



## ANÁLISIS DE DATOS

### FACULTAD DE INGENIERÍA Y NEGOCIOS

(ACI777)

Melissa Alegría Arcos, PhD.  
[malegria@udla.cl](mailto:malegria@udla.cl)

Agosto 2021

# Descripción del programa

Modalidad : E-SUPPORT

Horario : Martes 19:15 a 22:30

Distribución de horas de dedicación:

- Laboratorio: 54
- Trabajo Personal 126

La asignatura ANÁLISIS DE DATOS tiene por meta formativa que los estudiantes aprendan los fundamentos de análisis basado en datos para resolver problemas organizacionales y apoyar la toma de decisiones en base a la generación de modelos descriptivos y predictivos a partir de datos.

- Esta asignatura cuenta con un aula virtual e-support, la que será fuente oficial de material transversal para la asignatura.  
<https://udla.blackboard.com>

# Descripción del programa

## Contenido:

### 1) Estadística

- Estadística Descriptiva
- Analítica para toma decisiones.

### 2) Modelos Analíticos Descriptivos

- Análisis Exploratorio de datos

### 3) Aprendizaje No Supervisado

- Clustering
- Reducción de dimensionalidad
- Análisis de Componentes Principales (PCA)

### 4) Aprendizaje Supervisado

- Regresión Lineal y Regresión Logística
- Arboles de Decisiones
- Máquinas de Soporte Vectorial (SVM)

# Evaluaciones

Régimen	Ponderación	Componente	% Componente	Subcomponente	% Subcomponente
TODOS	24	EXAMEN	35	EXAMEN	100
		CATEDRA	50	CATEDRA 1	50
				CATEDRA 2	50
		EJERCICIO	15	EJERCICIO 1	25
				EJERCICIO 2	25
				EJERCICIO 3	25
				EJERCICIO 4	25

**Sobre eximición:**

El examen de esta asignatura es obligatorio para todo estudiante que la curse, por lo que no contempla eximición bajo ninguna circunstancia.

- El **plagio** y toda acción u omisión que vaya contra la ética, el reglamento y la normativa vigente será sancionada de acuerdo al Reglamento del alumno y normativa Institucional vigente. En particular, el plagio será sancionado con nota mínima 1,0 en la actividad evaluativa. Sin perjuicio de lo anterior y dependiendo de la falta cometida por el estudiante, el caso será evaluado por la Dir. de Escuela y autoridades FINE para determinar si corresponde la reprobación inmediata de la asignatura con nota mínima 1,0 como nota final de la asignatura, de acuerdo a la normativa y reglamento vigente.

# Profesora a cargo



**4 AÑOS**  
**UNIVERSIDAD ACREDITADA**  
DESDE MARZO 2019 / HASTA MARZO 2023  
GESTIÓN INSTITUCIONAL Y DOCENCIA DE PREGRADO



## Educación:

- Ingeniero en Bioinformática, Universidad de Talca.
- PhD en Ciencias m/ Biofísica y Biología Computacional, Universidad de Valparaíso.

## Posiciones:

- 2012-2015 Asistente de Investigación. Centro de Bioinformática y Biología Integrativa, UNAB
- 2017- 2019 Ingeniero en Bioinformática en uBiome Chile. Drug Design Group.
- 2019 -2021 Founder & CEO DrugDevs SpA. Start-up dedicada a Drug Repurposing . TSF-10 (start-up Chile, Corfo).
- 2021 Académico Investigador. Facultad de Ingeniería y Negocios, Universidad de Las Américas.

## Pasantías:

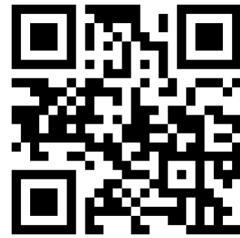
- Instituto de Bioinformática Europeo (EBI), grupo de la Dr. Janet Thornton, directora del EBI en Reino Unido. (2 meses)
- Octubre-Diciembre 2016 (2 meses) Grupo de la Dra. Patricia Babbitt., Universidad de California en San Francisco. Tema: Aprendizaje y perfeccionamiento en SSN (Redes de Similitud de Secuencia).

## Innovación:

Octubre 2015: Winner of the 2nd Competition R&D – “APLICA TU IDEA”, Santiago, Chile. → Reingeniería de enzimas.

Septiembre 2019 : DrugDevs, ganador TSF10 start-up Chile. (4 meses y posterior extensión) → Drug repurposing.

## Presentación con los estudiantes



# Introducción

## Análisis de datos.

También llamada analítica de datos, tiene como propósito examinar un conjunto o grandes cantidades de datos con el fin de encontrar tendencia o relaciones sobre determinada información para predecir, tomar decisiones o simplemente realizar un análisis exploratorio sobre un determinado tema.



# Introducción



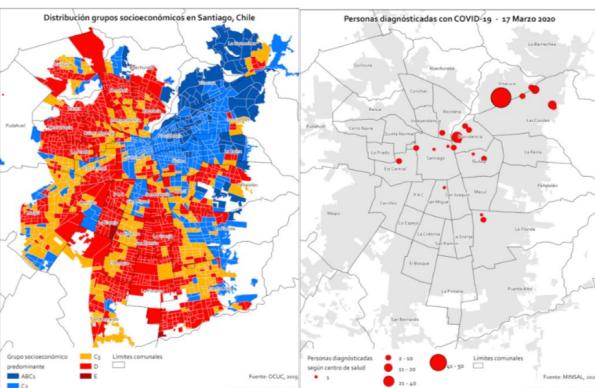
**Análisis de Mercado**



**Seguridad y Control**



**Académicos**

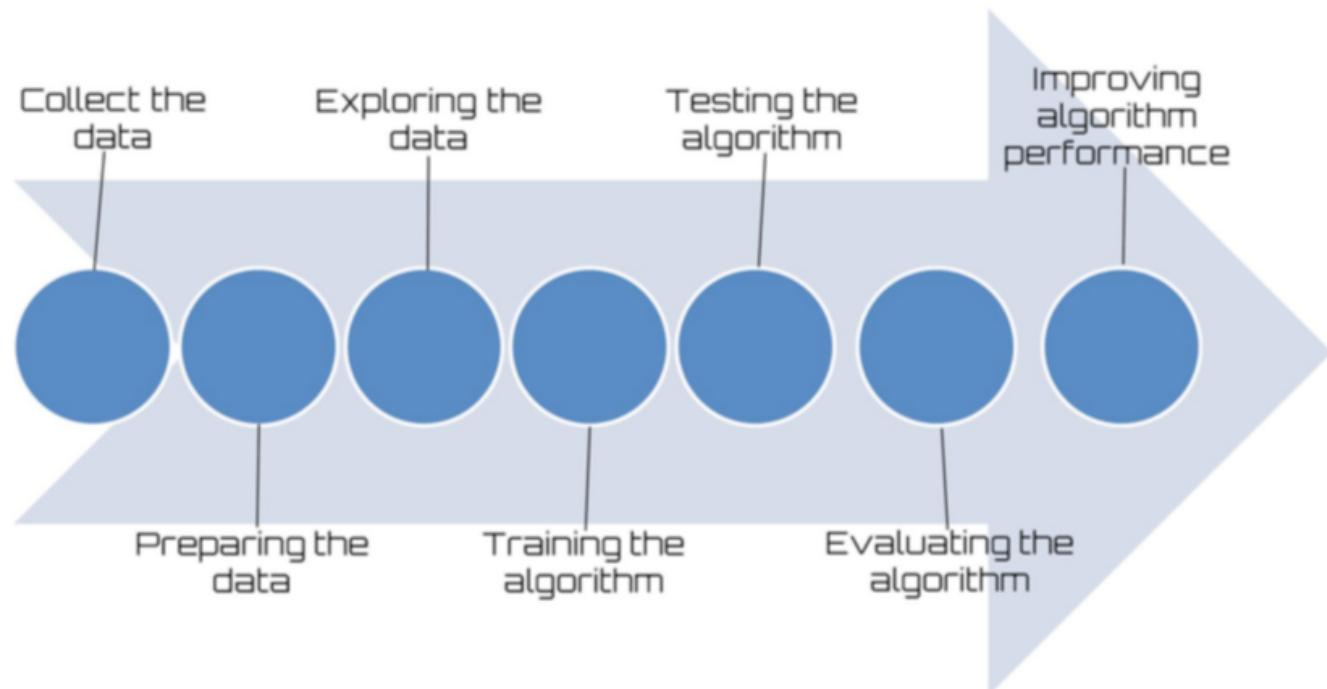


**Análisis Georreferenciados**



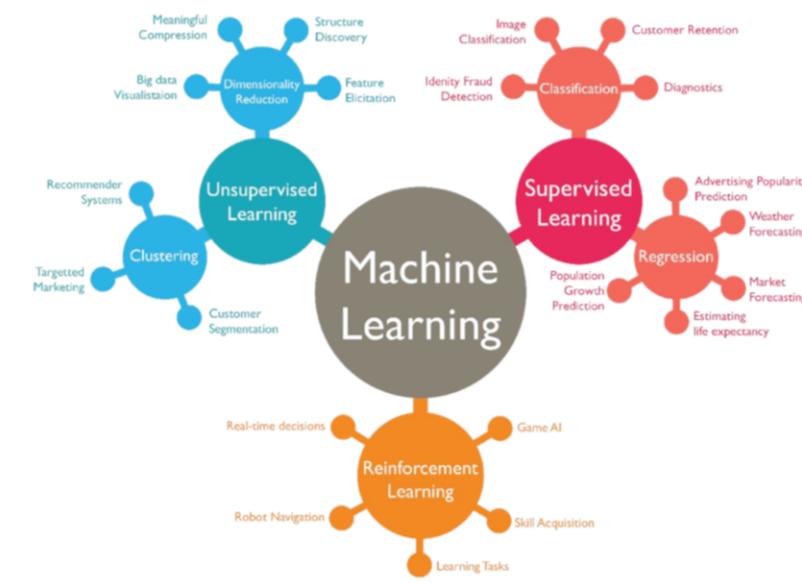
**Ciudades Inteligentes y el IoT**

# Proceso del análisis de Datos



# Machine Learning

El aprendizaje automático es el estudio científico de algoritmos y modelos estadísticos que los sistemas informáticos utilizan para realizar una tarea específica sin utilizar instrucciones explícitas, basándose en patrones e inferencias. Se considera un subconjunto de la inteligencia artificial



# Requerimientos para el curso



## Descargar Anaconda

Ingresá a <https://www.anaconda.com/download/#linux>



# Requerimientos para el curso

## Anaconda Individual Edition

[Download](#) 

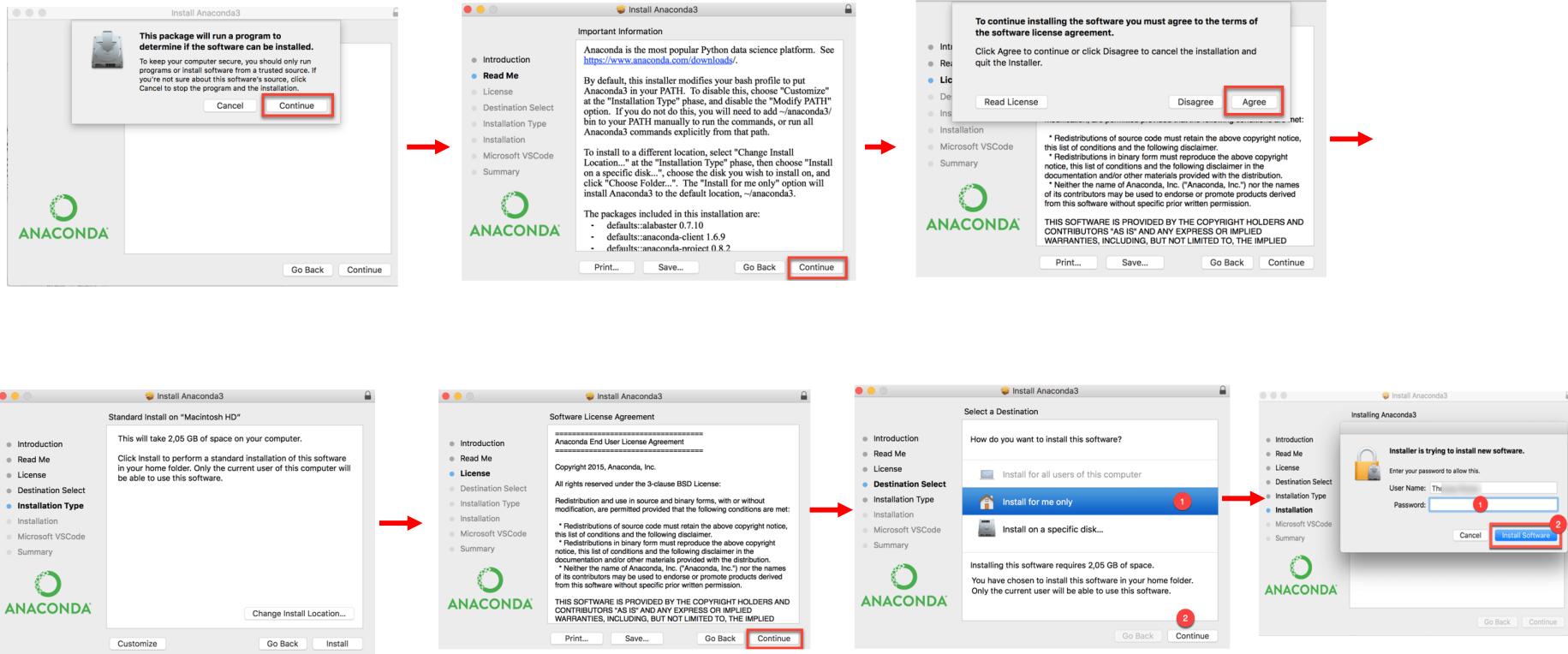
For MacOS

Python 3.8 • 64-Bit Graphical Installer • 440 MB

Get Additional Installers



# Requerimientos para el curso



# Requerimientos para el curso

## Windows

\*\*Verificar que su nombre de usuario en Windows no tenga espacios ni caracteres como tildes, comas, apóstrofes o eñes. Esto debido a que se pueden generar errores de codificación.

Ejemplos:

- C:\Users\Melissa : **Correcto**
- C:\Users\Noé : **Incorrecto**
- C:\Users\Melissa Alegría : **Incorrecto**

El siguiente es instalar la app en nuestro sistema. (Deberá tener permisos de Administrador si instala para todos los usuarios).

Ejecutamos el archivo que descargamos haciendo doble click.

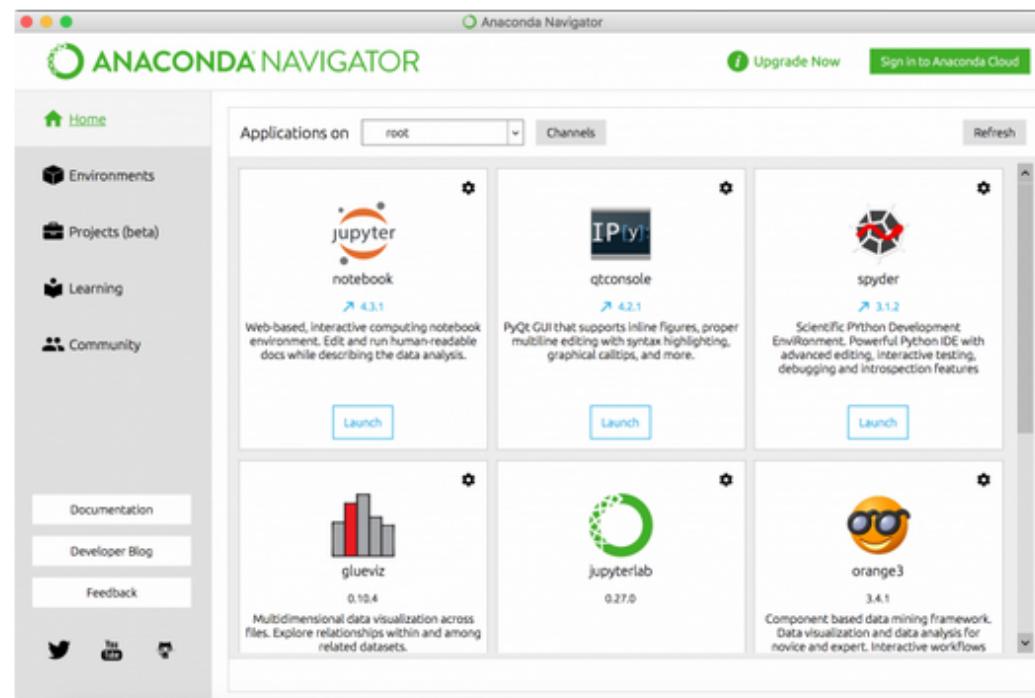
Se abrirá un “Típico Wizard” de instalación.

Seguiremos los pasos, podemos seleccionar instalación sólo para nuestro usuario, seleccionar la ruta en disco donde instalaremos y listo.

# Requerimientos para el curso

## Iniciar y Actualizar Anaconda

Anaconda viene con una suite de herramientas gráficas llamada "Anaconda Navigator". Cuando la abrimos se podra visualizar una pantalla como esta:



# Requerimientos para el curso

Comprobar la instalación abrimos una Terminal de Mac/Linux/Ubuntu o la Linea de Comandos de Windows.

Escribimos:

```
>conda -V
```

```
c(base) MacBook-Pro-de-Melissa:~ melissaalegria$ conda -V
conda 4.10.1
```

```
>python -V
```

```
● ● ● melissaalegria — bash — 80x24
(base) MacBook-Pro-de-Melissa:~ melissaalegria$ python -V
Python 3.8.5
```

*>conda update conda #debemos poner 'y' para actualizar y se descargarán.  
>conda update anaconda*

# Requerimientos para el curso

Comprobar la instalación este bien.

<https://raw.githubusercontent.com/jbagnato/machine-learning/master/versiones.py>

```
1 # scipy
2 import scipy
3 print('scipy: %s' % scipy.__version__)
4 # numpy
5 import numpy
6 print('numpy: %s' % numpy.__version__)
7 # matplotlib
8 import matplotlib
9 print('matplotlib: %s' % matplotlib.__version__)
10 # pandas
11 import pandas
12 print('pandas: %s' % pandas.__version__)
13 # statsmodels
14 import statsmodels
15 print('statsmodels: %s' % statsmodels.__version__)
16 # scikit-learn
17 import sklearn
18 print('sklearn: %s' % sklearn.__version__)
```

Luego ejecutamos:

```
>python versiones.py
```

# Requerimientos para el curso

## Actualizar libreria scikit-learn

Para actualizar esta librería muy usada en Machine Learning ejecutaremos los siguientes comandos:

```
> conda install scikit-learn. # instalar scikit-learn  
> conda update scikit-learn #Debemos confirmar la actualización poniendo 'y' en la terminal.
```

Volvemos a ejecutar el script versiones.py

```
> python versiones.py
```

# Instalación Jupyter Notebook o Jupyter Lab

Jupyter Notebook es un entorno de trabajo interactivo web que permite desarrollar código en Python de manera dinámica, a la vez que integrar en un mismo documento tanto bloques de código como texto, gráficas o imágenes. Te permite editar y ejecutar documentos de notebook a través de cualquier navegador web, e incluso dentro de algunas IDE.

Instalación y Uso de Jupyter Notebook



<https://jupyter.org/install>

# Instalación Jupyter Notebook o Jupyter Lab

<https://jupyter.org/install>

## Installation with mamba or conda

JupyterLab can be installed with `mamba` and `conda`:

```
mamba install -c conda-forge jupyterlab
```

or

```
conda install -c conda-forge jupyterlab
```

Note: If you have not installed mamba or conda yet, you can get started with the [miniforge](#) distribution.

## Installation with pip

If you use `pip`, you can install it with:

```
pip install jupyterlab
```

If installing using `pip install --user`, you must add the user-level `bin` directory to your `PATH` environment variable in order to launch `jupyter lab`. If you are using a Unix derivative (FreeBSD, GNU / Linux, OS X), you can achieve this by using `export PATH="$HOME/.local/bin:$PATH"` command.

## Run JupyterLab

Once installed, launch JupyterLab with:

```
jupyter-lab
```

# Instalación Jupyter Notebook o Jupyter Lab

<https://jupyter.org/install>

## Getting started with the classic Jupyter Notebook

### Installation with mamba or conda

The classic notebook can be installed with `mamba` and `conda`:

```
mamba install -c conda-forge notebook
```

or

```
conda install -c conda-forge notebook
```

### Installation with pip

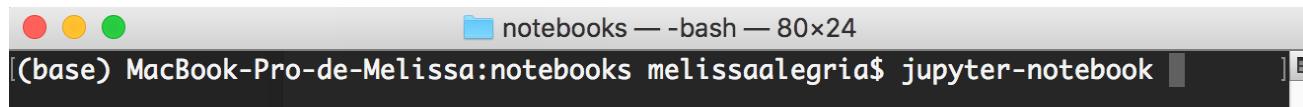
If you use `pip`, you can install it with:

```
pip install notebook
```

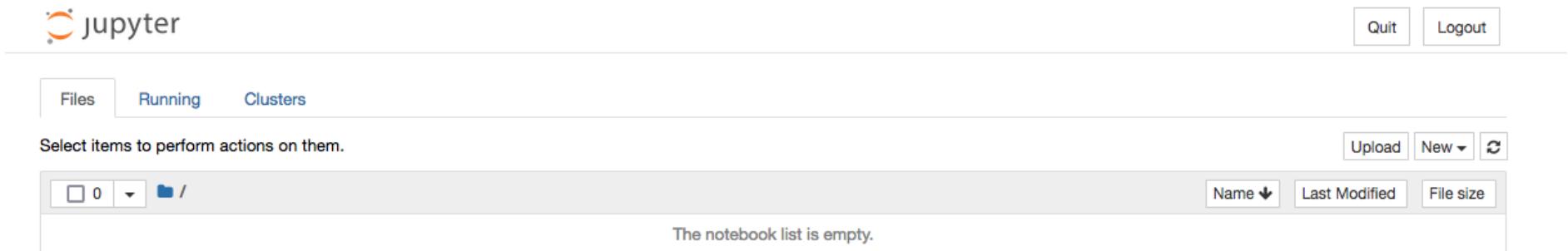
Congratulations, you have installed Jupyter Notebook! To run the notebook, run the following command at the Terminal (Mac/Linux) or Command Prompt (Windows):

```
jupyter notebook
```

# Verificar la instalación



```
notebooks — -bash — 80x24
(base) MacBook-Pro-de-Melissa:notebooks melissaalegria$ jupyter-notebook
```

A screenshot of a terminal window on a Mac OS X system. The title bar says "notebooks — -bash — 80x24". The command entered is "jupyter-notebook". The window has standard OS X window controls (red, yellow, green buttons) and a close button.

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New ↘

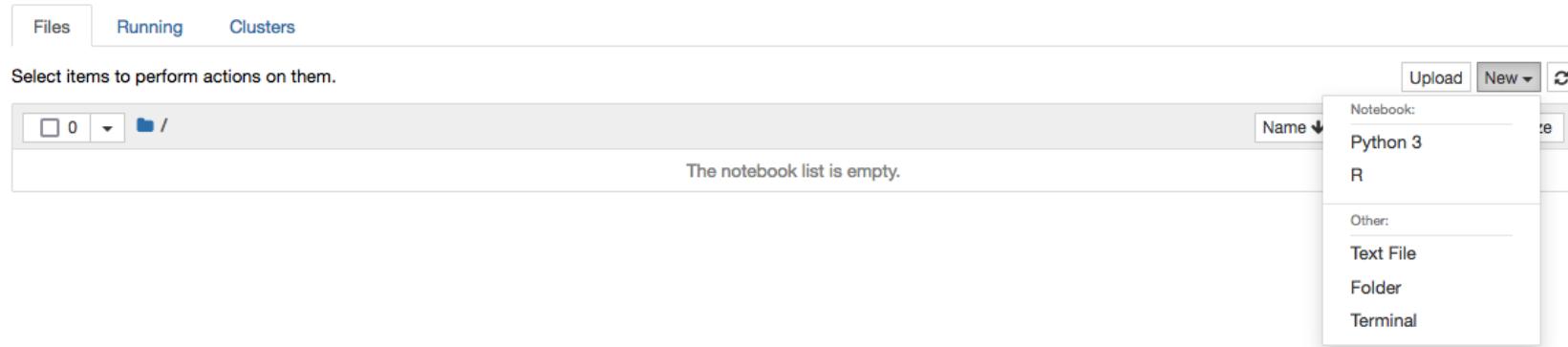
0 / Name ↓ Last Modified File size

The notebook list is empty.

The screenshot shows the Jupyter Notebook web interface. At the top, there's a header with the Jupyter logo, "jupyter", "Quit", and "Logout" buttons. Below the header, there are tabs for "Files", "Running", and "Clusters", with "Files" being active. A toolbar above the main area includes "Upload", "New" (with a dropdown arrow), and a refresh icon. The main content area shows a file browser with a directory structure: "0" files and "1" folder named "/". A message at the bottom of the list says "The notebook list is empty.".

# Verificar la instalación

Crear un nuevo proyecto en New--> Python3

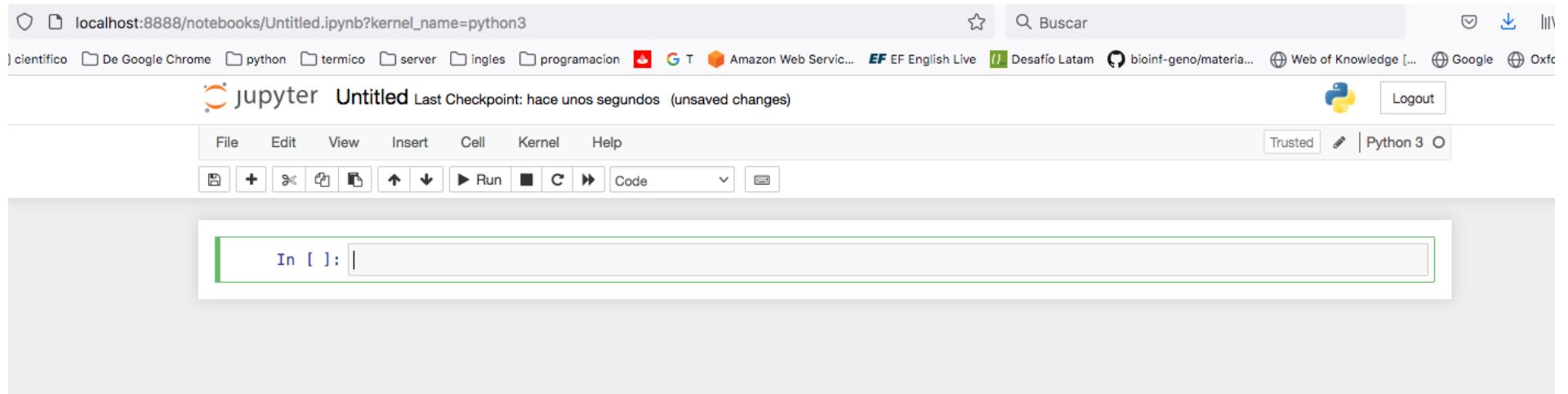


The screenshot shows a Jupyter Notebook interface. At the top, there are three tabs: 'Files' (selected), 'Running', and 'Clusters'. Below the tabs, a message says 'Select items to perform actions on them.' On the left, there's a file browser showing '0' files in the current directory '/'. In the center, it says 'The notebook list is empty.' On the right, there's a 'Name' dropdown and a 'New' button with a dropdown menu open. The menu contains:

- Upload
- New 
- 
- Notebook:
  - Python 3
  - R
- Other:
  - Text File
  - Folder
  - Terminal

# Verificar la instalación

Deben visualizar un nuevo notebook:



# Verificar la instalación

Ejecutar estos comandos

## Clase 1 Análisis de Datos

```
In [1]: print ("Clase 1 de Analisis de Datos")
```

```
Clase 1 de Analisis de Datos
```

```
In [6]: name = input("Cuál es tu nombre:\n")
```

```
Cuál es tu nombre:  
Melissa
```

```
In [7]: print("Hola %s" % name)
```

```
Hola Melissa
```

# Verificar la instalación

Ejecutar estos comandos

## Operaciones matemáticas

In [9]: `0+1`

Out[9]: 1

In [10]: `1+1`

Out[10]: 2

In [15]: `50+100`

Out[15]: 150

In [17]: `a = 100  
b = 50  
c = a+b  
print (c)`

150

# Verificar la instalación

## Importando la librería matemática

```
In [18]: import math
```

```
In [20]: variable = 60
c = 100
d = math.sqrt(100)
```

## Imprimiendo variables

```
In [22]: print(variable + 100)
print(d)
```

```
160
10.0
```

```
In [23]: print (d + c)
```

```
110.0
```

```
In [24]: # la potencia = 2 elevado a 4
print (2**4)
```

```
16
```

```
In [26]: #dividir
print(100/60)
```

```
1.6666666666666667
```

## Verificar la instalación

```
In [29]: # variables de texto
nombre = 'Carlos'
apellido = 'Santana'
# Interpolación
print("Mi nombre es {} {}".format(nombre, apellido))
```

```
Mi nombre es Carlos Santana
```

# Verificar la instalación

## Verificando instalación de Pandas

Librería orientada a la manipulación y limpieza de estructuras de datos, que nos permite trabajar fácilmente con tablas de datos.

Combina el alto desempeño de operaciones vectorizadas de numpy con las manipulaciones flexibles de hojas de cálculo y bases de datos relacionales.

- Se añade como “import pandas as pd”.

```
In [30]: import pandas as pd
```

**Dataset con esos tipos de datos:** <https://www.kaggle.com/lepcchenkov/usecdarscatalog>

```
In [32]: df = pd.read_csv('cars.csv')
df
```

# Verificar la instalación

	manufacturer_name	model_name	transmission	color	odometer_value	year_produced	engine_fuel	engine_has_gas	engine_type	engine_capacity	...
0	Subaru	Outback	automatic	silver	190000	2010	gasoline	False	gasoline	2.5	...
1	Subaru	Outback	automatic	blue	290000	2002	gasoline	False	gasoline	3.0	...
2	Subaru	Forester	automatic	red	402000	2001	gasoline	False	gasoline	2.5	...
3	Subaru	Impreza	mechanical	blue	10000	1999	gasoline	False	gasoline	3.0	...
4	Subaru	Legacy	automatic	black	280000	2001	gasoline	False	gasoline	2.5	...
...	...	...	...	...	...	...	...	...	...	...	...
38526	Chrysler	300	automatic	silver	290000	2000	gasoline	False	gasoline	3.5	...
38527	Chrysler	PT Cruiser	mechanical	blue	321000	2004	diesel	False	diesel	2.2	...
38528	Chrysler	300	automatic	blue	777957	2000	gasoline	False	gasoline	3.5	...
38529	Chrysler	PT Cruiser	mechanical	black	20000	2001	gasoline	False	gasoline	2.0	...
38530	Chrysler	Voyager	automatic	silver	297729	2000	gasoline	False	gasoline	2.4	...

38531 rows × 30 columns

# Requerimientos para el curso

## Instalacion de R y Rstudio

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

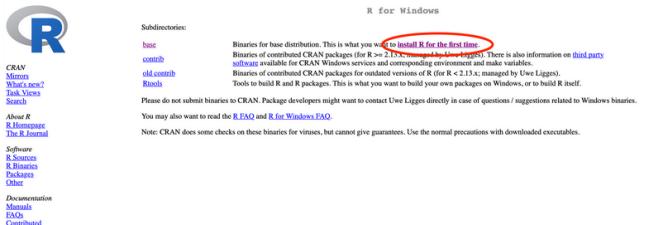
### Download and Install R

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

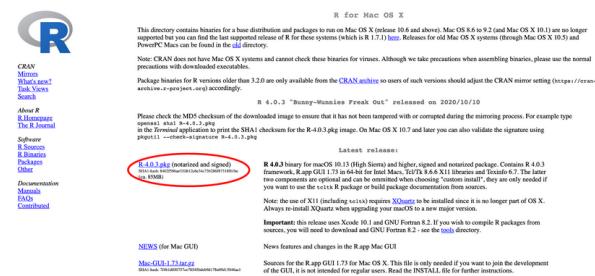
- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

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

Al hacer clic sobre "Download R for Windows" seremos dirigidos a la página que se reproduce a continuación. Allí haremos clic sobre **install R for the first time**.



The screenshot shows the CRAN Windows page. In the left sidebar, under 'Subdirectories', the 'base' directory is selected. A red circle highlights the link 'Install R for the first time'. Below the sidebar, there is a brief description of what the 'base' directory contains: 'Binaries for base distribution. This is what you want if you are installing R for the first time.' Other subdirectories listed include 'contrib', 'old contrib', and 'Rtools'.



The screenshot shows the same CRAN Windows page, but now the 'R 4.0.3' package is highlighted with a red circle. The package is described as 'R 4.0.3 binary for macOS 10.13 (High Sierra) and higher, signed and notarized package. Contains R 4.0.3 framework, R.app GUI 1.7.0 in 64-bit for Intel Macs, Tcl/Tk 8.6.8 X11 libraries and Tkroot 8.7. The later files are not needed by R itself, but they are only needed if you want to use the extra R package or build package documentation from source.' Below the package, the 'NEWS (for Mac GUI)' and 'Mac GUI 1.7.0.tar.gz' links are visible.

# Requerimientos para el curso

Instalar: Rstudio

<https://www.rstudio.com/products/rstudio/download/>



## Choose Your Version

The RStudio IDE is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace.

[LEARN MORE ABOUT RSTUDIO FEATURES](#)



RStudio Desktop	RStudio Desktop Pro	RStudio Server	RStudio Server Pro
Open Source License	Commercial License	Open Source License	Commercial License
<b>Free</b>	\$995 /year	<b>Free</b>	\$4,975 /year (5 Named Users)
<a href="#">DOWNLOAD</a>	<a href="#">BUY</a>	<a href="#">DOWNLOAD</a>	<a href="#">BUY</a>
<a href="#">Learn more</a>	<a href="#">Learn more</a>	<a href="#">Learn more</a>	<a href="#">Evaluation   Learn more</a>
Integrated Tools for R ✓	✓	✓	✓
Priority Support ✓	✓	✓	✓
Access via Web Browser		✓	✓
RStudio Professional Drivers	✓		✓
Connect to RStudio	✓		

# Requerimientos para el curso



RStudio Desktop 1.3.1093 - [Release Notes](#)

- 1.** Install R. RStudio requires [R 3.0.1+](#).
- 2.** Download RStudio Desktop. Recommended for your system:



**DOWNLOAD RSTUDIO FOR WINDOWS**  
1.3.1093 | 171.62MB

Requires Windows 10/8/7 (64-bit)

RStudio Desktop 1.3.1093 - [Release Notes](#)

- 1.** Install R. RStudio requires [R 3.0.1+](#).
- 2.** Download RStudio Desktop. Recommended for your system:



**DOWNLOAD RSTUDIO FOR MAC**  
1.3.1093 | 148.66MB

Requires macOS 10.13+ (64-bit)



# Requerimientos para el curso

## Abrir RStudio

The screenshot shows the RStudio interface. The top menu bar has tabs for 'Environment', 'History', 'Connections', and 'Presentation'. A red circle highlights the 'Environment' tab. The 'Global Environment' pane below it displays the message 'Environment is empty'. To the right, the word 'environment entorno' is written in red. The bottom navigation bar has tabs for 'Files', 'Plots', 'Packages', 'Help', and 'Viewer', with 'Packages' highlighted by a red circle. The 'R Script' pane on the left shows code related to natural language support and locale settings. The bottom status bar says 'output resultados'.

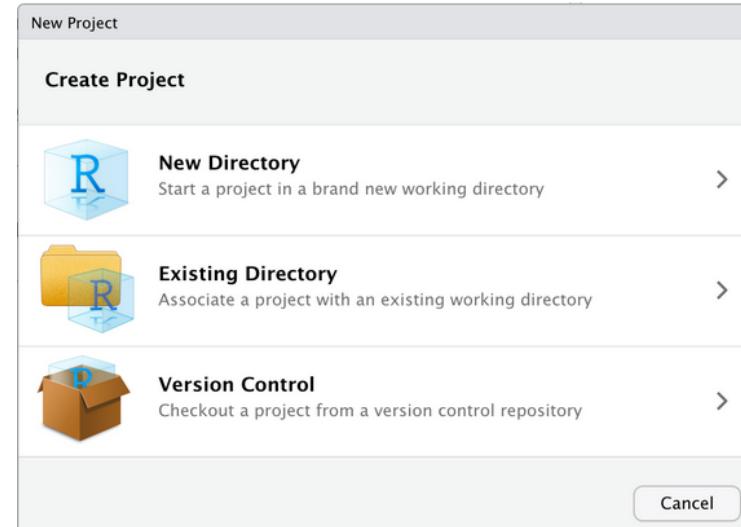
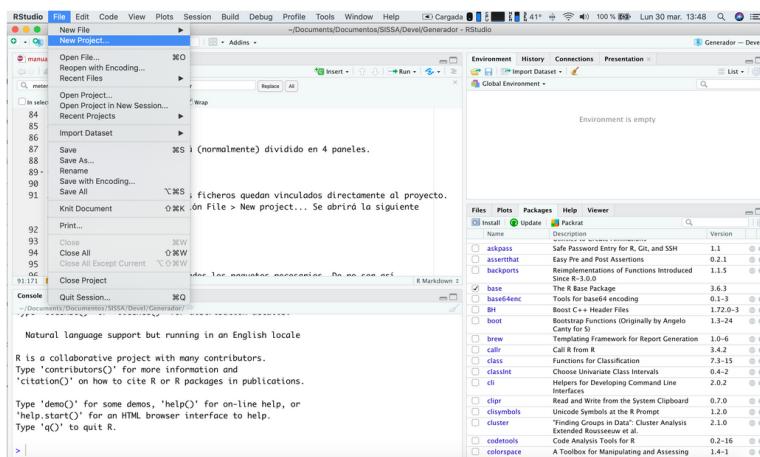
R Studio

The screenshot shows the RStudio interface with several panes:

- Code pane:** Displays R code for generating a scatter plot of diamond prices. The code uses ggplot2 to create a scatter plot of Price vs. Carat.
- Environment pane:** Shows the diamonds dataset with 53940 observations and 10 variables.
- Plots pane:** Displays a scatter plot titled "Diamond Pricing" showing the relationship between Price (Y-axis, 0 to 18,000) and Carat (X-axis, 0.0 to 3.5). The data points are colored by Clarity, with categories I1, SI2, SI1, VS1, VS2, VVS2, VVS1, and IF.

# Requerimientos para el curso

## Crear un proyecto



# Requerimientos para el curso

## Ejemplo de instalacion de paquetes

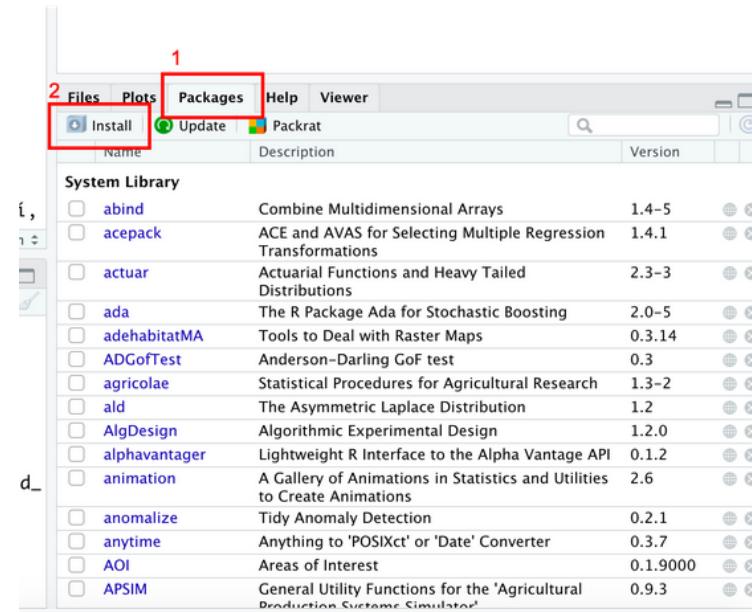
### Instalación de paquetes con código

```
#Instalar paquete
install.packages("<nombre del paquete>")
# Ejemplo1
install.packages("ggplot2")
# Instalar mas de un paquete a la vez
install.packages(c("<nombre del paquete 1>", "<nombre del paquete 2>"))
# Ejemplo2
install.packages(c("ggplot2", "dplyr"))
```

# Requerimientos para el curso

Instalar paquetes de forma visual en Rstudio

- 1.- hacer click en la pestaña **Packages** y seleccionar el paquete o buscarlo en la lupa
- 2.-Click en el botón **Install**.



## Tambien se puede tener una cuenta de RStudio Cloud

**R Studio Cloud** (<https://rstudio.cloud/>) es una plataforma web que permite crear, ejecutar y compartir proyectos implementados en lenguaje R sin necesidad de contar con ninguna infraestructura de hardware o software.

- 1) Ingrese a <https://rstudio.cloud/>
- 2) Hacer click en el link *Sign Up* que aparece en la parte superior derecha de la pantalla
- 3) Seleccione la opción *Cloud Free*
- 4) Haga click en el botón *Sign Up* que aparece en la parte inferior de la pantalla.
- 5) Complete el formulario y siga las instrucciones de registración.

PD1: Si se cuenta con una cuenta de *Google* o *GitHub*, se puede realizar una registración rápida utilizando alguna de estas cuentas (opciones *Sign Up with Google* o *Sign Up with GitHub*)

PD2: Cuentas gratuitas tienen ciertas limitaciones en cuanto a la cantidad de proyectos que puede crear y el tiempo disponible de ejecución sobre una base mensual --> Considerar para tareas y proyectos

## Crear una cuenta gratuita en google Colaboratory

<https://colab.research.google.com/notebooks/welcome.ipynb>



The screenshot shows the Google Colaboratory landing page. At the top, there's a navigation bar with links for Archivo, Editar, Ver, Insertar, Entorno de ejecución, Herramientas, and Ayuda. Below the navigation is a sidebar titled "Índice" containing links to Introducción, Ciencia de datos, Aprendizaje automático, Más recursos, Ejemplos de aprendizaje automático, and Sección. The main content area features a section titled "¿Qué es Colaboratory?" which explains that Colaboratory, also known as Colab, allows users to write and run Python code in a browser. It highlights three key features: no configuration required, free GPU access, and the ability to share. Below this, there's a "Introducción" section with a code cell that calculates the number of seconds in a day:

```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
86400
```

A note at the bottom of the code cell instructs users to click on the cell to select it and then press the play button to execute the code.

**Gracias por su atención**  
**Mail de contacto [malegriaa@udla.cl](mailto:malegriaa@udla.cl)**