



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Introducción a la Ciencia de Datos con R

## Miguel Jorquera

Educación Profesional  
Escuela de Ingeniería

El uso de apuntes de clases estará reservado para finalidades académicas. La reproducción total o parcial de los mismos por cualquier medio, así como su difusión y distribución a terceras personas no está permitida, salvo con autorización del autor.



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

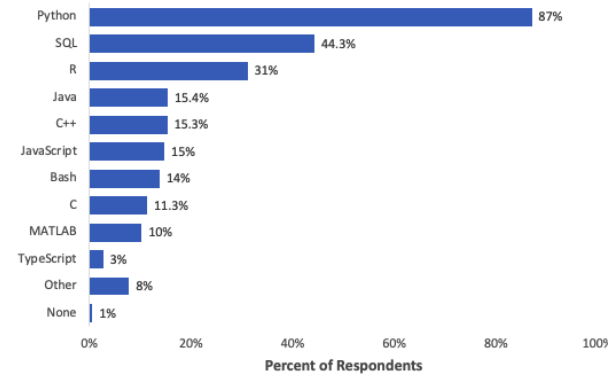
# INTRODUCCIÓN

# INTRODUCCIÓN

## ¿Qué es R?

- Software libre, basado en lenguaje S.
- Lenguaje funcional.
- Uso: Principalmente investigación, academia, análisis de datos.
- Ventajas (intuitivo, especializado, buena documentación, comunidad y CRAN)
- Desventajas (paradigma poco flexible, nativamente poco escalable\*)

What programming languages do you use on a regular basis?



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>. 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.



Copyright 2020 Business Over Broadway

## Características de R



### CRAN

Comunidad global



### Academia

Metodologías actualizadas



### Open Source

Colaboración masiva



### Empresa

Auge y soporte



## ¿Para qué es?

- Software altamente especializado en el análisis de datos.
- Ideal para análisis descriptivo, modelamiento predictivo e implementación de modelos de machine learning.
- Por ejemplo...





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# INSTALACIÓN LOCAL

## ¿Qué es R?

- **Instalación del R en local**  
Ultima versión disponible desde CRAN (<https://cran.r-project.org/>).
- **Interfaz de desarrollo local**  
Rstudio (<https://www.rstudio.com/>)
- **Interfaz cloud**  
Google Colab (R) ([colab.fan/r](https://colab.fan/r))  
Nuestro repositorio (<https://github.com/majorquev/Introduccion a ciencia de datos R.git>)



CRAN  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R  
[R Homepage](#)  
[The R Journal](#)

Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

### The Comprehensive R Archive Network

#### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

#### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-06-22, Taking Off Again) [R-4.0.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features](#) and [bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

#### Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

