

# Alternativas Open Source para ciencia de datos

Daniel E. de la Rosa

2020-11-07



# Contents

<b>Presentación</b>	<b>5</b>
Notas metodológicas . . . . .	6
<b>1 Sintaxis</b>	<b>7</b>
1.1 Secciones de código . . . . .	7
1.2 Comentarios . . . . .	7
<b>2 Gráficos</b>	<b>9</b>
2.1 Gráficos de funciones . . . . .	9



# Presentación

Este libro tiene como objetivo principal el de evaluar el uso en primer lugar de R como alternativa a softwares privativos para la realización de ciencia de datos. Este interés responde a múltiples motivaciones, detro de las que podemos mencionar:

- **Interoperabilidad:** cuanto menor sea el número de herramientas y/o lenguajes necesarios para resolver un problema, más eficiente y efectiva resulta la solución. Aún si múltiples lenguales son necesarios, la posibilidad de poder utilizar todos o la gran mayoría de ellos en un mismo documento tiene una gran importancia.
- **Costo:** el costo de las licencias de algunos de los softwares aquí cubiertos es quizás una las motivaciones más importantes para la búsqueda de alternativas open source.
- **Curva de aprendizaje:** como científicos de datos estamos llamados a manejar con cierta destreza los lenguajes de programación R y Python. No así el apredizaje de un nuevo lenguaje de programación o manejo de alguna herramienta tiene asociado un costo en tiempo y esfuerzo, que se convierte en una mativación más para aprender a resolver los problemas más comunes con las herramientas habituales.

Este documento, en principio, es por y para los estudiantes de la Maestría de Ciencias de Datos del INTEC y no pretende cubrir exhaustivamente ningun software o lenguaje de programación, sino más bien ofrecer soluciones alternativas, primordialmente en R, a los problemas y ejercicios de la maestría desarrollados en softwares propietarios.

La estructura del documento irá cambiando a lo largo del tiempo para ajustarse a las necesidades, siempre con el interés de agrupar los conceptos acorde a una temática en particular en un mismo capítulo.

## Notas metodológicas

Para mostrar los resultados de MatLab se utilizará la función **Publish** con el que MatLab genera un documento html que combina el código con los resultados. Posteriormente se utilizará la función **insertHTML** que toma un documento html e inserta su contenido en el libro.

# Chapter 1

## Sintaxis

Este capítulo reúne los aspectos de sintaxis de todos los lenguajes utilizados a lo largo del libro. Elementos como la inserción de comentarios, la creación de objetos básicos, entre otros que sean prácticamente común a cualquier implementación del lenguaje.

### 1.1 Secciones de código

#### 1.1.1 MatLab

En MatLab se crea una sección de código insertando dos signos de porcentaje `%`, y una subsección insertando tres `%%`.

#### 1.1.2 R

En R para insertar una sección de código inicias la línea de código con un signo de número `#` y la terminas con cuatro signos de `#` o guiones `-`.

Ejemplo:

```
# Sección 1 —
```

### 1.2 Comentarios

Los comentarios en los lenguajes de programación representan texto legible que sirve para orientar al programador pero que en principio son obviados por el lenguaje. Por lo general son utilizados para describir el código a su alrededor.

#### 1.2.1 MatLab

En MatLab un comentario se inserta utilizando el signo de porcentaje `%`.

### 1.2.2 R

En un script básico de R, un comentario puede ser insertado utilizando el símbolo de número #.



## Chapter 2

# Gráficos

### 2.1 Gráficos de funciones