



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.

# NOSOTROS

AYUDANTE



**Rodrigo Morales** · 1er

Estadístico | R programmer | Analista Cuantitativo

Chile · [Información de contacto](#)



Stratagemma Trading Systems



Pontificia Universidad  
Católica de Chile



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

# NOSOTROS

RELATOR



Miguel André Jorquera Viguera

Senior Data Scientist Consultant en EY - MetricArts

Chile · [Más de 500 contactos](#) · [Información de contacto](#)



EY



Pontificia Universidad  
Católica de Chile



# OBJETIVOS

- Revisar **conceptos básicos de programación** usando el lenguaje R.
- Usar el lenguaje R como una herramienta para **analizar datos** .
- Utilizar los sistemas gráficos de R para **visualización de datos** .
- Analizar conjuntos de datos utilizando los principios del **análisis exploratorio** de datos.
- Explicar la **información visual** contenida en los gráficos que generan R.
- Explicar y presentar los resultados de un **análisis de datos**.



# CONTENIDOS

## Clase 1:

- (3 hr) Elementos de programación en R
  - Tipo de datos en R
    - Vectores
    - Matrices
    - Listas
    - Dataframe
    - Base plot

## Clase 2:

- (1 hr) Adquisición de datos desde archivos csv/web
- (2 hr) Preparación de datos: limpieza y transformación
  - dplyr verbs (filter, arrange, mutate, group\_by, select)
  - Joins
  - Imputación de datos

## Clase 3:

- (2 hr) Adquisición mediante web scraping (.json)
- (1 hr) Interacción con bases de datos

- (1 hr) Estadística descriptiva (EDA)
- (2 hr) Visualización de datos en R

## Clase 5:

- (2 hr) Reglas de asociación
- (1 hr) Algoritmos de clustering

## Clase 6:

- (1 hr) Algoritmos de clustering
- (2 hr) Algoritmos de regresión

## Clase 7:

- (1 hr) selección y evaluación (regresión)
- (2 hr) Algoritmos de clasificación
  - KNN
  - Árboles y
  - Bagging
  - Random Forest
  - SVM

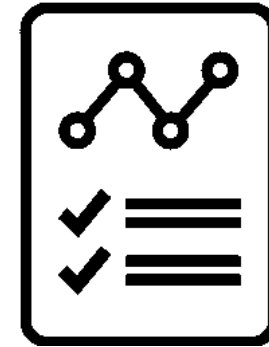
## Clase 8

- (2 hr) Selección y evaluación de modelos (clasificación)
- (1 hr) Validación cruzada



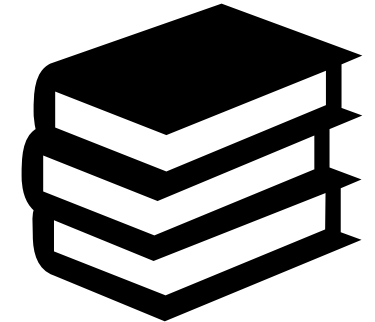
# EVALUACIÓN

- El programa consta de dos talleres evaluados cuyo horario de inicio será durante las sesiones de ayudantías
- Taller 1: Viernes 24 de sept.(50%)  
Inicio en Ayudantía 4
- Taller 2: Miércoles 13 de oct. (50%)  
Inicio en Ayudantía 6



# BIBLIOGRAFÍA

- [Wickham et al.2017]. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (1st ed.). O'Reilly Media, Inc..
- [Adler 2009] R in a Nutshell: A Desktop Quick Reference (1st ed.), O'Reilly Media, Inc..
- [Wickham 2009] Hadley Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer.
- [Dalgaard 2008] Peter Dalgaard, Introductory Statistics with R, Second Edition, Springer.
- [James et al. 2013]. An Introduction to Statistical Learning : with Applications in R. New York, Springer.
- [Munzert Simon et al. 2015] Automated Data Collection with R. A Practical Guide to Web Scraping and Text Mining



# COORDINACIÓN

- Elías Ortiz  
eortil@ing.puc.cl

