

# Diseño y Análisis de Datos Cuantiativos (en R)

Introducción al curso

2021-08-08



# Introducción

# Objetivo del curso

Entregar una comprensión general del proceso de investigación social cuantitativa y herramientas que permitan elaborar y ejecutar diseños de investigación utilizando el lenguaje R.

## Curso de múltiples dimensiones:

- ¿es posible medir la realidad social?, ¿que implica?
- ¿cómo medimos y registramos la realidad social?
- ¿cómo se hace esto para el caso chileno?
- ¿cuál es la mejor forma de describir cada tipo de dato cuantitativo?
- herramientas para analizar y reportar lo registrado en las bases de datos (R)

# Programa

El [programa](#) se divide en cuatro puntos:

- Epistemología y consideraciones generales sobre la ciencia
- Diseños de investigación social
- Tipos de datos cuantitativos, estructura de los datos y análisis pertinentes (estadísticas)
- Lenguajes y herramientas para el análisis de datos y reporte (R)

# Programa

Se esperan los siguientes resultados de aprendizaje con el curso:

- 
- 
-



# Fundamentos

## La sociología debe aspirar a ser una ciencia (social)

- Sistemática, reproducible, perfectible y explicativa (desmitificadora) de la realidad.
- La sociología debiese aportar a la resolución de problemas sociales. Juicio "práctico" de calidad.
- Hombros de gigantes



# Fundamentos

## La sociología necesita de la ciencia abierta para ser confiable y útil

- Transparentar y registrar cómo y porqué hacemos lo que hacemos; abrir a la comunidad científica y política los productos de las investigaciones; interacciones en base a argumentos y evidencia...

# Fundamentos

En sociología conviven y deben convivir diferentes enfoques.

- El enfoque cuantitativo es solo una opción. Es necesario reconocer su potencial y sus limitaciones.
- Enfoque cualitativo tiene potencia, puede interactuar con el cuantitativo.

# Fundamentos

## Relación de confianza e igualdad de trato con las y los estudiantes

- Con trabajo y motivación se puede llegar a elaborar productos de alta calidad
- opinión en clase es siempre válida y provechosa.
- Se entienden problemas personales (con fundamento se puede ser flexible)

# Fundamentos

## Reconocimiento y respeto por el trabajo e ideas de otros/as

- El plagio debe evitarse a toda costa.

## Apoyo constante en el aprendizaje y compañerismo

- R puede ser muy difícil si es su primera aproximación. Paciencia y apoyo desde los/as más avanzados.

# Sobre la web y contacto

Todo el material necesario para el desarrollo del curso se encontrará en la página web del curso (<https://dadc1ucsh2021.netlify.app/>).

Acá, podrán encontrar el calendario, la bibliografía para descargar, las presentaciones de las clases y las pautas de las tareas y trabajos.

A la web no se subirá ninguna información relativa a los y las estudiantes del curso. Notas y revisiones serán subidas a u-cursos, manteniéndose el anonimato.

El canal "oficial" de comunicación son los correos: [nicolasrattor@gmail.com](mailto:nicolasrattor@gmail.com) y [kevin.carrasco@ug.uchile.cl](mailto:kevin.carrasco@ug.uchile.cl) . Escribir con copia a ambos.

# Metodología

El curso se realizará por medio de la plataforma zoom, donde la metodología será una combinación de clases expositivas y de ejercicios prácticos.

# Evaluaciones

La evaluación final del estudiante se compone de:

- promedio simple de 8 tareas prácticas en R (50%)
- Nota por participación en clases (10%)
- 4 pruebas teórico-prácticas a realizar en horario de clases (24/09, 22/10, 26/11 y 10/12) (40%).

En pruebas solo se evalúa bibliografía obligatoria, lo pasado en clases y algunos problemas prácticos a resolver en R.

Pasa el ramo quien tenga nota 4,0 o superior.

# Herramientas del curso

- Página web <https://dadc1ucsh2021.netlify.app/>
- [RStudio](#) para datos y escritura.
- Buscador de bibliografía: [Google Scholar](#). Otros buscadores más refinadas.
- Gestor bibliográfico (Zotero)
- [Sci-hub](#) y [libgen](#)
- En el curso **no** se utilizará Word ni para la gestión bibliográfica ni para la escritura de textos.
- Tampoco usar excel para el procesamiento y análisis de datos.



# Introducción a R

## Instalación y cuestiones generales

# Introducción a R

R es un lenguaje de programación (1993), creado principalmente para el análisis de datos. Al principio cuesta comprenderlo, pero con tiempo y dedicación se pueden hacer grandes cosas. Por ejemplo, con lenguaje R y un poco de otras ayudas se hacen las presentaciones de este curso o incluso la [página web](#) del mismo.

Gracias a la Entorno de Desarrollo Integrado (IDE) **RStudio** tenemos un software "amigable" que soporta el lenguaje R y nos permite trabajar con éste.

R Funciona con un paquete base al que se le pueden ir agregando más funciones al instalar otros paquetes. Con esto se pueden hacer gráficos elegantes (Ggplot2), elaboración de documentos html, pdf o word (rmarkdown) o análisis estadísticos complejos como modelos factoriales (lavaan).

# Introducción a R

Pero si solo quiero hacer análisis descriptivos de datos, ¿por qué utilizar R y no continuar con Excel, SPSS u otros softwares que ya sabemos ocupar?

- R es gratis, su código es abierto y se encuentra disponible en la mayoría de las plataformas (Windows, Mac OS, Linux).
- Actualmente son bien valorados los profesionales de las ciencias sociales que utilizan este lenguaje.
- Permite registrar procesamientos y analisis estadísticos, posibilitando que nuestra investigación sea "reproducibile". Mi yo del futuro, un supervisor o un colega puede leer nuestro código y volver a correrlo si le facilitamos los insumos.
- Por lo anterior, facilita el trabajo colaborativo (sobre todo al agregar otros *software* como Github).

# Introducción a R

## Ventajas de R (continuación)

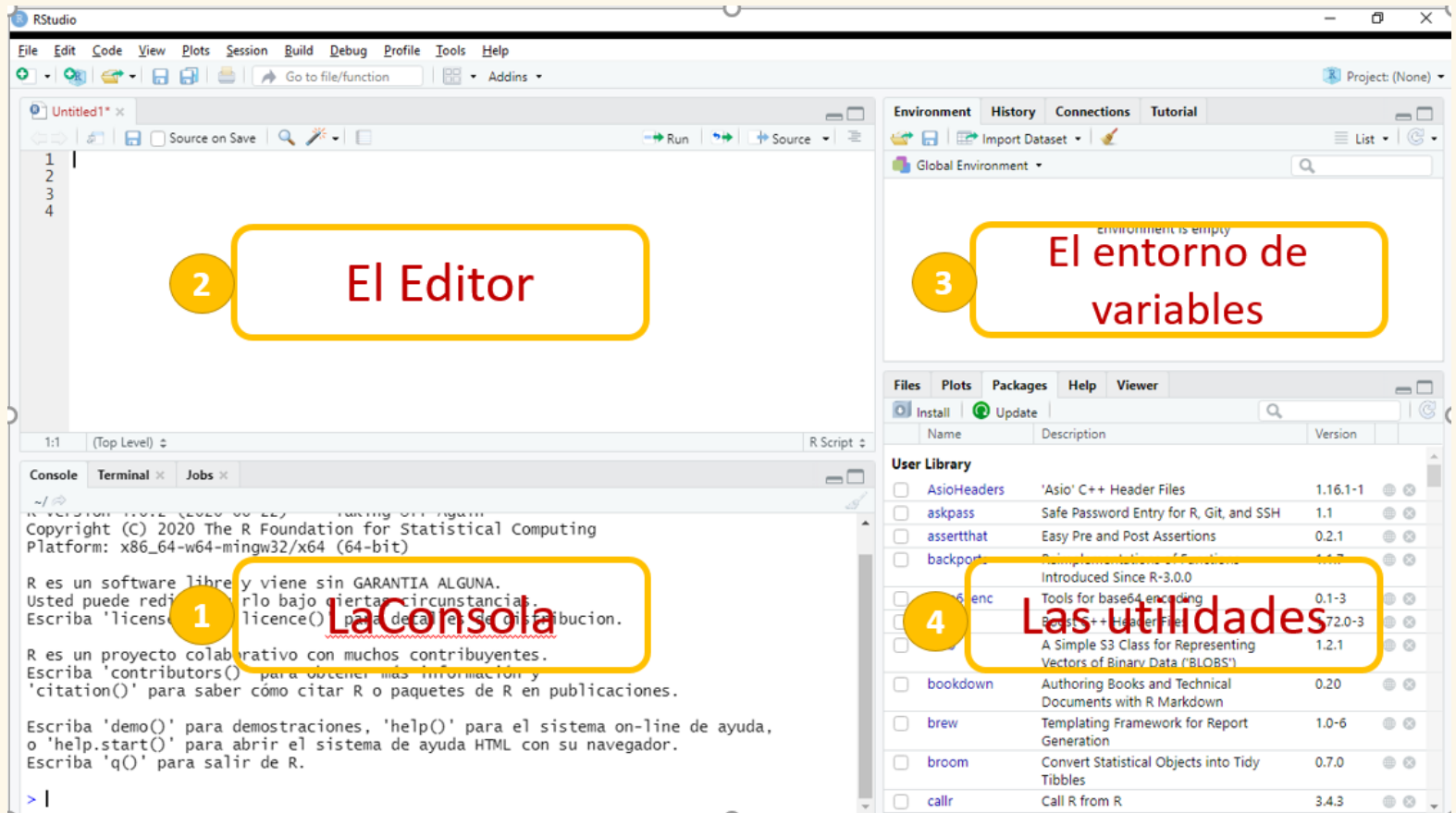
- Tiene una comunidad de usuarios activa, muy dispuesta a ayudar. Por lo general, todas nuestras "panas" ya fueron preguntadas y respondidas en foros-comunidades como [stackoverflow](#).
- Consume pocos recursos informáticos (no te "pega" el computador como otros programas de análisis estadístico).
- Interactúa con otros lenguajes. Se pueden leer bases en spss, sas, stata o excel (hasta se pueden leer pdfs). También se puede exportar bases de datos a esos lenguajes.
- Esto facilita el trabajo colaborativo incluso con quienes no usan R.

# Introducción a R

Debilidades del lenguaje (todas superables):

- Curva de aprendizaje "empinada" al inicio.
- La mayoría de consultas, respuestas, manuales, registros, etc. están en inglés.
- Un problema puede ser resuelto de muchas formas. A veces, sobre todo al inicio, se seleccionan caminos deficientes que terminan siendo incongruentes con otros procesos.
- La comunidad está más centrada en resolver problemas que en el proceso que permite su solución. A veces las soluciones no son las más elegantes y concisas.

# Abramos R



The image shows the RStudio desktop application with four numbered callouts highlighting its main components:

- 1 La Consola**: Points to the console window at the bottom left, which displays the R startup message and instructions.
- 2 El Editor**: Points to the source editor window at the top left, which is currently empty.
- 3 El entorno de variables**: Points to the Environment pane at the top right, which shows that the environment is empty.
- 4 Las Utilidades**: Points to the Packages pane at the bottom right, which lists installed and available packages.

The console window contains the following text:

```
R version 4.0.2 (2020-08-15) -- "Fishing on the Moon"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' para obtener más información.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

The Environment pane shows:

Environment is empty

The Packages pane shows a list of installed and available packages:

Name	Description	Version
asioheaders	'Asio' C++ Header Files	1.16.1-1
askpass	Safe Password Entry for R, Git, and SSH	1.1
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Implementations of Functions Introduced Since R-3.0.0	1.1.7
base64enc	Tools for base64 encoding	0.1-3
bit	Safe C++ Header Files	72.0-3
bit64	A Simple S3 Class for Representing Vectors of Binary Data (BLOBS)	1.2.1
bookdown	Authoring Books and Technical Documents with R Markdown	0.20
brew	Templating Framework for Report Generation	1.0-6
broom	Convert Statistical Objects into Tidy Tibbles	0.7.0
callr	Call R from R	3.4.3

# Interactuando con R

Dos formas:

## 1) A través de la consola.

El símbolo > significa que R está listo para que le entreguemos una instrucción a ejecutar.

Los comandos que se escriban aquí **no** serán grabados para un nuevo inicio de sesión.

El código escrito en consola se ejecuta con la tecla Enter.

# Interactuando con R

## 2) Usando archivos *script*

Una práctica habitual y recomendada es guardar rutinas completas en archivos script. Con esto guardamos el código.

Texto plano que contiene nuestro código, siendo los archivos .R los más comunes.

Piensen estos archivos como un "bloc de notas".

Para ejecutar el código en script puedes utilizar el *shortcut* **Ctrl + Enter**.

El signo gato (#) sirve para "silenciar" códigos o comentar.



# Primeras operaciones con R

R puede ser una calculadora

```
## ¿Cuánto es 5 + 5?
```

```
5 + 5
```

```
## [1] 10
```

```
5 * 5
```

```
## [1] 25
```

```
(400+6783) * 2.7
```

```
## [1] 19394.1
```

# Primeras operaciones con R

Podemos crear objetos y asignarles valores

```
a <- 7  
b <- 3  
a + b
```

```
## [1] 10
```

incluso el resultado lo podemos asignar a un objeto

```
c <- a+b  
c
```

```
## [1] 10
```

Si observamos en nuestro "ambiente", vemos que ahora hay 3 nuevos objetos.

# Primeras operaciones con R

La mayor parte del tiempo usaremos funciones integradas.

Pueden ser entendidas como scripts que funcionan "tras bambalinas".

```
sum(5,10)
```

```
## [1] 15
```

```
# aproximar
```

```
round(9.556789)
```

```
## [1] 10
```

Gran parte de las funciones que utilizamos en R vienen contenidas en paquetes (packages).

Las funciones que acabamos de conocer pertenecen al paquete base, que viene con el software R.

Existen más de 10.000 paquetes en R -> funciones y datos.

# RMarkdown

# Introducción a RMarkdown

R Markdown provee de un marco para generar **documentos** en el contexto del análisis de datos. Estos documentos pueden estar básicamente en formato pdf, html o word.

La idea es "integrar" código de análisis de datos con texto plano (R + Markdown).

En un único documento de RMarkdown podemos:

- Guardar y ejecutar código.
- Generar reportes de alta calidad que pueden ser compartidos fácilmente.
- Los documentos creados con R Markdown son completamente reproducibles.

# Introducción a RMarkdown

La lógica de los archivos `.rmd` es cercana a un bloc de notas, con la excepción de un "encabezado" (*yaml*). Todo su contenido es texto plano, pudiendose agregar "bloques" (*chunks*) de código `.R`

File -> New File -> RMarkdown -> OK -> knit



# RMarkdown

Para ser esquemáticos: los archivos de R Markdown en general tienen 3 partes:

(1). Un encabezado que permite configurar inicialmente el documento que vamos a escribir (**yaml**)

# R Markdown

(2). "Pedazos de código" (o en inglés, *chunk codes*).



# R Markdown

(3). Los cuerpos de texto donde podemos hacer comentarios sin necesidad de usar "#"

```
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF,
15 and MS word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 when you click the knit button a document will be generated that includes both content as well
18 as the output of any embedded R code chunks within the document. You can embed an R code chunk
19 like this:
20
21 {r cars}
22 summary(cars)
23
24
25 ## Including Plots
26
27 You can also embed plots, for example:
28
29 {r pressure, echo=FALSE}
30 plot(pressure)
31
32
33 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R
34 code that generated the plot.
```

# R Markdown

En los cuerpos de texto se pueden agregar de forma simple y rápida propiedades y elementos. Siendo los principales:

## (4). Énfasis

*\*cursiva\**     **\*\*negrita\*\***

## (5). Encabezados

# Capítulo

## Subcapítulo

## (6). Imágenes

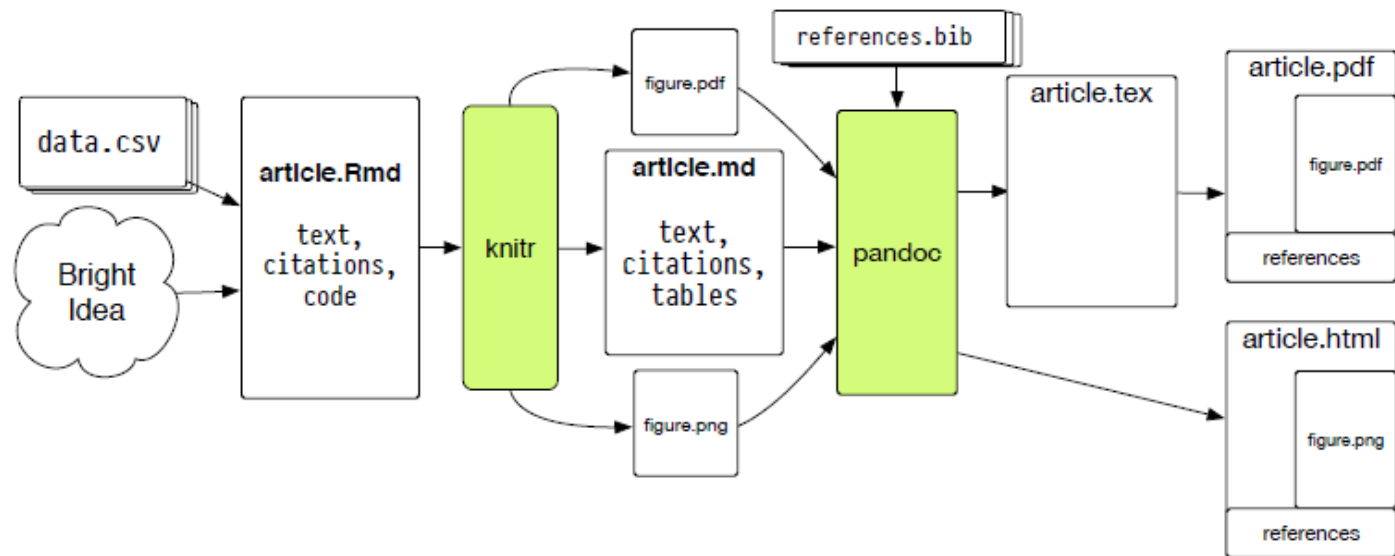
Imágenes en la web o archivos locales en el mismo directorio:

``

# R Markdown

Para generar el documento se puede utilizar el comando render o bien utilizar el botón Knit (Tejer).

¿Como funciona?



# Tarea práctica N°1

Generar en RMarkdown un reporte simple, en formato html, que contenga los siguientes elementos:

- Nombre del estudiante
- presentación de quién escribe (qué cosas te gusta hacer; por qué decidiste estudiar sociología; si es que has trabajado, en qué cosas; que cosas has investigado en trabajos de la universidad; a futuro, en que te ves trabajando)
- Foto de tú sociólogo o cientista social favorito

Adjuntar comprimido con código en `.rmd`, archivos `.html` y foto utilizada. Mandar todo comprimido en un `.zip` o `.rar`.

Entrega a más tardar el **lunes 16 de agosto hasta las 23:59 hrs**, vía correo electrónico

Envíos y consultas a mail profesor y ayudante

# Recursos web utilizados

- Capacitaciones en R del INE.
- R Markdown. Hoja de referencia.
- Xaringan: Presentation Ninja, de Yihui Xie. Para generar esta presentación.

## Para reforzar y seguir aprendiendo

- Introducción a RMarkdown.

## Bibliografía

Wickham, H. (2021). *R Para Ciencia de Datos*. URL: <https://es.r4ds.hadley.nz/>.