



Curso de programación en R

Act. Carlos Ernesto López Natarén



¿Qué es R?

- Lenguaje de programación.
- Ambiente de trabajo.
- Sistema de gráficación.
- En corto, lenguaje para poder manipular el “asset” más importante de una empresa:
¡DATOS!



Objetivos

Al terminar el curso, debemos:

- Entender la estructura del ambiente R
- Ser capaces de meter comandos en R
- Ser capaces de cargar e importar datos en R
- Ser capaces de producir y modificar gráficos en R



Objetivos

- Ser capaces de llevar a cabo análisis con un buen nivel de sofisticación superior a Excel
- Ser capaces de crear nuestras propias funciones en R
- Ser capaces encontrar ayuda si no conocemos alguna función, paquete o característica más avanzada, de manera sistemática y rápida.



¿Por qué usarlo?

- Tiene muchas funciones estadísticas
- Podemos implementar de manera rápida procedimientos estadísticos
- Buenos gráficos “base” y más que excelentes gráficos avanzados
- Los autores de funciones pueden ser contactados directamente



¿Por qué usarlo?

- ¡¡¡Una comunidad activa!!!
- No solo estadística, sino actuarial, ingenieril computacional, bioinformática, financiera, etc, etc, etc!!!
- Cuesta \$0.00 :-)



Los creadores

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS My Account | Welcome, natorro | Log Out | Help

The New York Times Search All NYTimes.com Go

Business Computing

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JOBS | REAL ESTATE | AUTOS

Search Technology Go **Inside Technology** **Bits Blog** **Personal Tech** »
Internet | Start-Ups | Business Computing | Companies | Cellphones, Cameras, Computers and more

Data Analysts Captivated by R's Power



Stuart Isett for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.



intel
Sponsors of Tomorrow.

The 2nd generation Intel® Core™ processors are here.

Get Core
intel inside™



¿Quién está detrás?

- Desde 1997 se mantiene con un equipo llamado el “R Development Core Team”
- Financiado por la R Foundation, non-profit, basada en Viena
- Más información en www.r-project.org



Diferencias con otros

Digamos... SPSS:

- GUI, más reducida en R
- Salida: Menos estadístico, no tablas formateadas
- Sintaxis
- Selección de variables
- Nuevos conceptos: Paquetes, Objetos



R GUI



R Console
Archivo Editar Misc Paquetes Ventanas Ayuda

```
R version 2.14.0 (2011-10-31)  
Copyright (C) 2011 The R Foundation for Statistical Computing  
ISBN 3-900051-07-0  
Platform: i386-pc-mingw32/i386 (32-bit)  
  
R es un software libre  
Usted puede redistribuirlo y/o modificarlo bajo los términos de la Licencia Pública General de GNU.  
Escriba 'license()' o 'licence()' para ver el texto de la licencia.  
  
R es un proyecto colaborativo.  
Escriba 'contributors()' para ver los nombres de los colaboradores.  
'citation()' para saber cómo citar R.  
  
Escriba 'demo()' para ejecutar una demostración.  
o 'help.start()' para abrir el sitio web de ayuda.  
Escriba 'q()' para salir de R.  
  
Loading required package: graphics  
Loading required package: stats  
> utils::arrangeWindows  
> curve(dnorm, -3, 3)  
> |
```

R Graphics: Device 2 (ACTIVE)
Archivo Histórico Redimensionar

The figure shows a plot of the standard normal distribution density function, $dnorm(x)$, against x . The x-axis ranges from -3 to 3, and the y-axis ranges from 0.0 to 0.4. The curve is a symmetric bell shape centered at 0, with a peak height of approximately 0.4.

Sin nombre - Editor
Archivo Editar Paquetes

Windows taskbar: Inicio | R Console | Sin nombre - Editor R | R Graphics: Device 2 ... | ES | 0:22



Está horrible

- Se tiene muy mala interacción
- Se puede uno olvidar fácilmente de qué variables lleva uno
- Las gráficas aparecen en ventanas aparte



R Studio



The screenshot displays the RStudio application window. The top menu bar includes File, Edit, View, Workspace, Plots, Tools, and Help. The main interface is divided into several panes:

- Console:** Shows the R startup sequence, including the version (2.14.0), copyright information, and platform details. It also displays the R license text in Spanish and instructions for using help functions. The user has entered the command `curve(dnorm, -3, 3)` and the plot has been generated.
- Workspace:** Contains buttons for Load, Save, Import Dataset, and Clear All.
- Plots:** Contains buttons for Zoom, Export, and Clear All.

The plot shows a normal distribution curve (bell curve) centered at 0. The x-axis is labeled 'x' and ranges from -3 to 3. The y-axis is labeled 'dnorm(x)' and ranges from 0.0 to 0.4.

```
R version 2.14.0 (2011-10-31)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

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.

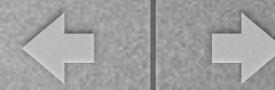
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-
line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su
navegador.
Escriba 'q()' para salir de R.

Loading required package: rcom
Loading required package: rscproxy
> curve(dnorm, -3, 3)
> |
```



Dos aspectos importantes

1. Aprenda a programar en 10 años (Peter Norvig)
2. Código hermoso



Teach Yourself Programming in Ten Years

Peter Norvig

Why is everyone in such a rush?

Walk into any bookstore, and you'll see how to *Teach Yourself Java in 7 Days* alongside endless variations offering to teach Visual Basic, Windows, the Internet, and so on in a few days or hours. I did the following [power search](#) at [Amazon.com](#):

[pubdate: after 1992 and title: days and
\(title: learn or title: teach yourself\)](#)

and got back 248 hits. The first 78 were computer books (number 79 was [Learn Bengali in 30 days](#)). I replaced "days" with "[hours](#)" and got remarkably similar results: 253 more books, with 77 computer books followed by [Teach Yourself Grammar and Style in 24 Hours](#) at number 78. Out of the top 200 total, 96% were computer books.

The conclusion is that either people are in a big rush to learn about computers, or that computers are somehow fabulously easier to learn than anything else. There are no books on how to learn Beethoven, or Quantum Physics, or even Dog Grooming in a few days. Felleisen *et al.* give a nod to this trend in their book [How to Design Programs](#), when they say "Bad programming is easy. *Idiots* can learn it in *21 days*, even if they are *dummies*."

Let's analyze what a title like [Learn C++ in Three Days](#) could mean:

- **Learn:** In 3 days you won't have time to write several significant programs, and learn from your successes and failures with them. You won't have time to work with an experienced programmer and understand what it is like to live in a C++ environment. In short, you won't have time to learn much. So the book can only be talking about a superficial familiarity, not a deep understanding. As Alexander Pope said, a little learning is a dangerous thing.



No es lo mismo esto:



```
#set.seed(10)
n <- 1000
# Lets create a Progress Bar:
pb <- txtProgressBar(min = 0, max = n, style = 3)
zeros <- 1
values <- rep(0, n)
changes <- rep(0, n)
for (j in 1:n) {
  accrued <- rep(0, n)
  accrued[2] <- sample(c(1, -1), 1)
  #To be sure we are looking at a change in sign, we need to
  #be sure that the value we are seeing at time i is different from zero
  # and that the sum of the actual time and two before are zero.
  changes <- rep(0, n)
  for (i in 3:n){accrued[i]<-accrued[i-1]+sample(c(1,-1),1)
  if( (accrued[i] + accrued[i-2]) == 0 && accrued[i] != 0)
  #if(accrued[i]+accrued[i-1]+accrued[i-2] == 0)
  changes[i-2] <- 1
  }
  zeros <- sum(accrued == 0)
  values[j] <- sum(changes)
  setTxtProgressBar(pb, j)
}
close(pb)
# zeros - 1
```



que esto:



```
#set.seed(10)
n <- 1000

# Lets create a Progress Bar:
pb <- txtProgressBar(min = 0, max = n, style = 3)

zeros <- 1
values <- rep(0, n)
changes <- rep(0, n)

for (j in 1:n) {

  accrued <- rep(0, n)
  accrued[2] <- sample(c(1, -1), 1)

  # To be sure we are looking at a change in sign, we need to
  # be sure that the value we are seeing at time i is different from zero
  # and that the sum of the actual time and two before are zero.
  changes <- rep(0, n)
  for (i in 3:n) {
    accrued[i] <- accrued[i-1] + sample(c(1, -1), 1)
    if( (accrued[i] + accrued[i-2]) == 0 && accrued[i] != 0)
      #if(accrued[i]+accrued[i-1]+accrued[i-2] == 0)
      changes[i-2] <- 1
  }
}
```



Temario

1. Introducción
2. Introducción al lenguaje R
3. Programando datos estadísticos
4. Programando con R
5. Simulación



Temario

6. Algebra lineal computacional
7. Optimización numérica
8. Data mashup
9. Graficación con ggplot
10. Excel y R, ¿cómo los integro?



Día 1 - Introducción

- Revisar el temario, ¿algún tema en particular que les interese?
- Instalación de R
- Instalación de R-Studio
- R-Studio, keystrokes, organización
- Indentación
- Algunos comanditos de R



Día 1 - Introducción

La dinámica a seguir:

- Todas las clases, explicaré conceptos nuevos, haremos algunos ejercicios de nivel medio para practicar entre todos
- Se dejará una tarea, esta será para la clase siguiente, si el problema es muy difícil, se resolverá en clase, si no, recibiré preguntas los primeros 15 minutos de la clase.
- Se tratará (porque es experimental) de grabar la clase, tanto lo que digo como lo que aparece en la pantalla.



Bibliografía

Cualquier libro de R de la Serie Use R! de Springer es buena.

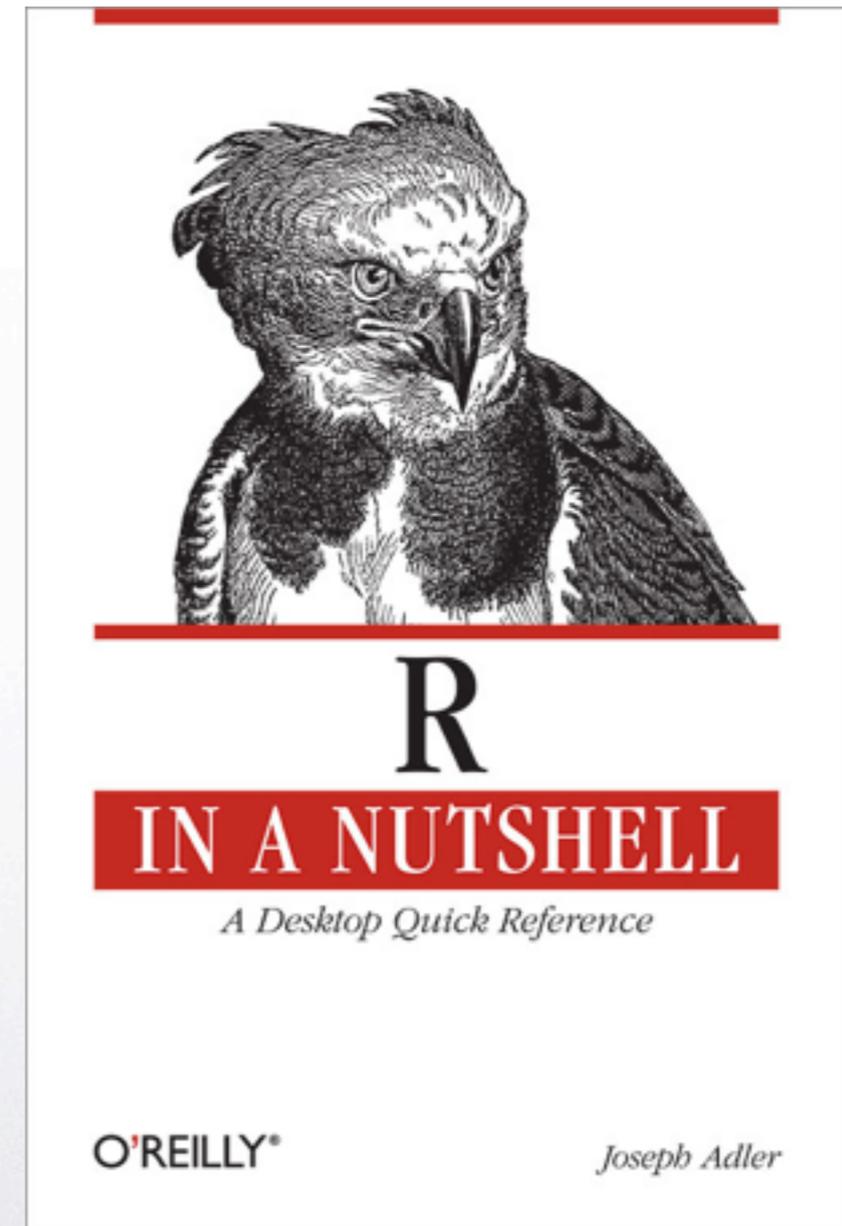
Cualquier libro de R de la serie de Estadística de O'Reilly también es bueno.

Tendremos como base los siguientes:



Bibliografía

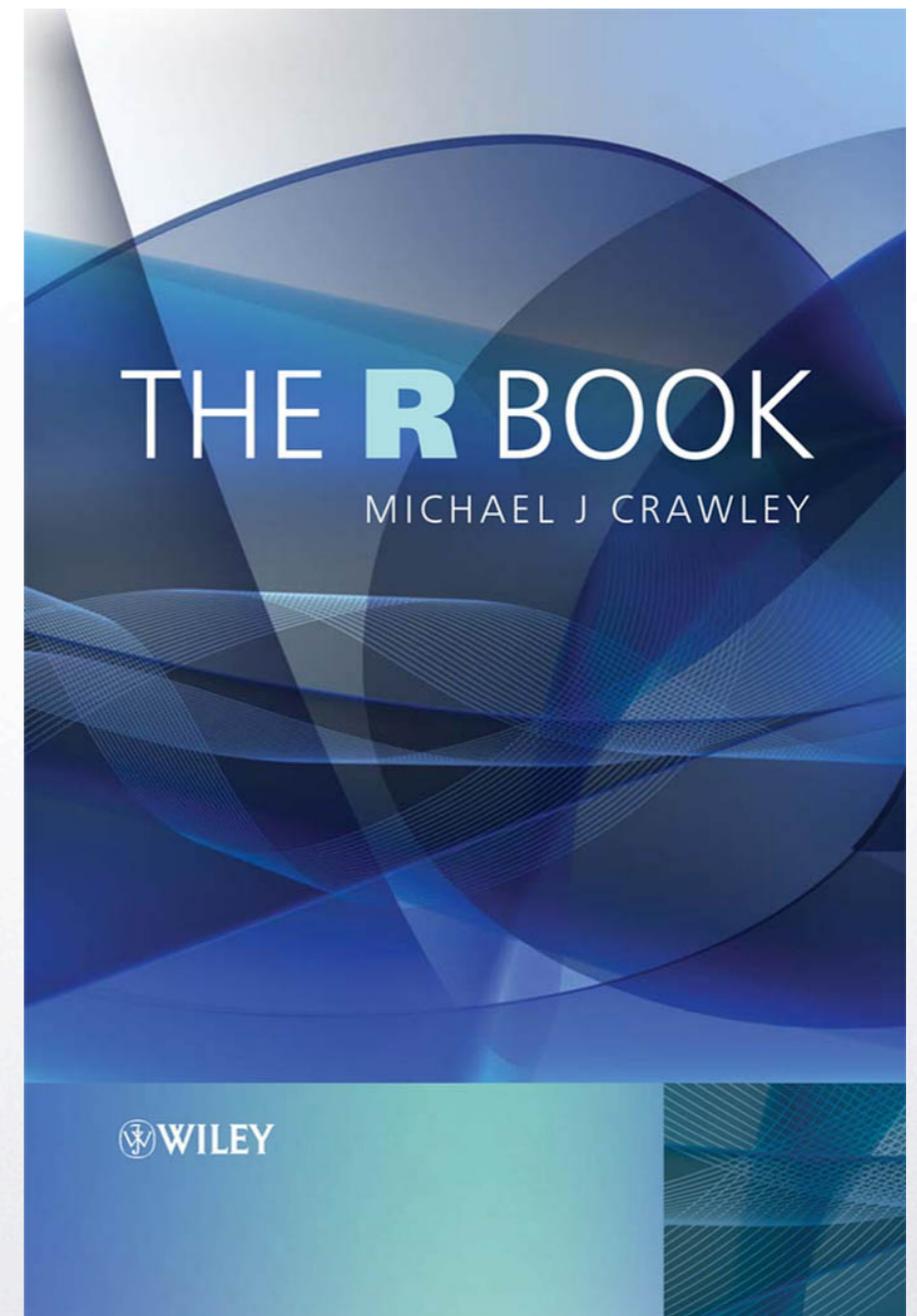
R in a
nutshell de
Joseph Adler





Bibliografía

The R Book
de
Michael J.
Crawley

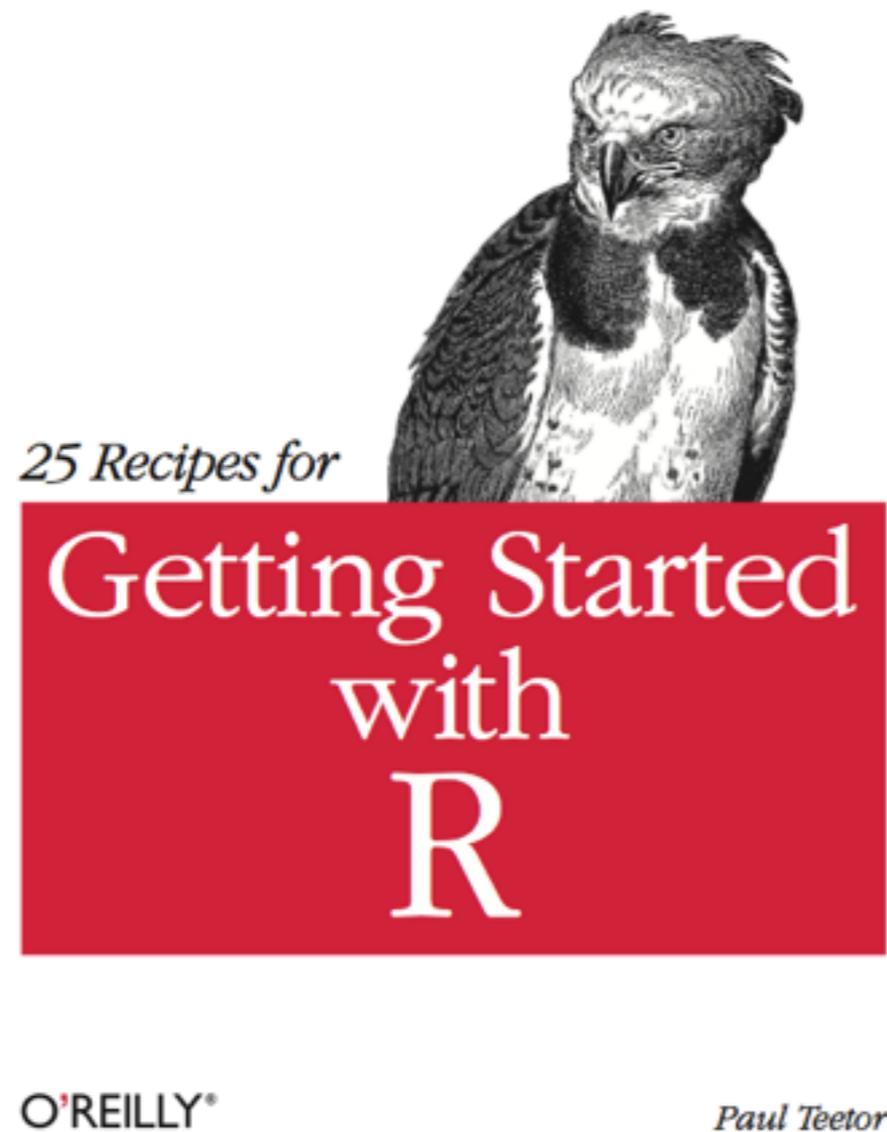




Bibliografía (complementaria)

Excerpts from the R Cookbook

An Integrated Development Environment for R





Manos a la obra :-)

- **Vamos a instalar sus máquinas virtuales**
- **Vamos luego a ver cómo se instala, R, RExcel y RStudio**



Muchas gracias :-)