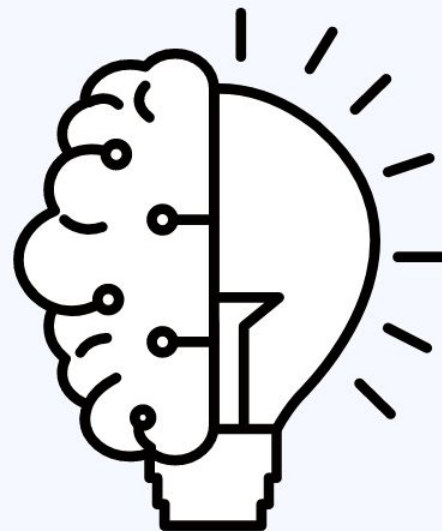


Ciencia de datos

Para el sector público de salud



Introducción a R y RStudio

En colaboración con...



@RLadiesConce



MÓDULO 1: Nivelación y conceptos básicos

Actividades sincrónicas (2 hrs cada uno)

Fecha	Hora	Tema
09/12/2020	18.45h	Aspectos generales del curso
15/12/2020	18.45h	Introducción a R y RStudio
17/12/2020	18.45h	Estructura de datos y operadores
22/12/2020	18.45h	Importación de datos
29/12/2020	18.45h	Análisis prefactibilidad y valor público

Avisos

Revisen frecuentemente la plataforma Twist

Ajusten sus nombres y coloquen su foto

Ajusten las notificaciones a sus necesidades

Reuniones autoconvocadas

La participación es fundamental

Agenda

Introducción a R

Funcionalidades RStudio

Packages (librerías)

Recomendaciones

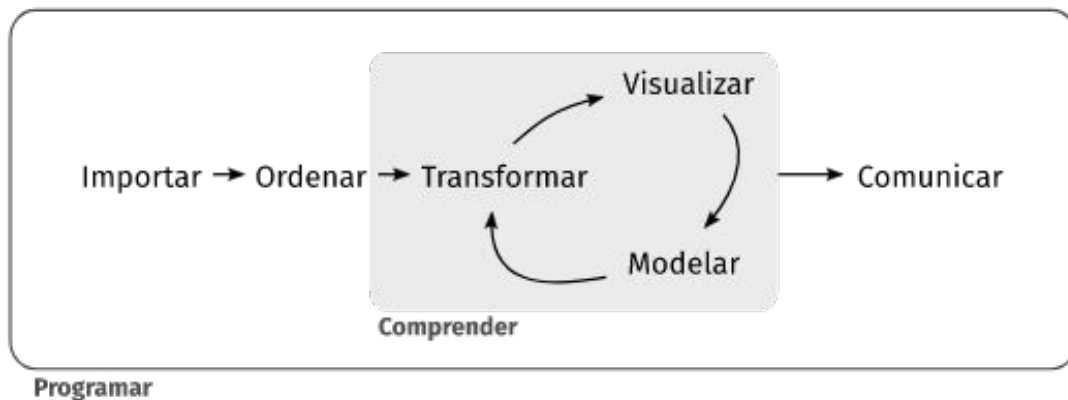
O'REILLY®



R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Grolemund



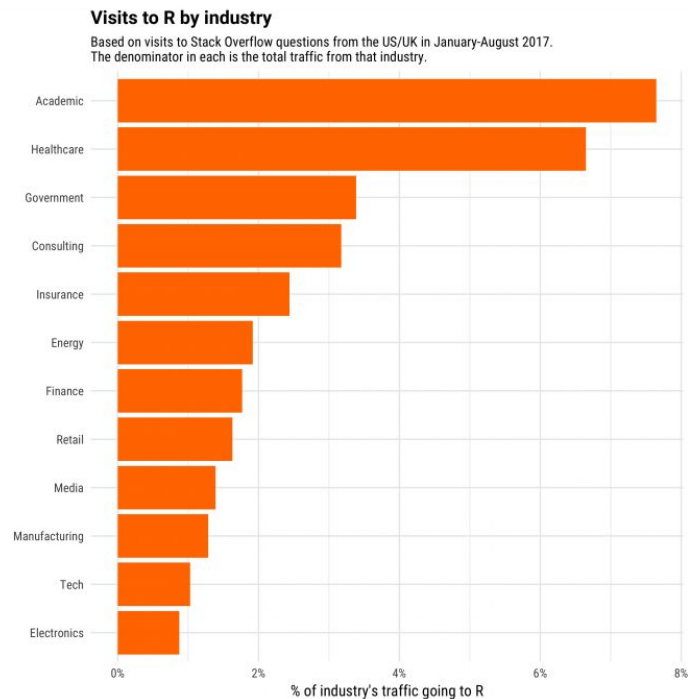
<https://es.r4ds.hadley.nz/>

Introducción a R



R

R es uno de los lenguajes más usados para Data Science



Aug 2020	Aug 2019	Change	Programming Language	Ratings	Change
1	2	▲	C	16.98%	+1.83%
2	1	▼	Java	14.43%	-1.60%
3	3		Python	9.69%	-0.33%
4	4		C++	6.84%	+0.78%
5	5		C#	4.68%	+0.83%
6	6		Visual Basic	4.66%	+0.97%
7	7		JavaScript	2.87%	+0.62%
8	20	▲▲	R	2.79%	+1.97%
9	8	▼	PHP	2.24%	+0.17%
10	10		SQL	1.46%	-0.17%

The top ten languages in TIOBE's Programming Community index for August 2020.
Image: TIOBE



Lo mejor es usar cada uno según necesitemos hacer algo.

R

- Enfoque estadístico
- Basado en el lenguaje S (70's) John Chambers (laboratorios Bell)
- Creado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland (1993)
- Open source y gratuito (GNU General Public License)
- R Development Core Team (1997, versión 1.0 / 2000), financiados por R Foundation
- Multiparadigma (OOP / funcional)
- +10.000 librerías (packages)
- Gran comunidad

Desventajas R

- Lenguaje ya algo antiguo
- Pobre en análisis 3D
- Funcionalidades basado en usuarios
- Los objetos deben estar en RAM
- No sirve para todo

¿Hay otros lenguajes para DS?

Python

Scala

Julia

SQL

JavaScript



R: Engine



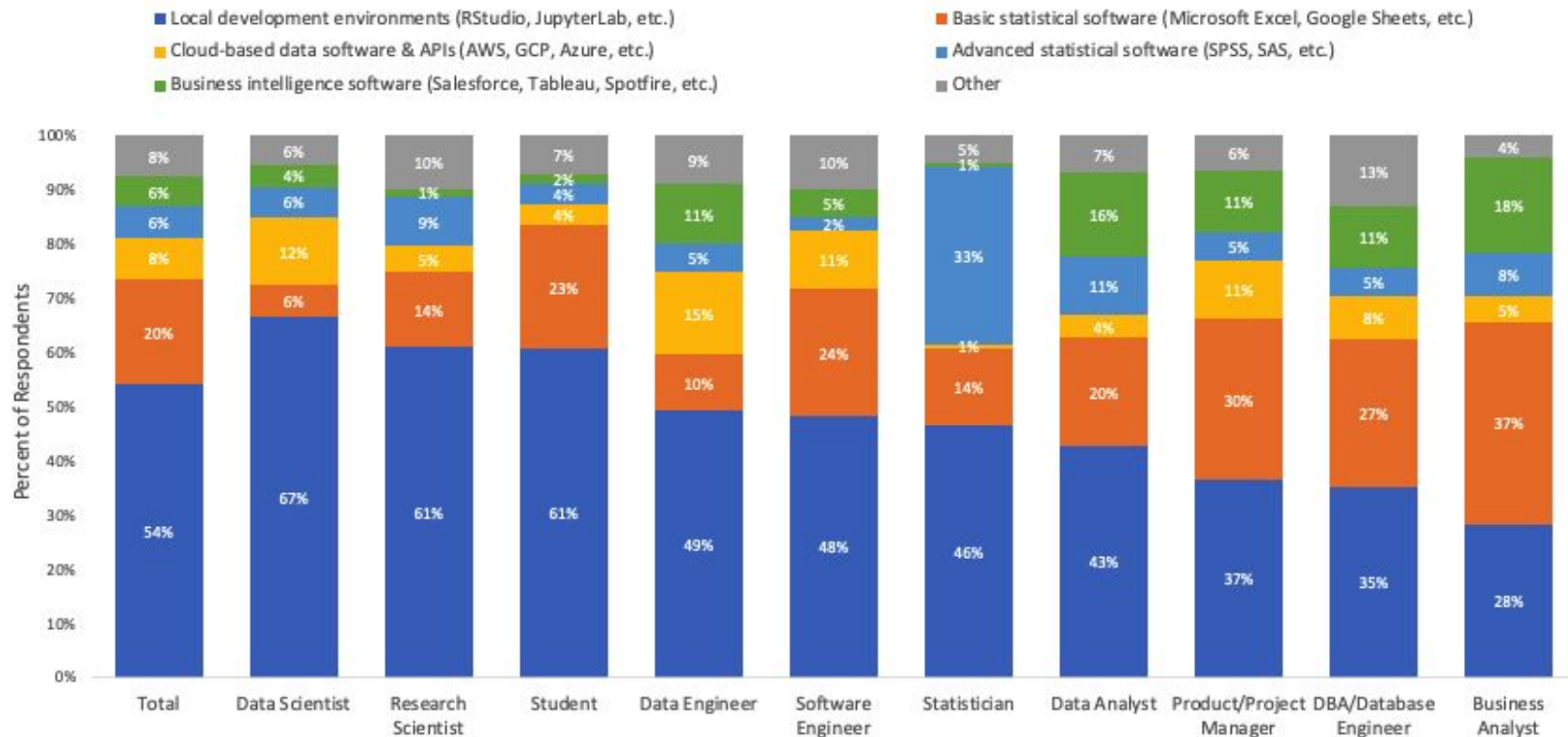
RStudio: Dashboard



FIGURE 1.1: Analogy of difference between R and RStudio.

Imagen de Modern Dive

What is the primary tool that you use at work or school to analyze data?



Note: Data are from the 2019 Kaggle ML and Data Science Survey. You can learn more about the study here: <https://www.kaggle.com/c/kaggle-survey-2019/data>; A total of 19717 respondents completed the survey.

RStudio

IDE (integrated development environment)

Interfaz unificada

Documentación y ayuda integrada

Revisión de sintaxis y autocompletado

v1.3 (1.4.1087-9 preview)

Packages



Packages

CRAN: The Comprehensive R Archive Network

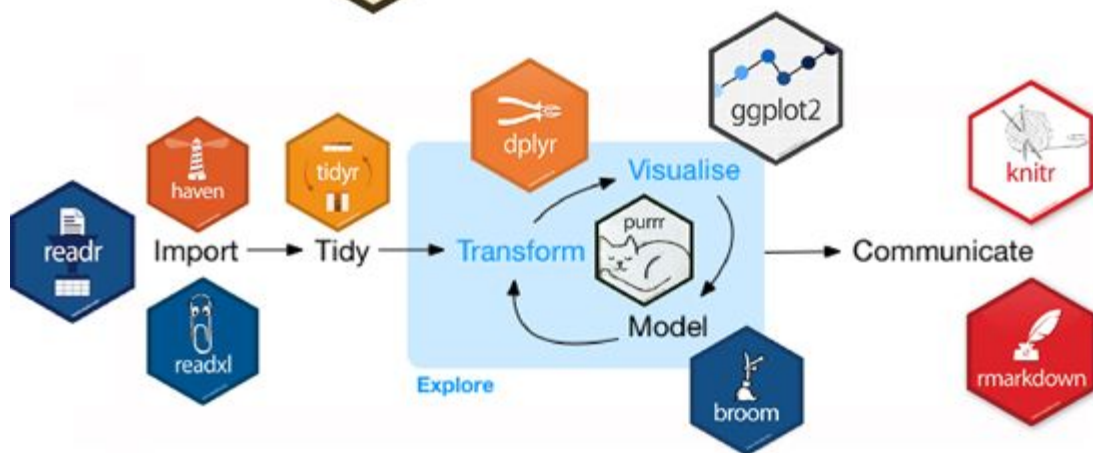
<https://cran.r-project.org/>

GitHub

Tidyverse

<https://www.tidyverse.org/>





data.frames



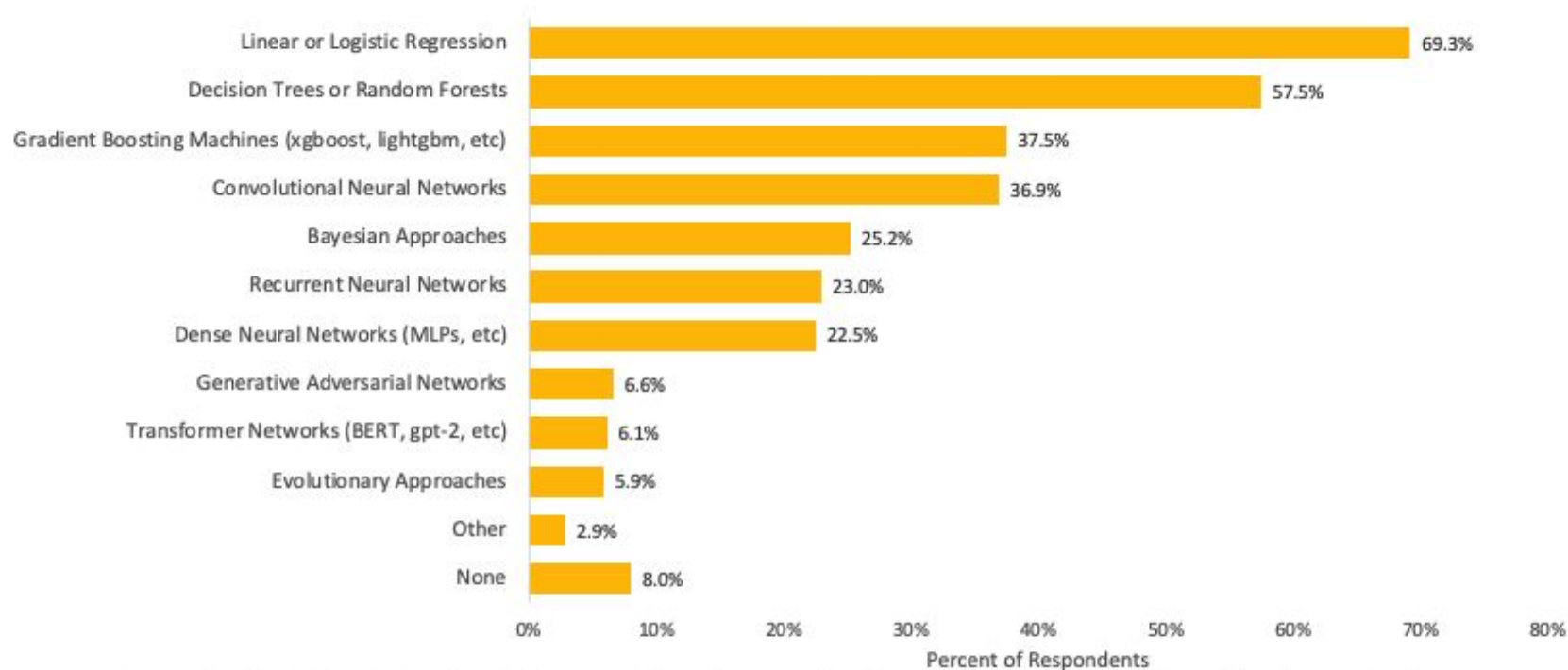
factors



strings



Which of the following ML algorithms do you use on a regular basis? (Select all that apply)



Note: Data are from the 2019 Kaggle ML and Data Science Survey. You can learn more about the study here: <https://www.kaggle.com/c/kaggle-survey-2019/data>. A total of 19717 respondents completed the survey; the percentages in the graph are based on a total of 14762 respondents who provided an answer to this question.

Comunidades, recursos y ayuda

StackOverflow

Twitter: @RLadiesConce, #Rstats, #RstatsES, @hadleywickham, @rstats4ds

Towardsdatascience.com

RStudio community

RStudio resources

R4DS community

YouTube

Google

Recomendaciones

Recomendaciones 1

Usar modo oscuro

Usar proyectos en RStudio

Ser consistente con el tipo de nombres

Usar carpetas para los proyectos:

- Crea una subcarpeta para tus datos originales
- Crea una subcarpeta para los datos limpios
- Crea una subcarpeta para tus scripts
- Crea una carpeta para cada documento
- Crea una carpeta si necesitas guardar figuras

Recomendaciones 2

Usar nombres claros y reconocibles para los scripts:

- 00_load.R, puedes cargar tus datos en varios formatos
- 01_clean.R, para limpiar los datos
- 02_eda.R, para en análisis exploratorio y gráficos
- 03_analysis_cor.R, por ejemplo

Agregar comentarios al código

Usar <- y no = para asignaciones

Indentación

Recomendaciones 3

Usar atajos de teclado

- Ctrl / Cmd + Enter ejecuta la línea / bloque de código R en la consola R
- Ctrl + 1 salta al script
- Ctrl + 2 saltar a la consola R
- Ctrl + Shift + N nuevo script
- Ctrl + Shift + C comentar la línea
- Ctrl + Shift + R insertar salto de sección
- Ctrl + Alt + i insertar un fragmento Rmd vacío
- Alt + - inserta el <-
- Ctrl + Shift + M inserta el operador de la tubería %>%
- Alt + ↑/ ↓ Permite mover líneas de código