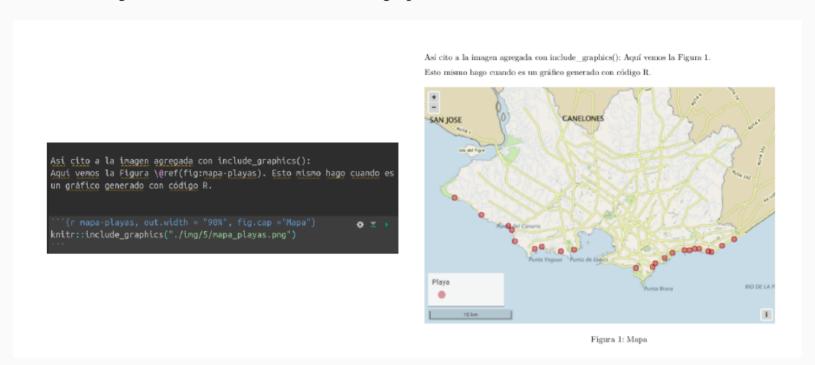
```
675
676
677 Como se muestra en la tabla \@ref(parGraf) ...
678
```

Como se ve en el archivo compilado:

Como se muestra en la tabla 10 ...

Referencias cruzadas para imágenes

- Fundamental agregar parámetro fig.cap=... en el chunk
- Para imágenes incluidas con knitr::include_graphics() se utiliza \@ref(nombre_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