

# Mini curso marzo 2021: Reportes y documentación con R markdown



Dra. Leticia Vega



Dr. Alejandro Ponce-M

# Índice

- Mi experiencia con R Markdown
  - Alex
  - Leti
- Fuentes y ejemplos
- Proceso de crecimiento de R Markdown
- Literatura del curso
- Partes del R Markdown
- ¡Vámonos a R Markdown en RStudio!!

# Mi experiencia con R Markdown



- Reportes internos de Trabajo
- Reportes para agricultores de los resultados obtenidos
- Análisis de datos

**Reporte Sobre los indicadores para PPR**

Conabio  
25. January 2021

---

**Dentro del PIR responde al OUTPUT 1.1.1**

Número de áreas de implementación con proyectos en marcha

En el proyecto se han trabajado en cinco de las seis áreas de implementación. Del 2020-07-01 al 2020-12-31 se ha trabajado en 0 áreas. Queda pendiente el área de la Ciudad de México.

Los avances en este período con respecto a las metas se pueden ver en la siguiente tabla:

Metas_Avances	Avance antes de junio del 2020	Avance en el semestre	Meta Final
Resultados	5	0	6
Avances	83	0	83

El valor del avance para la Meta Final = (Avances hasta el semestre anterior + Avances en el semestre) / Meta Final  
Avances = el valor esta en porcentaje

---

**Dentro del PIR responde al OUTPUT 1.1.1**

Número de publicaciones

Del 2020-07-01 al 2020-12-31 se han hecho 1 publicaciones.

Los avances con respecto a las metas en este período se pueden ver en la siguiente tabla:

Metas_Avances	Avance antes de junio del 2020	Avance en el semestre	Meta Final
Resultados	6	1	3
Avances	200	33	233

El valor del avance para la Meta Final = (Avances hasta el semestre anterior + Avances en el semestre) / Meta Final  
Avances = el valor esta en porcentaje

## Dentro del PIR responde al OUTPUT 1.1.2

### Protocolo diseñado, aprobado y adoptado

El protocolo se puede encontrar en alguna de las siguientes fases:

- 0 - Protocolo no existe
- 1 - Protocolo en diseño
- 2 - Protocolo diseñado
- 3 - Protocolo aprobado
- 4 - Protocolo adoptado

Como se puede ver en la siguiente tabla, la meta se cumplió en el 2019:

FASES	INICIO	FIN	Valor
Protocolo en diseño	2018-08-01	2019-01-31	25
Protocolo diseñado	2019-01-01	2019-07-31	50
Protocolo aprobado	2019-08-01	2019-01-31	75
Protocolo adoptado	2019-09-18	2019-10-30	100

La columna Valor está en porcentaje

Este OUTPUT ya se cumplió desde el año 2019 al 100%.

Línea base	Meta intermedia	Meta final
Protocolo no existe	Sistema diseño/diseñado	Sistema implementado/adoptado

## Dentro del PIR responde al OUTPUT 1.1.2

### Fase de desarrollo del SIAgroBD

El SIAgroBD se puede encontrar en alguna de las siguientes fases:

- 0 - SIAgroBD no existe
- 1 - SIAgroBD en diseño
- 2 - SIAgroBD diseñado
- 3 - SIAgroBD aprobado
- 4 - SIAgroBD adoptado

Del 2020-07-01 al 2020-12-31 se han el SIAGRO se implementó y adoptó. Por lo que ya se cumplió con la meta al 100% como se muestra en la siguiente tabla:

Metas_Avances	Avance antes de junio del 2020	Avance en el semestre	Meta Final
Resultados	6	1	3
Avances	200	33	233

El valor del avance para la Meta Final = (Avances hasta el semestre anterior + Avances en el semestre) / Meta Final  
Avances = el valor esta en porcentaje

Desde el semestre anterior se terminaron las fases de prueba y se implementó y adoptó que corresponden a un avance del 75 y 100%. Los avances con respecto a las metas se ve en la siguiente tabla:



## Grupos de Enfoque en San Andrés Tzirondaro, Michoacán

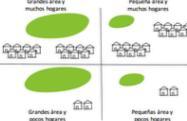
La Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Comabio), a través del proyecto Agrobiodiversidad mexicana y con la colaboración de la UNAM esta cuantificando y valorando la diversidad agrícola de Michoacán. Para cumplir con este objetivo se están llevando a cabo distintas estrategias, entre las que se encuentran los Grupos de Enfoque.

### ¿Qué son los Grupos de Enfoque?

Los Grupos de Enfoque son reuniones entre agricultores, hombres y mujeres, donde se discute y conversa sobre un tema de interés, en este caso sobre las plantas usadas para la alimentación de una localidad o comunidad. La metodología se basa en cuatro preguntas:

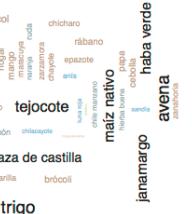
1. ¿Cuales son las plantas, para la alimentación, que se cultiva o recolectan en su comunidad?
2. De estas plantas, ¿cuales son consumidas en la comunidad?
3. De estas plantas, ¿cuales se compran en la comunidad?
4. De estas plantas, ¿cuales se venden en la comunidad?

A cada pregunta las plantas se van colocando en los diferentes cuadrantes de acuerdo a su área de siembra o recolecta y a su uso por los habitantes de la comunidad como se muestra en la siguiente figura:



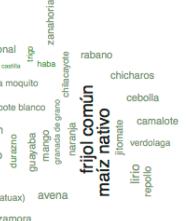
En julio del 2019 en la comunidad de San Andrés Tzirondaro, se llevó a cabo el Grupo de Enfoque con la participación de 9 hombres y 16 mujeres.

Como parte de los resultados de San Andrés Tzirondaro se tuvo un total de 112 especies reportadas por su nombre común. En el grupo de los hombres se tuvo un total de 82 especies que se muestran en la siguiente nube de palabras, donde las plantas o cultivos con letras más grandes son aquellas que fueron sembradas o recolectadas en grandes extensiones por muchos agricultores:



Respondiendo a las cuatro preguntas mencionadas previamente, en el grupo de los hombres las plantas o cultivos más producidos fueron capulín, tejocote, higo, calabaza de castilla, frijol común, maíz nativo, nopal de castilla, alfalfa, aveña, haba verde, janamargo, trigo; las más utilizados en el autocultivo (o consumo de sus animales) fueron capulín, tejocote, higo, calabaza de castilla, frijol común, maíz nativo, nopal de castilla, alfalfa, aveña, haba verde, janamargo, trigo; las que se compran principalmente fueron chabacano, chirimoya, ciruela, durazno, membrillo, nispero, pera, pitaya, granada, higo, limón, maracuya, naranja, naranja lima, calabacita, chayote, chile manzano, frijol común, joconal, maíz nativo, nopal de castilla, tomate verde, hierba buena, jitomate, rábano, ruda, sandia, alfarroba, brócoli, cebolla, chícharo, cilantro, col, colliflor, fresa, garbanzo, haba verde, janamargo, lenteja, papá, pepino, trigo, zanahoria; y, finalmente, las que se venden más fueron durazno, limón, calabaza de castilla, chayote espinoso, chile manzano, maíz nativo, frijol común, alfalfa, haba verde.

En el grupo de las mujeres se tuvo un total de 71 especies que se muestran en la siguiente nube de palabras, donde las plantas o cultivos con letras más grandes son aquellas que fueron sembradas o recolectadas en grandes extensiones por muchos agricultores:



en cuanto a las preguntas mencionadas previamente, en el grupo de las mujeres las plantas o cultivos más producidos fueron frijol común, maíz nativo; las más utilizados en el autocultivo (o consumo de sus animales) fueron frijol común, maíz nativo; las que se compran principalmente fueron capulín, muez, pitaya, ciruela, durazno, granada de grano, granada moqui, guayaba, higo, limón, mango, naranja, plátano, calabaza de castilla, chile, frijol común, maíz nativo, tomate verde, alfalfa, alfarroba, anís, chilacayote, haba, jitomate, joconal, cebolla, chícharos, cilantro, repollo, trigo; y, finalmente, las que se venden más fueron maíz nativo. Finalmente, se muestran algunas fotos del evento en San Andrés Tzirondaro



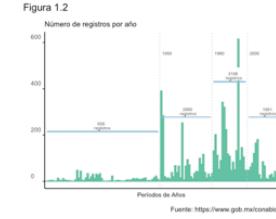
# Mi experiencia con R Markdown



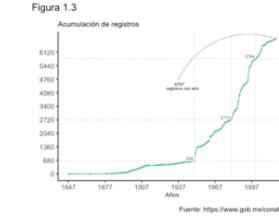
- Reportes internos de Trabajo
- Reportes para agricultores de los resultados obtenidos
- Análisis de datos



Fuente: <https://www.gob.mx/conabio>



Fuente: <https://www.gob.mx/conabio>



Fuente: <https://www.gob.mx/conabio>

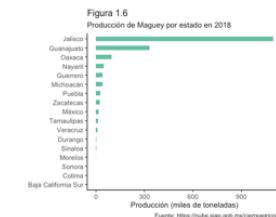


Fuente: <https://www.gob.mx/conabio>



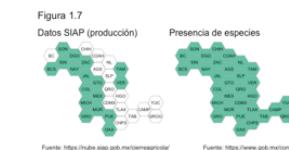
Fuente: <https://www.gob.mx/conabio>

**1.4 Datos oficiales del SIAP**  
La producción de agave para el 2018 se concentró principalmente en Jalisco con 109 mil toneladas, seguidos por Guanajuato y Oaxaca (Figura 1.6).



Fuente: <https://hub.siaip.gob.mx/mexmagueytotal>

**1.5 SIAP versus Conabio**  
En la siguiente figura se puede ver los estados donde se produce magueyes o agaves (figura 1.7 a mano izquierda), aunque la distribución de las 196 especies de magueyes se encuentra en todo el país (figura 1.7 a mano derecha).



Fuente: <https://hub.siaip.gob.mx/mexmagueytotal>

Fuente: <https://www.gob.mx/conabio>



Fuente: <https://www.gob.mx/conabio>

**2.2 Registros a través de los años**  
Se tienen registros desde el año 1853 y el período con mayor cantidad de registros fue de 1980 al 2000. Hoy en día se tiene un total de 2386 registros para las 40 especies registradas (Figuras 2.2 y 2.3).

# ¡Inspiration!



History of Teacup Giraffes

Modules ▾

About the Authors



Welcome to the Wonderful World of

## TEACUPS, GIRAFFES, & STATISTICS

A delightful series of modules to learn statistics and R coding for students,  
scientists, and stats-enthusiasts.

<http://bit.ly/giraffe-stats>

Github: <https://github.com/tinystats/teacups-giraffes-and-statistics>

# Mi experiencia con R Markdown



- Manual de usuario de IDEAMEX
- Notas de cursos cortos de Bioinformática
- Análisis de datos

IDEAMEX: Integrative Differential Expression Analysis for Multiple EXperiments

User's Guide

Verónica Jiménez-Jacinto  
Alejandro Sánchez-Flores  
Leticia Vega-Alvarado

Last revised February 2019

<b>Contents</b>	
1 Introduction	2
1.1 Citation	2
2 Accessibility	2
3 Starting an analysis	3
3.1 Raw count table format	3
3.1.1 Sample names format	4
3.2 Uploading data file	4
3.3 Methods selection	4
3.4 Samples names validation and edition	5
3.5 Comparison definition	6
3.5.1 Setting cutoffs parameters	6
3.6 Submit the analysis	6
4 Data Analysis	7
4.1 Pseudo-counts	7
4.2 Plots	8
4.2.1 Barplot	8
4.2.2 Density plots	8
4.2.3 Count Per Million plot	8
4.2.4 Principal Components Analysis plot	8
4.2.5 Multi-Dimensional Scaling plot	10
5 Differential Expression Analysis	10
5.1 Raw data filtering	11
5.1.1 Count Per Million	11
5.2 Evaluation of the differential expression	11
5.3 Output Plots	11
5.3.1 Expression plot	11
5.3.2 MA plot	12
5.3.3 MD plot	12

Table 3: Plots				
Plot	edgeR	limma	NOISeq	DESeq2
Expr	X	X	Yes	X
MA	X	X	X	Yes
MD	X	Yes	Yes	X
MDS	Yes	Yes	X	X
PCA	X	X	Yes	Yes
Smear	Yes	X	X	X
Volcano	Yes	X	X	X

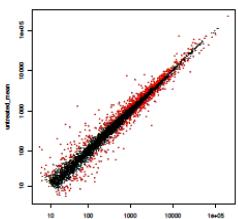


Figure 12: Expression plot

## 5.3.2 MA plot

This plot represents each gene with a dot. The x axis is the average expression over the mean of normalized counts, the y axis is the log2 fold change between conditions. Features declared as differentially expressed are highlighted in red (Gonzalez 2014). (see Figure 13)

## 5.3.3 MD plot

The mean-difference (MD) plots show average expression (mean: x-axis in limma or D for NOISeq) against log-fold-changes (difference: y-axis in limma or M for NOISeq). Features declared as differentially expressed are highlighted. (see Figure 14).

## Análisis de resultados de Qiime2 usando el paquete qiime2R

Leticia Vega Alvarado \*<sup>1</sup>

<sup>1</sup>Instituto de Ciencias Aplicadas y Tecnología - UNAM

[leticia.vegacit@unam.mx](mailto:leticia.vegacit@unam.mx)

23 Oct 2020

## Índice

Objetivo	3
1 Introducción	3
1.1 ¿Qué es qiime2R?	3
1.2 Instalando qiime2R.	3
1.2.1 Iniciando RStudio desde el servidor del curso.	3
2 Leyendo archivos .qza	3
2.1 Leyendo la tabla de abundancias	4
2.2 Leyendo el archivo de taxonomía	5
3 Compactando la tabla de abundancias y generando gráficas	6
3.1 Compactando la tabla de abundancias a nivel género	6
3.2 Generando gráficas de barras a un nivel género	6
3.3 Creando el heatmap de abundancia a nivel de género	8
4 Gráficas de pie a nivel phylum	9
4.1 Gráficas de pie con ggplot	9
5 Manipulación de [Dataframes] con dplyr	11
5.1 ¿Qué es dplyr?	11
5.2 Gramática de dplyr	11
5.3 La función select()	12
5.4 La función filter()	12
5.5 La función	12

### 1 Objetivo

#### 2 Introducción

#### 3 Métodos para el análisis de Expresión diferencial

#### 4 Pipeline para el análisis de expresión diferencial partiendo de una tabla de conteos.

#### Referencias

## Análisis de resultados de Qiime2 usando el paquete qiime2R

### Objetivo

Tener una visión general de qué es el paquete qiime2R, para qué sirve y cómo instalarlo, así como conocer algunas de sus funciones principales (lectura de archivos, compactar datos por nivel, gráficas de barras y de tipo "pie"), con la finalidad de que los asistentes puedan utilizarlos en el manejo y análisis de datos de metagenómica. Además, proporcionar los conceptos generales del paquete dplyr para el manejo de datos.

### 1 Introducción

#### 1.1 ¿Qué es qiime2R?

qiime2R es un paquete desarrollado por Jordan Bisanza, Anders Kiledal y Rodrigo O. Polo; y sirve para exportar los "artifacts" (objetos que genera qiime2) a R. Los datos leídos se devuelven como un data.frame, lo cual permite manejárselos y analizarlos con la amplia gama de funciones, para análisis y visualización, que tiene R. Además, qiime2R ofrece funciones para compactar datos a un nivel taxonómico, generar gráficas de barras, heatmaps de abundancia, entre otras. Este paquete se encuentra todavía en desarrollo.

### 1.2 Instalando qiime2R

qiime2R está disponible actualmente a través de github, y se puede instalar en R mediante el siguiente comando:

```
# (si requiere namespaces("devtools", quietly = TRUE))  
install.packages("devtools")  
devtools::install_github("jbisanza/qiime2R")
```

#### 1.2.1 Iniciando RStudio desde el servidor del curso.

Recordemos que para nuestro taller trabajaremos en RStudio desde un servidor. Para acceder a RStudio debemos:

1. Ingresar a <http://122.248.220.22:8787>
2. Proporcionar nuestro usuario y password, y de manera automática se abrirá el ambiente de trabajo de RStudio.

### 2 Leyendo archivos .qza

## Análisis de expresión diferencial

Leticia Vega-Alvarado, Dra.

Instituto de Ciencias Aplicadas y Tecnología, UNAM.

[leticia.vegacit@unam.mx](mailto:leticia.vegacit@unam.mx)

26 Mar 2020

### 1 Objetivo

Brindar los conceptos básicos, de los paquetes edgeR, DESeq2 y NOISeq (del proyecto Bioconductor), que nos permiten comparar datos de RNASeq provenientes de diferentes condiciones, para detectar aquellos genes que se expresan con una diferencia estadísticamente significativa.

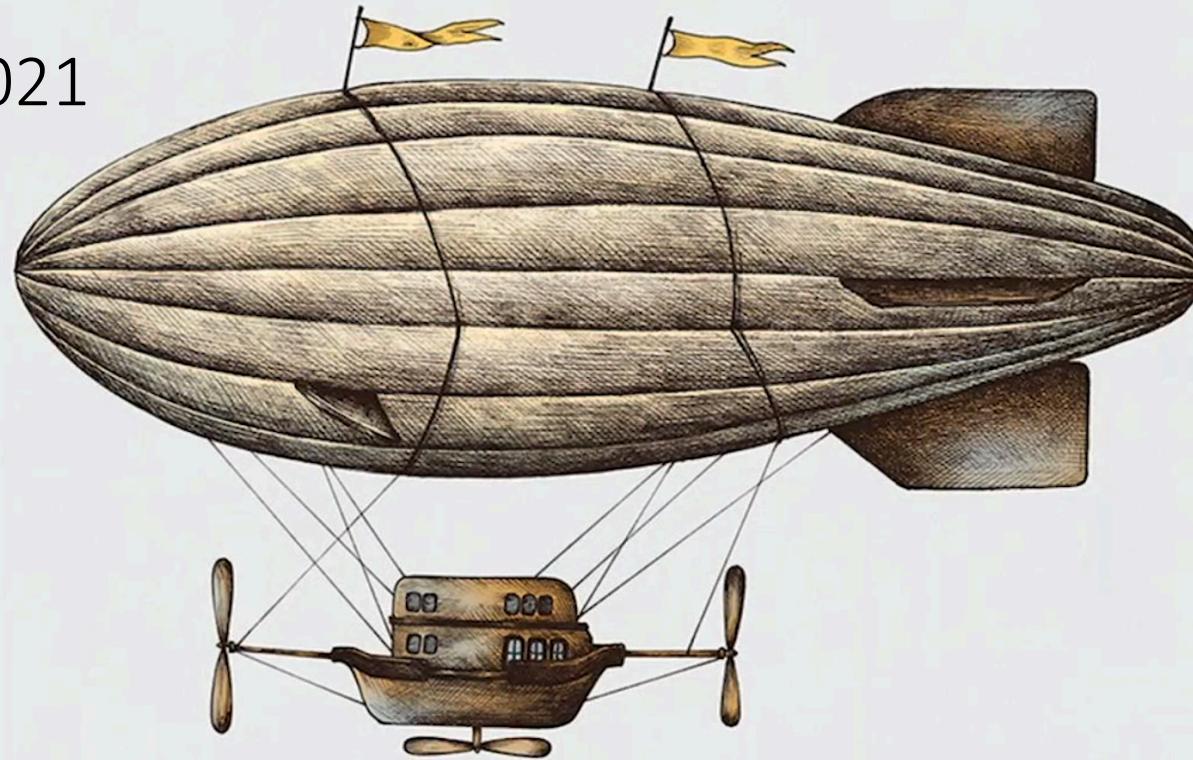
### 2 Introducción

Un organismo puede tener miles de genes, algunos de estos siempre están activos y son necesarios para el uso constante del organismo. Sin embargo, existen otros genes cuya expresión se "apaga" o "prende" (como una serie de luces navideñas) en función de ciertas condiciones ambientales o funcionales.



# Fuentes y ejemplos

RStudio::conf 2021



## Sharing on Short Notice

HOW TO GET YOUR TEACHING MATERIALS ONLINE WITH R MARKDOWN

Alison Hill · Desirée De Leon

[rstd.io/sharing](https://rstd.io/sharing)

<https://rstudio.com/resources/webinars/sharing-on-short-notice-how-to-get-your-materials-online-with-r-markdown/>

# Fuentes y ejemplos

- <http://bit.ly/giraffe-stats>
- <https://education.rstudio.com/>
- <https://bookdown.org/yihui/blogdown/>
- <https://stat545.com/>
- <https://conf20-intro-ml.netlify.com/>
- <https://summer-of-blogdown.netlify.com/>
- <https://apreshill.github.io/data-vis-labs-2018>

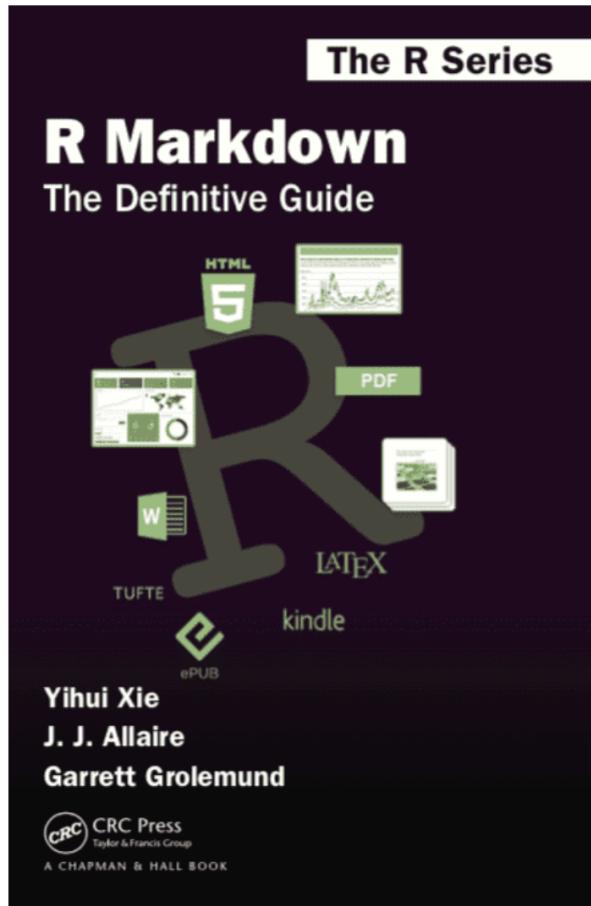
# Proceso de crecimiento de R Markdown



<https://rstudio.com/resources/webinars/sharing-on-short-notice-how-to-get-your-materials-online-with-r-markdown/>

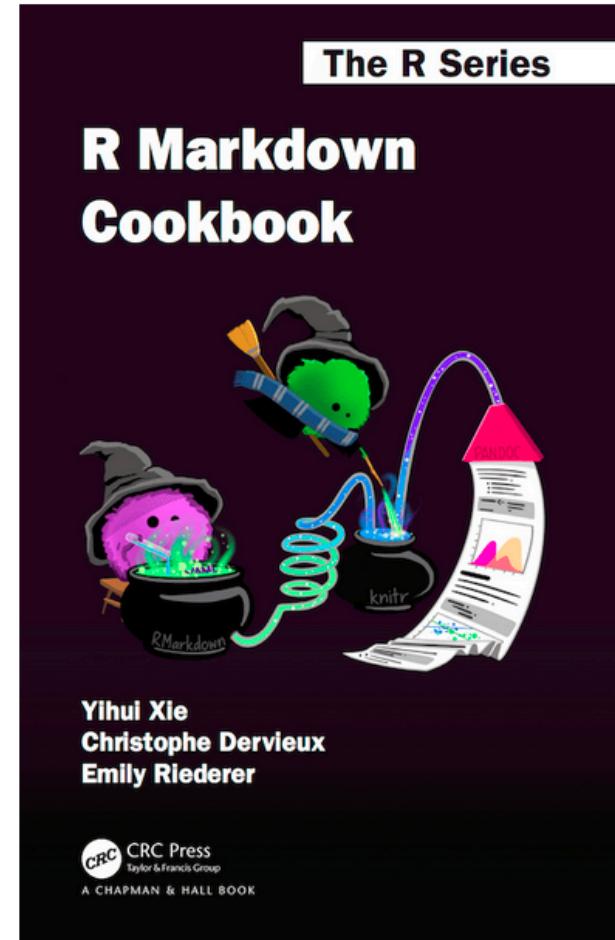
# Literatura del curso

## R Markdown



<https://bookdown.org/yihui/rmarkdown/>

## R Markdown Cookbook



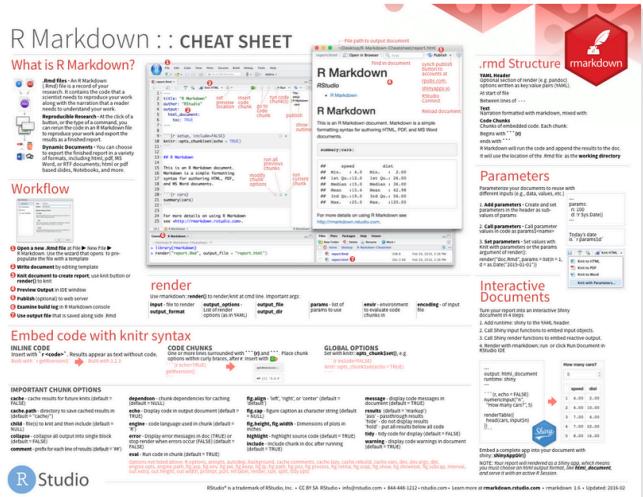
<https://bookdown.org/yihui/rmarkdown-cookbook/>

# Literatura del curso

## R Markdown Cheatsheet

R Markdown is an authoring format that makes it easy to write reusable reports with R. You combine your R code with narration written in markdown (an easy-to-write plain text format) and then export the results as an html, pdf, or Word file. You can even use R Markdown to build interactive documents and slideshows. Updated February 16. ([Old Version](#).)

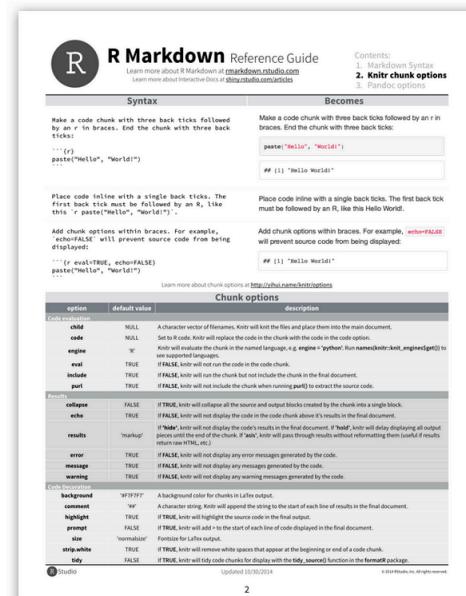
[DOWNLOAD](#)



## R Markdown Reference Guide

R Markdown marries together three pieces of software: markdown, knitr, and pandoc. This five page guide lists each of the options from markdown, knitr, and pandoc that you can use to customize your R Markdown documents. Updated October 14.

[DOWNLOAD](#)



# Rstudio:: Cheatsheets

<https://rstudio.com/resources/cheatsheets/>

# Partes del R Markdown

The screenshot shows an RStudio interface with an R Markdown file named 'index.Rmd' open. The file contains YAML metadata at the top, followed by R code chunks and a list of goals. Handwritten annotations with colored curly braces group the code into three categories:

- metadata = YAML**: Groups the YAML configuration at the top of the file.
- code = R**: Groups the R code chunks (lines 11-17).
- text = markdown**: Groups the explanatory text and lists (lines 19-31).

The RStudio environment includes a Global Environment panel showing an empty environment, a File browser panel listing the project files, and a Console panel at the bottom.

```
1 ---  
2 title: "Lab 02: MoMA Museum Tour"  
3 subtitle: "CS631"  
4 author: "Alison Hill"  
5 output:  
6   html_document:  
7     theme: flatly  
8     toc: TRUE  
9     toc_float: TRUE  
10 ---  
11 ```{r setup, include = FALSE, cache = FALSE}  
12 knitr::opts_chunk$set(  
13   comment = NA,  
14   warning = FALSE,  
15   message = FALSE,  
16   fig.path = "moma-figs/")  
17 ...  
18  
19 # Goals for Lab 02  
20  
21 - Review `dplyr` functions learned in last lab and DataCamp course  
22 - Practice using `dplyr` functions to get to know a new dataset  
23 - Map global plot aesthetics to variables in `ggplot2`  
24 - Create faceted plots with `ggplot2`  
25  
26 # Inspiration + data  
27  
28 We'll use data from the Museum of Modern Art (MoMA)  
29  
30 - Publicly available on [GitHub](https://github.com/MuseumofModernArt/collection)  
31 - As analyzed by [fivethirtyeight.com](https://fivethirtyeight.com/features/a-new-moma-data-set-is-now-available/)
```