

# R Markdown

Programación para el análisis de datos

Departamento de Ciencias Sociales, UCU - Martín Opertti

# Introducción

# Introducción a R Markdown

- R Markdown permite crear documentos combinando texto, código y sus resultados. Los documentos de R Markdown tienen extensión `.Rmd` y admiten muchos formatos de salida como archivos PDF, word, HTML o diapositivas.
- Los reportes creados con R Markdown pueden usarse como procesador de texto sin necesidad de incluir código (ej. escribir artículos) o incluir texto y código para mostrar también el proceso detrás de los resultados.
- R Markdown integra un conjunto de paquetes y herramientas externas. Necesita el paquete `rmarkdown` pero RStudio lo descarga de forma automática así que no es necesario instalar.
- [R Markdown Cheatsheets](#)

# ¿Por qué usar R Markdown?

- Es libre y gratuito
- Es compatible con varios lenguajes de programación
- Muchos formatos de salida
- Facilita la investigación reproducible y la divulgación del conocimiento

# Formatos de salida

- Formato PDF
- Formato HTML personalizado
- Formato HTML personalizado
- Esta misma presentación (con el paquete `xaringan`)

# ¿Cómo se ve un archivo .Rmd?

```
1 ---
2 title: "Template con datos de Gapminder"
3 author: "Martín Opertti"
4 date: "2022-10-10"
5 output: html_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 # Cargo paquetes
11 library(tidyverse)
12 library(gapminder)
13
14
15 ## Gapminder
16
17 La Fundación [Gapminder](https://www.gapminder.org/) es una empresa sin fines de lucro registrada en Estocolmo, Suecia,
18 que promueve el desarrollo global sostenible y el logro de los Objetivos de Desarrollo del Milenio de las Naciones
19 Unidas mediante un mayor uso y comprensión de las estadísticas y otra información sobre el desarrollo social,
20 económico y ambiental a nivel local, nacional. y niveles globales
21
22 {r gap}
23 # Asigno y filtro data
24 df_gap <- gapminder %>%
25   filter(year == 2007)
26
27
28 ## Gráfico
29
30 Incluyo el siguiente código y resultado
31
32 {r pressure, echo=FALSE}
33 hist(gapminder$lifeExp)
```

# Resultado (output)

## Template con datos de Gapminder

Martin Opertti

2022-10-10

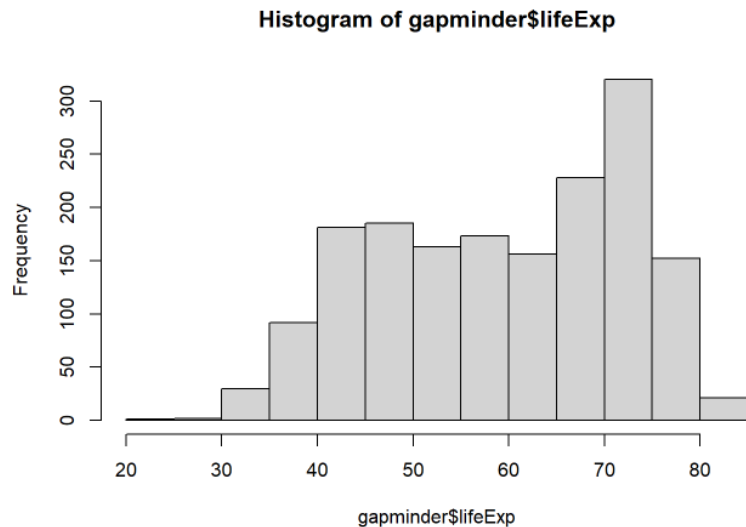
### Gapminder

La Fundación [Gapminder](#) es una empresa sin fines de lucro registrada en Estocolmo, Suecia, que promueve el desarrollo global sostenible y el logro de los Objetivos de Desarrollo del Milenio de las Naciones Unidas mediante un mayor uso y comprensión de las estadísticas y otra información sobre el **desarrollo social, económico y ambiental** a nivel local, nacional, y niveles globales

```
# Asigno y filtro data
df_gap <- gapminder %>%
  filter(year == 2007)
```

### Gráfico

Incluyo el siguiente código y resultado



# .Rmd y output

The image displays two side-by-side windows. The left window, titled 'example.Rmd', shows the source R Markdown code. The right window, titled 'example.html', shows the rendered HTML output of the same document.

**Left Window (example.Rmd):**

```
1 # Header 1
2
3 This is an R Markdown document. Markdown is a
4 simple formatting syntax for authoring webpages.
5 Use an asterisk mark to provide emphasis, such
6 as italics or bold.
7 Create lists with a dash:
8
9 - Item 1
10 - Item 2
11 - Item 3
12
13 ```
14 Use back ticks to
15 create a block of code
16 ```
17
18 Embed LaTeX or MathML equations,
19 
$$\frac{1}{n} \sum_{i=1}^n x_i$$

20
21 Or even footnotes, citations, and a
22 bibliography. [1]
23
24 [^1]: Markdown is great.
```

**Right Window (example.html):**

## Header 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring web pages.

Use an asterisk mark to provide emphasis, such as *italics* or **bold**.

Create lists with a dash:

- Item 1
- Item 2
- Item 3

Use back ticks to create a block of code

Embed LaTeX or MathML equations, 
$$\frac{1}{n} \sum_{i=1}^n x_i$$

Or even footnotes, citations, and a bibliography. <sup>1</sup>

---

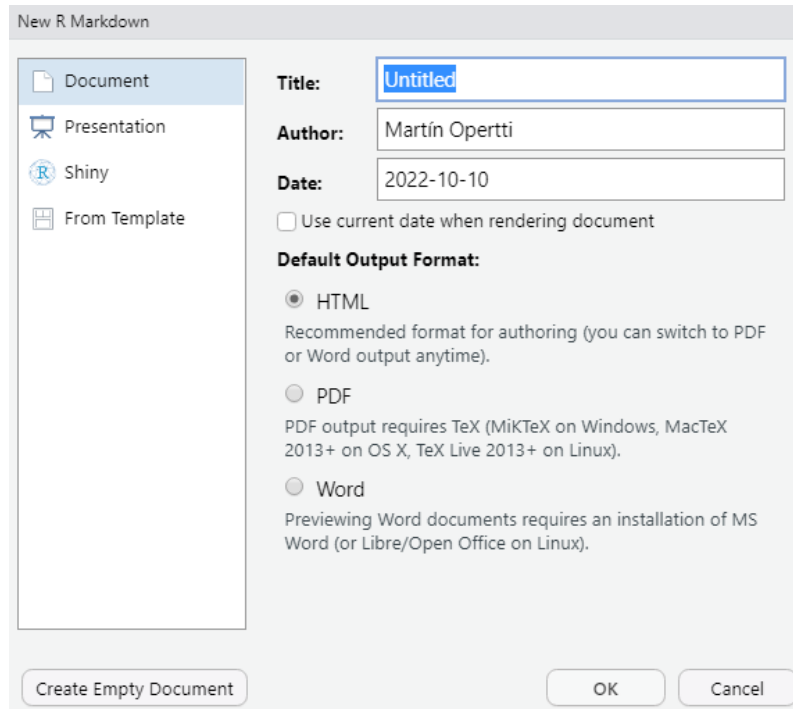
1. Markdown is great. ↩



¿Cómo crear un archivo .Rmd?

# ¿Cómo crear un archivo .Rmd?

- Para crear un archivo de R Markdown seleccionamos `File/New File/R Markdown`
- Luego ingresamos nombre, autor y formato de salida (aunque podemos modificarlo luego en el código)



The screenshot shows the 'New R Markdown' dialog box. On the left, there is a sidebar with four options: 'Document' (selected), 'Presentation', 'Shiny', and 'From Template'. The main area contains the following fields and options:

- Title:** A text box containing 'Untitled'.
- Author:** A text box containing 'Martín Operti'.
- Date:** A text box containing '2022-10-10'.
- ☐ Use current date when rendering document
- Default Output Format:**
  - ☒ **HTML**  
Recommended format for authoring (you can switch to PDF or Word output anytime).
  - ☐ **PDF**  
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).
  - ☐ **Word**  
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

At the bottom, there are three buttons: 'Create Empty Document', 'OK', and 'Cancel'.

# Compilar

Cuando queremos ver el resultado de nuestro archivo .Rmd tenemos que compilarlo (render). Para ello arriba del script veremos un botón que dice `Knit` y luego seleccionamos `knit`.

Esto generará otro documento con el formato de salida seleccionado. También podemos hacerlo con (`ctrl + shift + k`)

# YAML header

En la parte de arriba de documento se encuentra el YAML donde definimos el título, autor, la fecha, formato de salida entre otros argumentos, demarcado encima y debajo por ---

A modo de ejemplo, si queremos exportar un documento en formato HTML, podemos incluir un TOC (table of contents) para ordenar el documento. También podemos elegir un **tema**. Para ello, simplemente incluimos lo siguiente:

```
1 ▾ ---
2 title: "Template con datos de Gapminder"
3 author: "Martín Opertti"
4 date: '2022-10-10'
5 output:
6   html_document:
7     toc: true
8     toc_float: true
9     theme: cerulean
10 ▲ ---
```

# Títulos

```
34
35 ▼ # Encabezado 1
36
37 ▼ ## Encabezado 2
38
39 ▼ ### Encabezado 3
40
```

Encabezado 1

Encabezado 2

Encabezado 3

# Formato

Con asteriscos podemos poner texto en negrita y también armar listas

```
43  **Lista**
44
45  * Items
46    * Item 1
47    * Item 2
48    * Item 3
49
50  * Items
51    * Item 1
52    * Item 2
53    * Item 3
```

## Lista

- Items
  - Item 1
  - Item 2
  - Item 3
- Items
  - Item 1
  - Item 2
  - Item 3

# Fragmentos de código

- Para insertar un fragmento de código (code chunk) podemos `Ctrl + Alt + I`
- Dentro de el podemos escribir código (que se visualizará de forma distinta en el reporte). Podemos correr el código localmente en el archivo `.Rmd` o que se corra al compilar
- En la primer línea del fragmento de código podemos definir argumentos. Por ejemplo si incluimos `echo = FALSE`, el código no aparecerá pero si el resultado, o si definimos `results = "hide"` veremos el código pero no el resultado. Con `message = FALSE` suprimen los mensajes en la consola (recomendado) y con `eval = FALSE` aparecerá el código (si `echo = TRUE`) pero no será evaluado por R.
- Se ven de esta forma:

```
20 {r gap}
21 # Asigno y filtro data
22 df_gap <- gapminder %>%
23   filter(year == 2007)
24
25
```

# Tablas e imagenes

- Hay muchas formas de incluir tablas en un documento de R Markdown. Una de ellas es `knitr::kable()`. Pueden ver más detalles [aquí](#)
- Para incluir imágenes pueden utilizar `knitr::include_graphics()`.
- Para ambos casos hay un ejemplo en el template



# Trabajo final

# Pauta

Escribir un informe (con R Markdown) que integre código y texto detallando algún procesamiento, análisis o recolección de datos -en su acepción más amplia- que no haya sido visto en el curso. El objetivo es que busquen tutoriales, artículos o libros de análisis de datos en R que no hayan sido vistos en el curso, y que lo apliquen a datos que les interesen. Tienen que describir en qué consiste el análisis, procesamiento o recolección de datos pero lo central es que expliquen cómo funciona el código.

El trabajo debe:

- Describir el tipo de análisis, procesamiento o recolección de forma conceptual.
- Describir uso y argumentos de las funciones utilizadas.
- Aplicar las funciones a datos que no sean los mismos que usan los recursos que toman como fuente.
- Citar todos los recursos utilizados tanto para código como para la descripción conceptual.
- Realizar un procesamiento de los datos a utilizar aplicando lo aprendido en el curso

# Formalidades

- Fecha de entrega y presentación: 29 de noviembre
- Se entrega una carpeta con los siguientes elementos:
  - Proyecto de R
  - Archivo de R Markdown (.Rmd)
  - Archivo de salida de R Markdown (HTML o PDF, 2 a 5 carillas)
  - Script con procesamiento de los datos a utilizar
  - Data (si aplica)
- Grupos de a 2 o individual
- Taller para consultas 24 de noviembre

# Criterio de evaluación

- El trabajo final es el 40% de la nota del curso
- La calificación es sobre 12:
  - 2 puntos por cumplir con todos los elementos requeridos y usarlos de forma correcta
  - 2 puntos por usar de forma adecuada el R Markdown (integrar texto, código, imágenes, tablas, etc.)
  - 2 puntos por crear un script para procesar los datos utilizando lo visto en el curso (con anotaciones sobre por qué estamos haciendo esas transformaciones)
  - 6 puntos por la descripción y aplicación de las funciones o paquetes que no vimos en el curso.

# Ideas

- Visualizaciones de datos
  - Gráficos de densidad
  - Mapas y mapas de calor
  - Gráficos animados
- Utilizar datos de encuesta en R
  - Tablas ponderadas
  - Manejar etiquetas
  - Crear ponderadores
- Análisis de texto
  - Palabras más frecuentes en data de texto
  - Análisis de sentimiento
- Conectarse con APIs para extraer datos
  - Google perspectives
  - Twitter
  - Banco Mundial
  - Our World in Data
- Webscrapping
- Estadística avanzada
  - Modelos específicos
  - Machine Learning
- Paquetes para manejar variables de cadenas, fechas, etc.

# Datos

- [Kaggle](#)
- [LAPOP](#)
- [Latinobarómetro](#)
- [Encuesta Mundial de Valores](#)
- [Banco Mundial](#)
- [Our World in Data](#)
- [Otras](#)