



# **Dia 1. Introducción al R básico**

¿Qué es R?

**Principios de la investigación reproducible.**

**Tipos y estructuras de datos en R.**

---

Universidad de Guadalajara, CUCSH–CUCBA  
Viacheslav Shalisko

16–24.07.2019

One of the core principles of the scientific process is that other scientists are able to repeat your experiment and either confirm or refute your results.

This is referred to as **reproducibility** or **replication**.





# Investigación reproducible / repetible

WORLD VIEW • 24 MAY 2018

<https://www.nature.com/articles/d41586-018-05256-0>

## Before reproducibility must come preproducibility



*Instead of arguing about whether results hold up, let's push to provide enough information for others to repeat the experiments, says Philip Stark.*

“Just as I have pledged not to review papers that are not preproducible, I have also pledged not to submit papers without providing the software I used, and — to the extent permitted by law and ethics — the underlying data. I urge you to do the same.”

SCIENCE  
SHOULD BE  
**‘SHOW ME’,**  
NOT  
**‘TRUST ME’.**

# Investigación reproducible / repetible

Hacer *investigación reproducible* es poder permitir a ti mismo y a los demás repetir y obtener los mismos resultados de un trabajo científico.

Para esto es necesario que el producto de la investigación incluya:

- Publicación en forma de un artículo científico, un preprint o tesis, etc.
- Descripción completa de materiales y métodos
- Datos utilizados
- Código de cómputo utilizado
- Versión del software utilizado
- Cualquier otra información necesaria para repetir los experimentos y análisis

# Investigación reproducible / repetible

## ¿Reproducible para quién?

1. Quién hizo la investigación, 6 meses (o 6 días, o algunos años) después

### *Evitar:*

- ¿Cómo, por Dios, era esto?
- El archivo bueno es final.xlsx. No, espera, tal vez sea final\_elbueno.xlsx, o final\_3\_revisado.xlsx. Deja reviso la última fecha de modificación. Hum....
- Ahorita me acuerdo

### *Ganar:*

- Volver a correr fácilmente los mismos datos con pequeñas modificaciones a los parámetros o datos de entrada
- Reutilizar código o datos para nuevos análisis
- Confianza en tus resultados



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc



# Investigación reproducible / repetible

## ¿Reproducible para quién?

### 2. Tus colegas y asesor/a

*No temer escuchar:*

- ¿Me ayudas a hacer este análisis?
- ¿Puedes explicarme como obtuviste esto?
- ¿Puedes repetir esta gráfica agregando este dato nuevo?

### 3. La comunidad científica

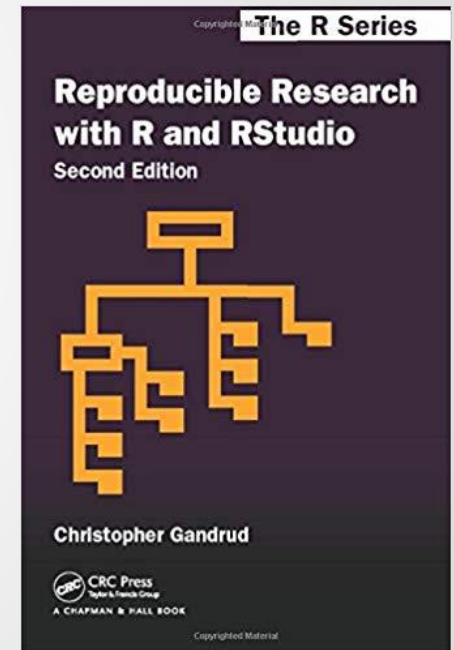
- Corroborar resultados
- Construir sobre lo construido
- Aportar al futuro de la Humanidad



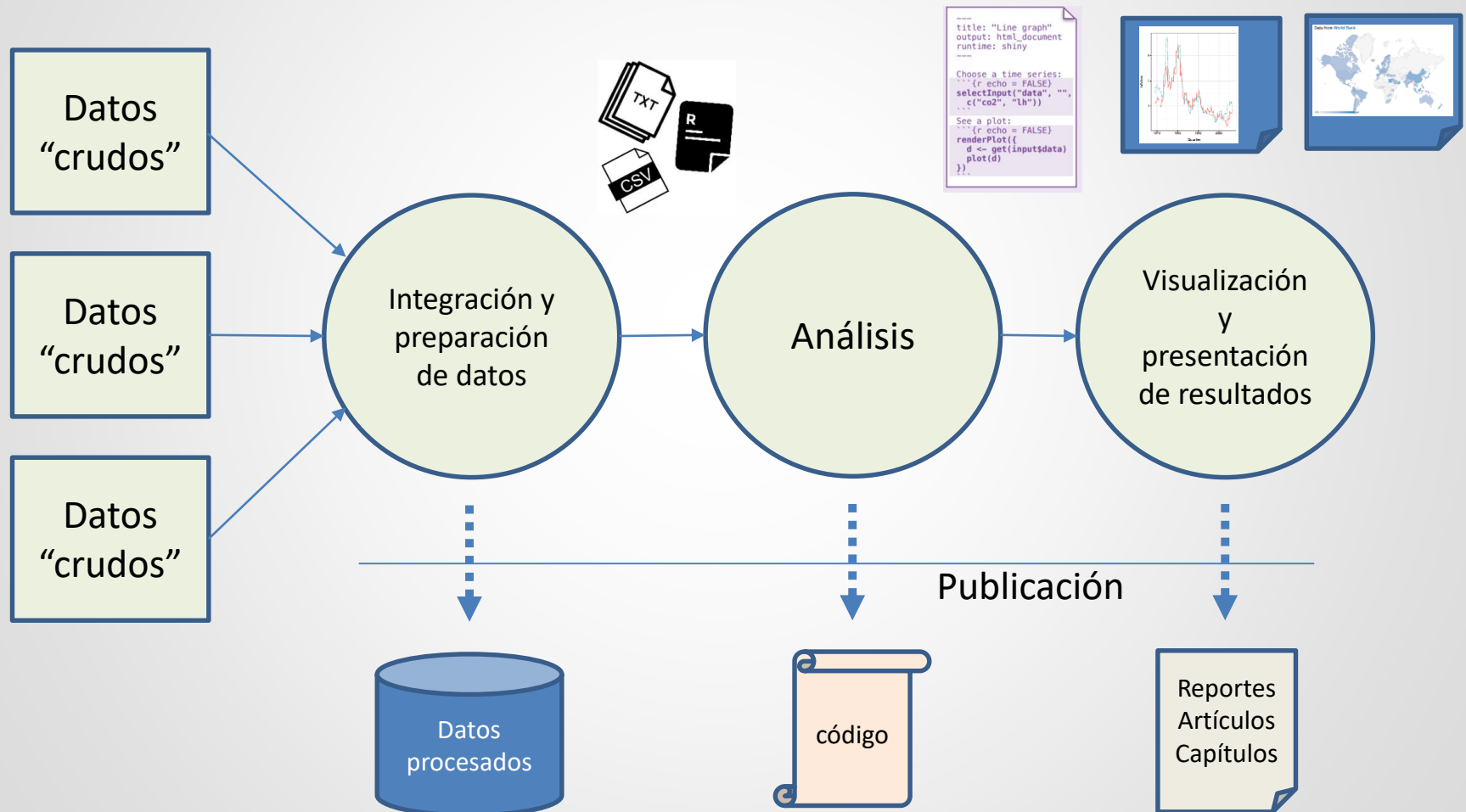
# Investigación reproducible / repetible

Los “tips” prácticos para realizar la investigación y análisis de datos reproducibles (Gandrud 2015)

- Documenta todo
- Todo flujo de trabajo debe estar reflejado en los archivos de texto
- Todos archivos con análisis deben ser legibles por una persona
- Vincula los archivos de datos con el análisis de forma explícita
- Planifica desde principio como vas a organizar, almacenar y publicar tus archivos

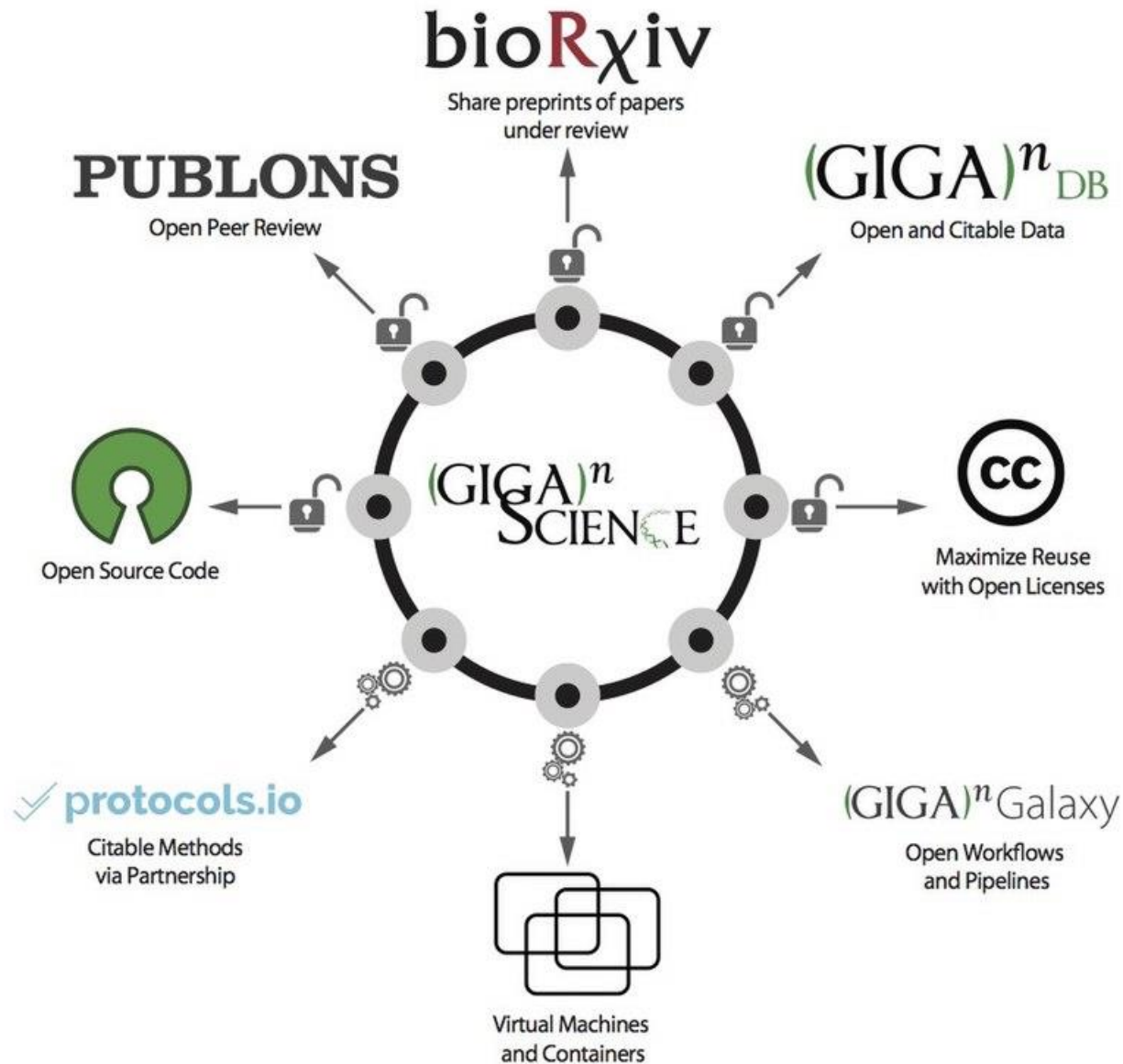


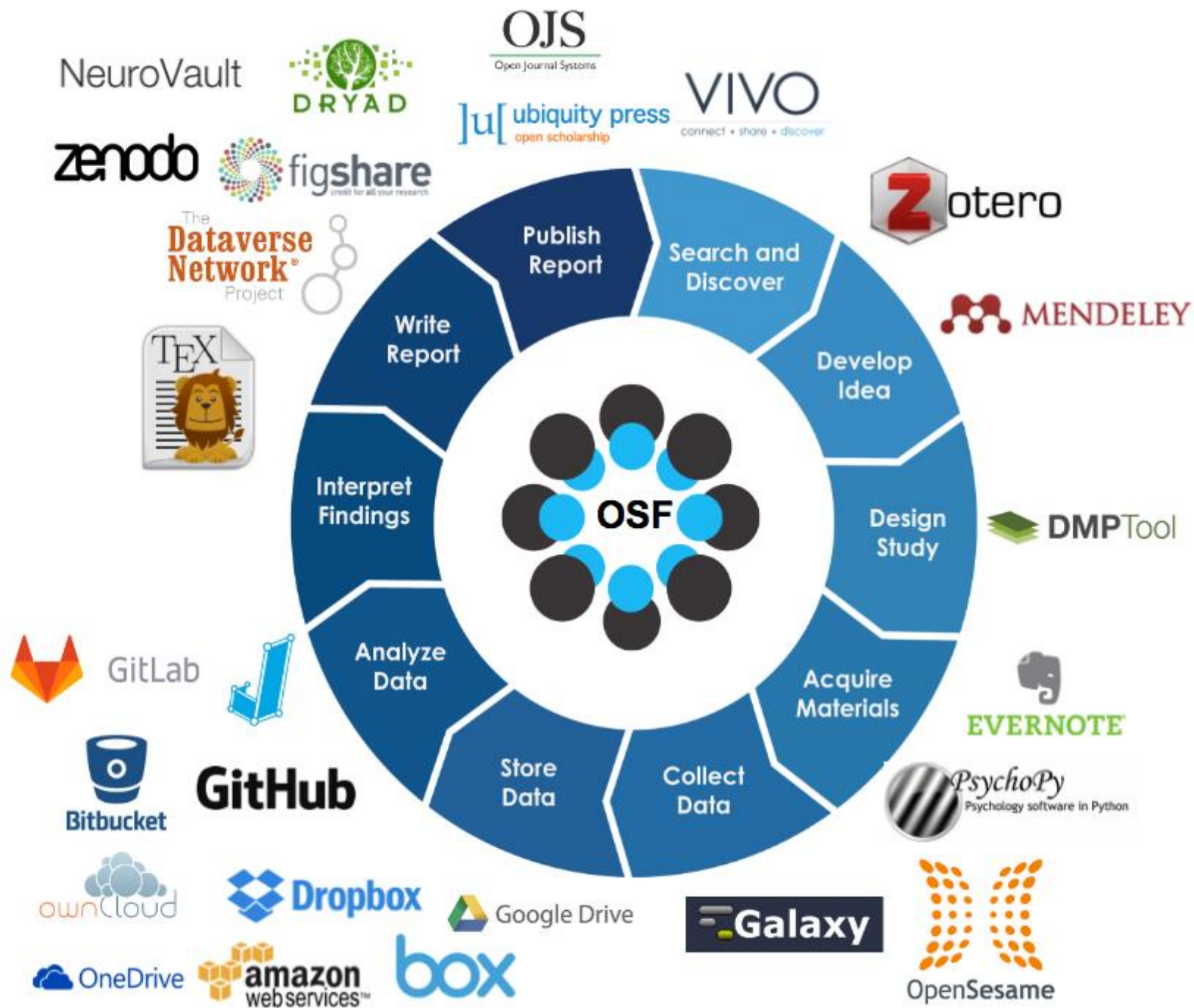
# Investigación reproducible / repetible





GigaScience integrates and publishes all research objects  
to maximize reproducibility, transparency and reuse





# Investigación reproducible / repetible

## Excusas comunes para no compartir nuestro código

- Me da pena que vean mi código
- No quiero que otros saquen provecho de mi código, me pertenece o a mi institución
- Otros no publican su código ¿por qué yo sí?
- Me da flojera pulir mi código para publicarlo
- Si publico mi código le van a encontrar errores y demandar correcciones o ayuda

Si respondiste sí (o tus colaboradores) a cualquier de los puntos anteriores checa esta lectura recomendada:

***Publish your computer code: it is good enough*** de Nick Barnes

<https://www.nature.com/news/2010/101013/full/467753a.html>

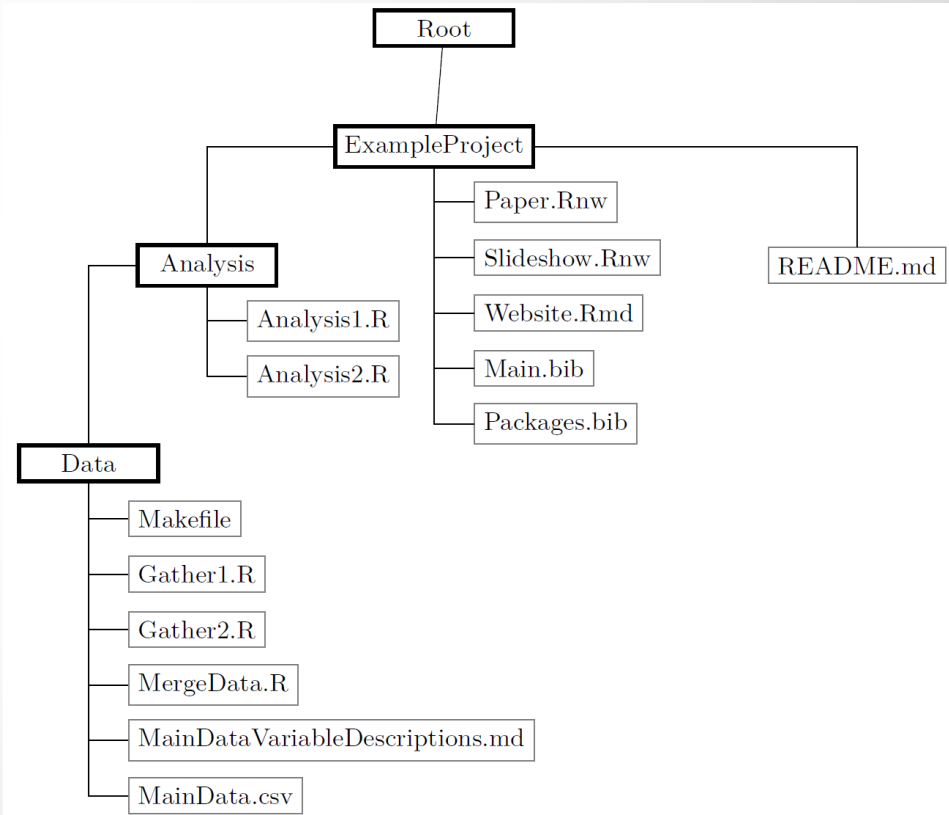


# Investigación reproducible / repetible

## ¿Cómo compartir código?

- Scripts comentados y con un README
- En lo posible utilizar formato Markdown
- Los archivos del proyecto organizados en una estructura transferible
- Crear repositorio de código
- DataDryad (como parte del repositorio de datos)
- GitHub (mejor para funciones y proyectos que continuarán actualizándose)

Ejemplo de estructura de catálogos y archivos del proyecto  
(Gandrud 2015)



# Investigación reproducible / repetible

## GitHub

Es un repositorio de código abierto

Utiliza el protocolo

**git** para llevar un sistema de control de versiones



Su símbolo es un gatopulpo

Tiene una interfase  
Web pública

Se puede bajar como  
una aplicación de  
escritorio

Permite escribir/  
revisar código en  
equipo

<https://github.com/>

# Scripts

***Un script*** es una recopilación por escrito de las instrucciones que queremos que la computadora corra, de modo que al tener esas instrucciones cualquiera pueda repetir el análisis tal cual se hizo.

El script consta de dos tipos de texto:

1. ***El código*** (comandos) que queremos que se ejecute, en el orden que queremos que lo ejecute.

Es decir lo mismo que escribiríamos en la Terminal de computadora para hacer un análisis, pero guardado en un archivo de texto que tiene todos los comandos juntos y que podemos abrir para repetir o compartir el análisis.

2. ***Comentarios*** escritos para un ser humano en un lenguaje de humanos, dígase no solo en español/ingles, sino que nos permita entender qué hace el código, qué tipo de información requiere y cualquier otra cosa que una persona cualquiera necesite para poder utilizar el código del script de forma correcta.



# Markdown

¿Que es el Markdown?

*“Markdown es la herramienta de conversión de texto al formato HTML pensada para los quien escriben para la web. Markdown permite escribir utilizando el formato de texto plano fácil para leer y escribir, para después convertirlo al formato XHTML (o HTML) valido estructuralmente.”*

- John Gruber, creador de Markdown

What is Markdown?

*“ Markdown is a text-to-HTML conversion tool for web writers. Markdown allows you to write using an easy-to-read, easy-to-write plain text format, then convert it to structurally valid XHTML (or HTML). ”*

- John Gruber, creator of Markdown

# R Markdown

How it works



R Markdown from  Studio

[Get Started](#)

[Gallery](#)

[Formats](#)

[Articles](#)



Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown.  
Turn your analyses into high quality documents,  
reports, presentations and dashboards.

Consulta la pagina web <http://rmarkdown.rstudio.com> para conocer la ideología R Markdown

# R Markdown

```
---  
title: "Line graph"  
output: html_document  
runtime: shiny  
---  
  
Choose a time series:  
r echo = FALSE  
selectInput("data", "",  
  c("co2", "lh"))  
...  
  
See a plot:  
r echo = FALSE  
renderPlot({  
  d <- get(input$data)  
  plot(d)  
})
```



D:/GoogleDrive/UdeG\_Docencia/CUCBA\_Curso\_R/script\_dia\_3.html

script\_dia\_3.html Open in Browser Find Publish

## Taller de R básico. Dia 3.

*Viacheslav Shalisko*  
19 de octubre de 2016

### Prueba 1

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

El formato *R Markdown* permite elaborar los documentos que contienen el código fuente en R, los resultados de su ejecución y el texto de comentarios.

Inclusión del código junto con los resultados en el mismo documento es parte de la estrategia para realizar la investigación reproducible. La entrega del código fuente de análisis junto con los resultados permite a otros científicos comprender la estructura del procedimiento y repetirlo. El formato es lo suficientemente flexible para habilitar o deshabilitar inclusión del código y de los resultados específicos.



# R Markdown

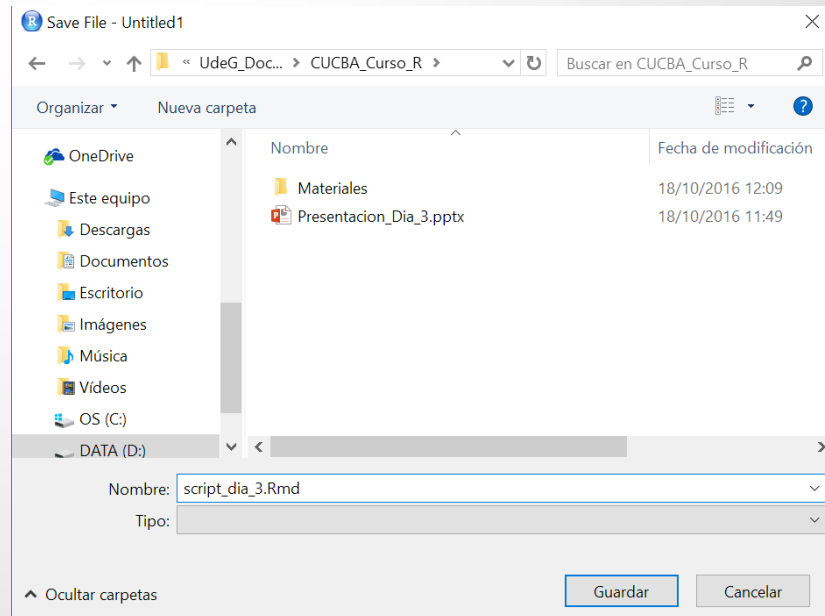
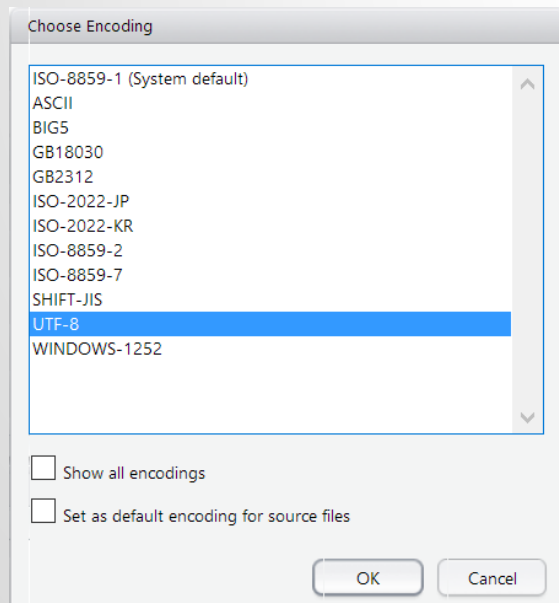
```
---
title: "Line graph"
output: html_document
runtime: shiny
---

Choose a time series:
<<
{r echo = FALSE}
selectInput("data", "",
  c("co2", "lh"))

See a plot:
{r echo = FALSE}
renderPlot({
  d <- get(input$data)
  plot(d)
})
```

## Notas:

- Para poder representar de forma correcta los símbolos especiales (letras con acento, letras del alfabeto griego, etc.) a veces se requiere especificar la codificación UTF-8.
- Guardar el script como el archivo en el formato *R Markdown* con la extensión *.Rmd*
- *Knit* as HTML (o *Knit* as PDF)
- Primera vez que se ejecuta *Knit* puede ser necesario instalar las bibliotecas adicionales (`knitr` y dependencias)



# R Markdown

```
---
title: "Line graph"
output: html_document
runtime: shiny
---

Choose a time series:


```

  (r echo = FALSE)
  selectInput("data", "",
    c("co2", "lh"))

```



See a plot:


```

  (r echo = FALSE)
  renderPlot({
    d <- get(input$data)
    plot(d)
  })

```


```

## Ejemplo de R Markdown

Así se ve el código en el editor de RStudio

El resultado de ejecución con el compilador *Knit*

```
11
12 ### A. Cargar la tabla de datos
13 #### Estructura de datos (tabla `Datos_del_censo.csv`):
14 1. *Centro* - código del CU
15 2. *Especie* - nombre científico
16 3. *Codigo* - identificador único del árbol
17 4. *AB* - área basal del árbol (dm2)
18 5. *DTr* - diametro del tronco equivalente (cm)
19 6. *Alt* - estatura del árbol (m)
20 7. *DCop* - diametro promedio de la copa (m)
21 8. *ExcCop* - excentricidad de la copa
22
23 ```{r}
24 arbolado <- read.csv("Materiales/Datos_del_censo.csv")
25 dim(arbolado)
26 ```
27
28
29
30
31
32
33
```

fragmento del código R ("chunk")

### A. Cargar la tabla de datos

Estructura de datos (tabla `Datos_del_censo.csv`):

1. *Centro* - código del CU
2. *Especie* - nombre científico
3. *Codigo* - identificador único del árbol
4. *AB* - área basal del árbol (dm<sup>2</sup>)
5. *DTr* - diametro del tronco equivalente (cm)
6. *Alt* - estatura del árbol (m)
7. *DCop* - diametro promedio de la copa (m)
8. *ExcCop* - excentricidad de la copa

```
arbolado <- read.csv("Materiales/Datos_del_censo.csv")
dim(arbolado)
```

```
## [1] 4785 8
```

# R Markdown

## R Markdown Cheat Sheet

learn more at [rmarkdown.rstudio.com](http://rmarkdown.rstudio.com)

rmarkdown 0.2.50 Updated: 8/14



Paquete rrticles



Otro Markdown Cheatsheet

<https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

Github's Markdown Guide

<https://help.github.com/en/categories/writing-on-github>



# PROFACAD: Producto 2 para el portafolio

## **Producto para el portafolio: Producto 2. Ejercicio 1.**

Fecha de entrega del producto: 21 de julio 2019

Ejercicio 1: En forma individual realizar diseño de una plantilla en el formato R Markdown que puede servir como base para la presentación del proceso y los resultados del análisis de datos en conformidad con los principios de la investigación repetible. La plantilla debe contener las secciones (comentarios y espacio para código 'code chunk' donde se aplica):

1) Encabezado YAML, 2) Descripción, 3) Parámetros generales, 4) Datos fuente, 5) Análisis y resultados, 6) Exportación de resultados

Producto para el portafolio: Plantilla en el formato R Markdown PDF que determina la estructura de presentación de los procesos y resultados de análisis de datos científico, documento que cumple con las características definidas en las condiciones del ejercicio 1.

# Estructuras de datos en R

## Bibliografía complementaria

Gandrud, C. (2015). *Reproducible research with R and RStudio* (2<sup>nd</sup> ed.). Chapman and Hall/CRC.

Xie, Y., Allaire, J., Golemund, G. (2018) *R Markdown: The Definitive Guide*. Chapman & Hall/CRC  
<https://bookdown.org/yihui/rmarkdown/>

Nüst, D. et al. (2018). *Writing reproducible geoscience papers using R Markdown, Docker, and GitLab*.  
<https://vickysteeves.gitlab.io/repro-papers/index.html>