

# Apertura



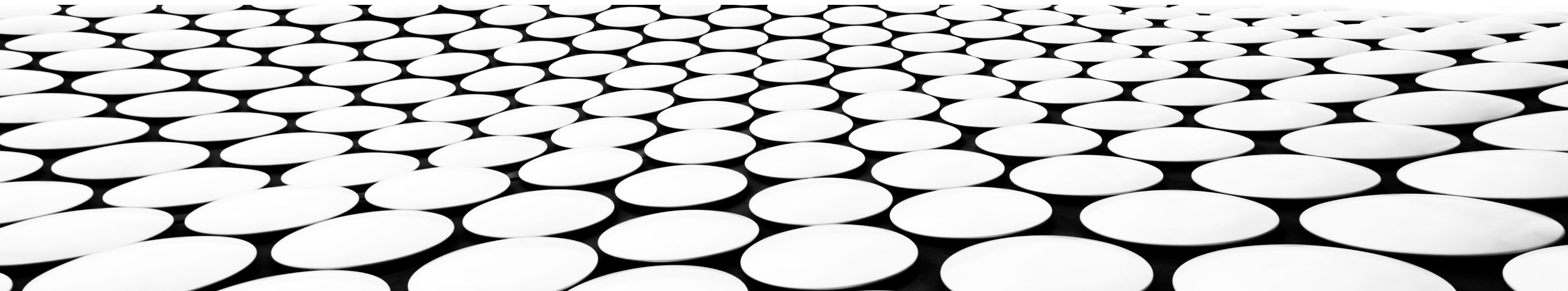
UNIVERSIDAD  
AUTÓNOMA  
METROPOLITANA  
Unidad Xochimilco



DEPARTAMENTO DE SISTEMAS BIOLÓGICOS  
ÁREA DE PRODUCTOS BIOLÓGICOS

# INTRODUCCIÓN A BIOPYTHON Y BIOSERVICES

Dr. Eduardo Zúñiga León



# Recomendaciones

## REQUERIMIENTOS PARA TRABAJO EN ZOOM

Conexión a Internet, al menos 20 MB de velocidad.  
Micrófono y cámara.  
Tener instalado Zoom.

## REQUISITOS COMPUTACIONALES

Conocimientos básicos de computación.  
Sistema operativo Windows (10), Mac, Linux (Ubuntu versión 14.04, 16.04, 18.04).  
6 GB de RAM o más, 50 GB de espacio en disco y al menos 4 núcleos (cores).

# Objetivo del curso

Sesión 1 (02-08-2021)

INTRODUCCIÓN A BIOPYTHON Y BIOSERVICES



**Proporcionar al participante las herramientas elementales del lenguaje de programación**

**Python para el análisis básico de secuencias y acceso a bases de datos usando las librerías de Biopython y el paquete Bioservices en el entorno interactivo Jupyter Notebook.**



# Objetivos del curso



- Desarrollar habilidades de programación.
- Reconocer las estructuras de datos usadas en Python.
- Implementar las funciones más relevantes de los módulos Biopython y Bioservices.
- Aprovechar todo el recurso depositado en bases de datos biológicas.
- Desarrollar visualizaciones estáticas e interactivas apropiadas para publicaciones científicas.



# Planeación

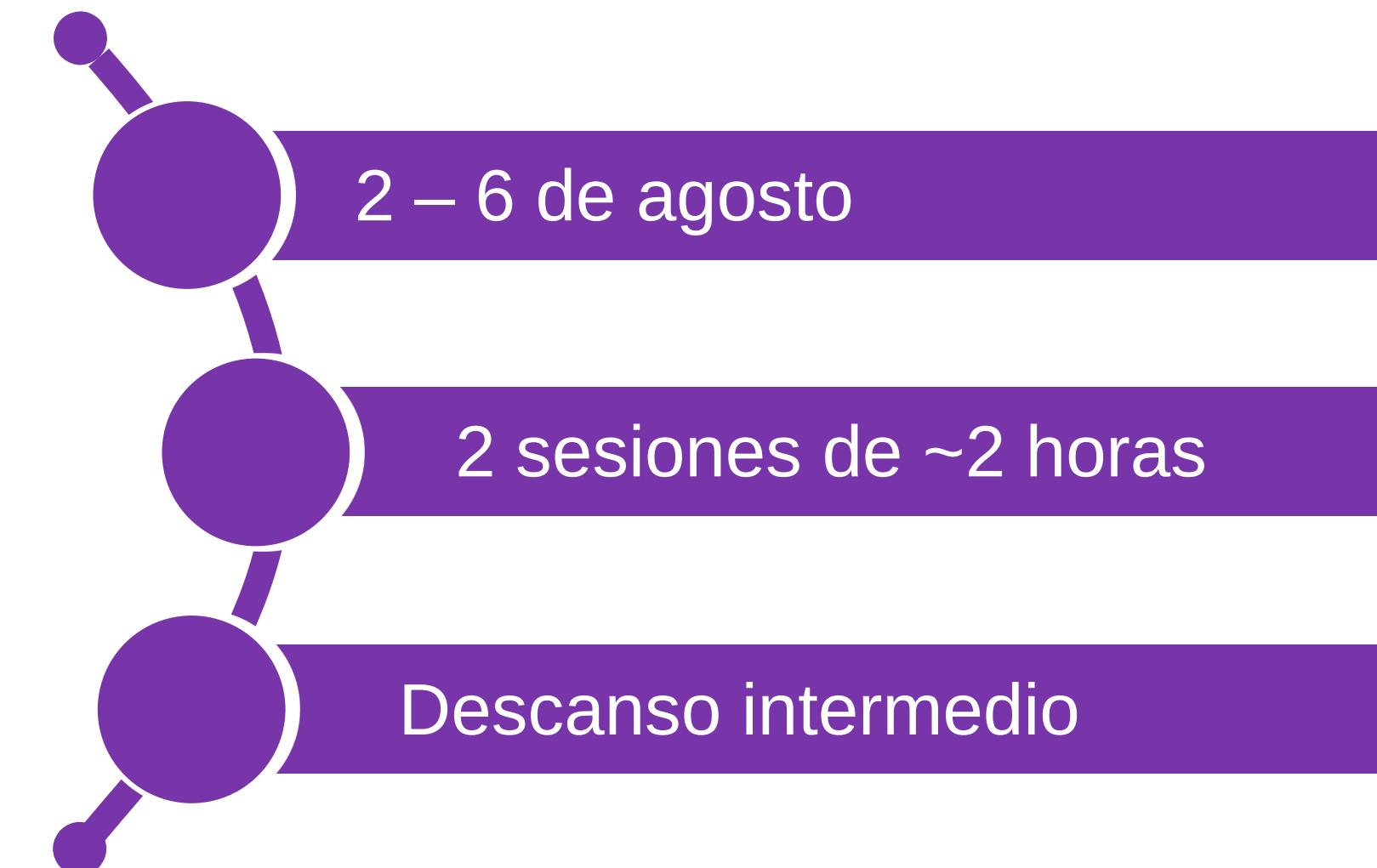


## DURACIÓN, FECHAS Y HORARIOS

Duración: 20 horas, cinco sesiones de 4 horas, con un receso de 20 minutos

Fechas: 2 al 6 de agosto de 2021.

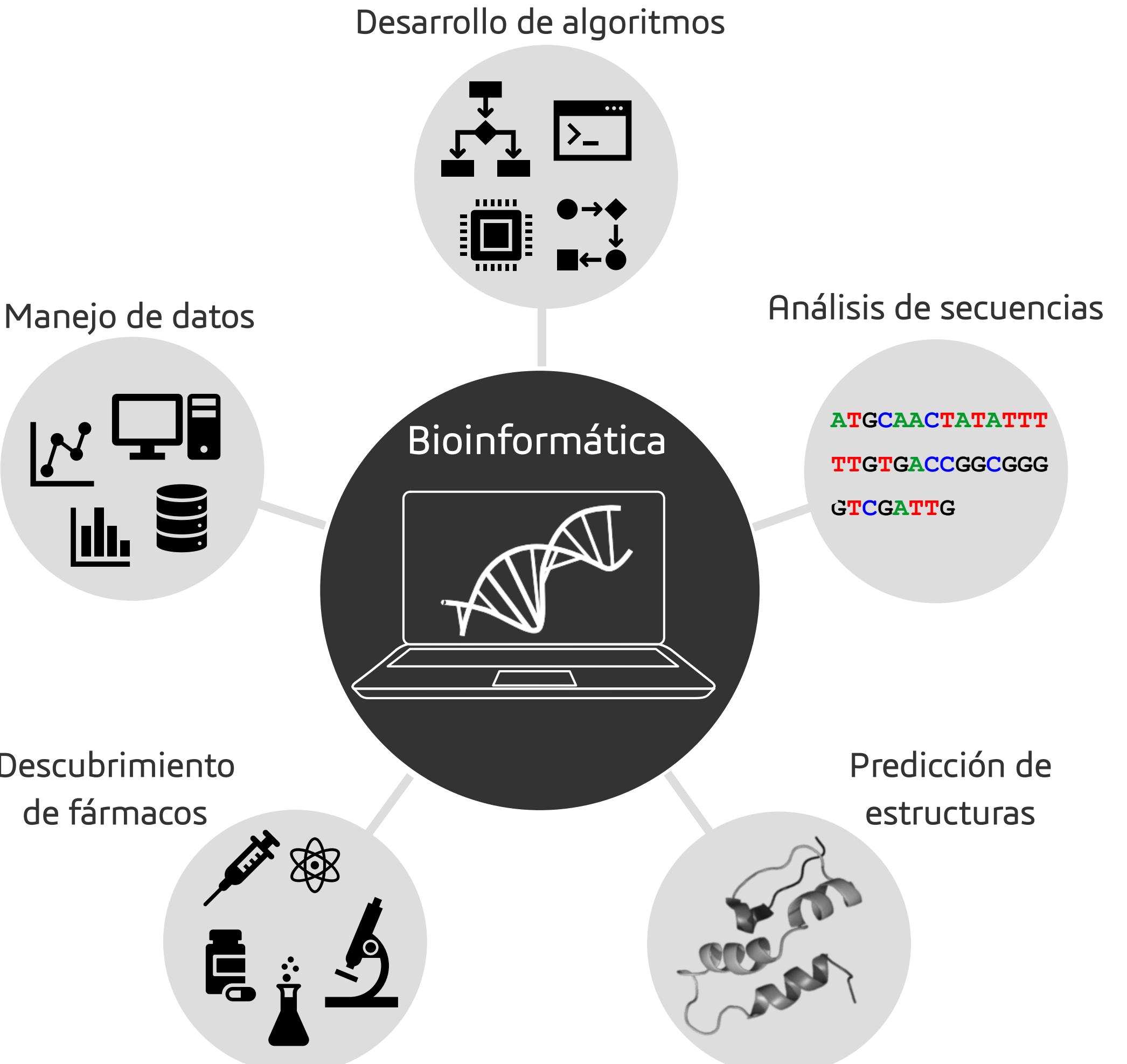
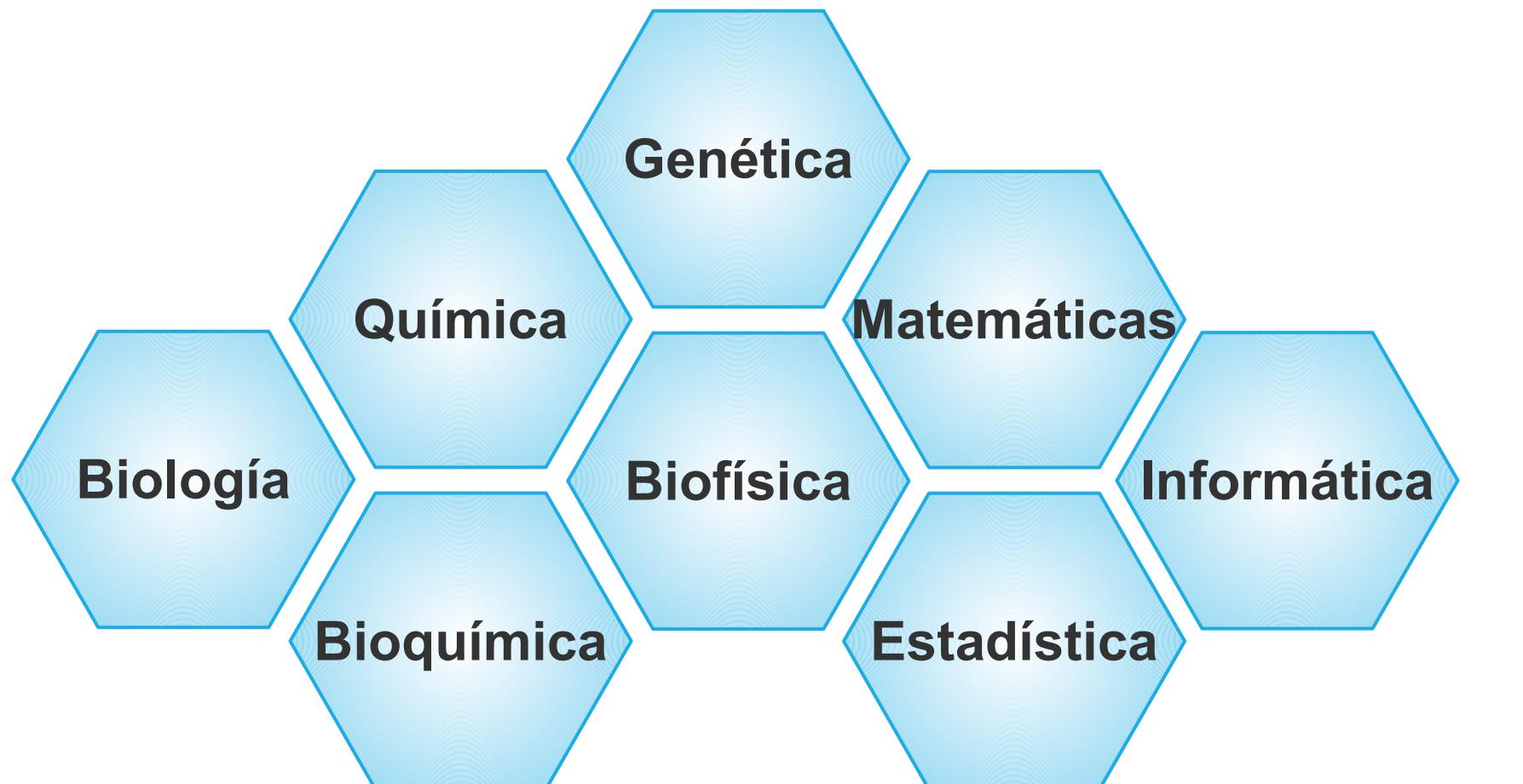
Horarios: de 16:00 a 20:00 horas.



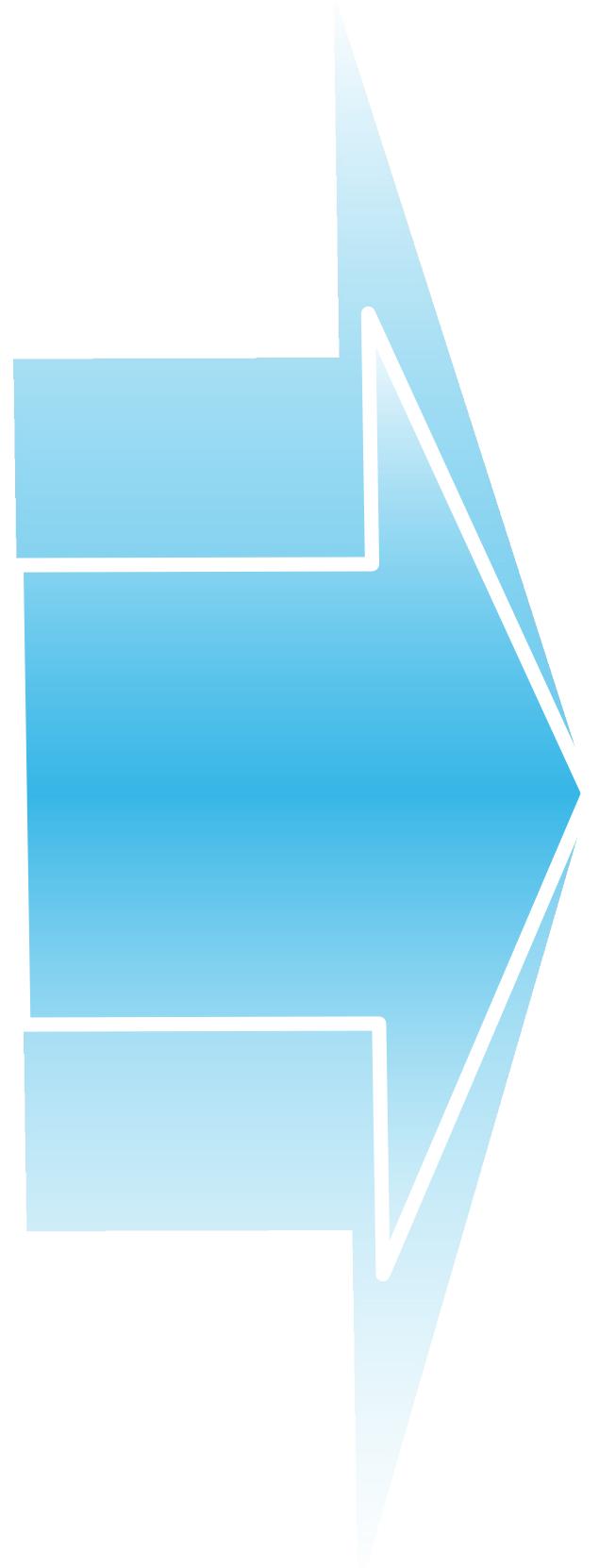
# Bioinformática

Es una disciplina científica **multidisciplinaria** encargada de la aplicación de herramientas computacionales para **organizar, analizar, interpretar, visualizar y almacenar información biológica** a gran escala.

Genet. Mol. Res, 2017;16(1): gmr16019645

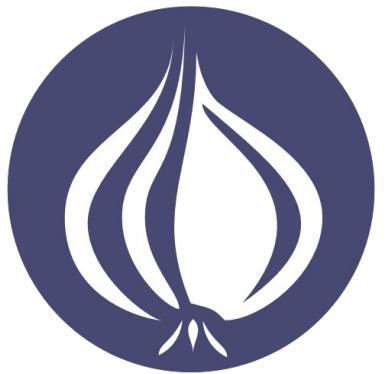


# Lenguajes



python

Java™



JS

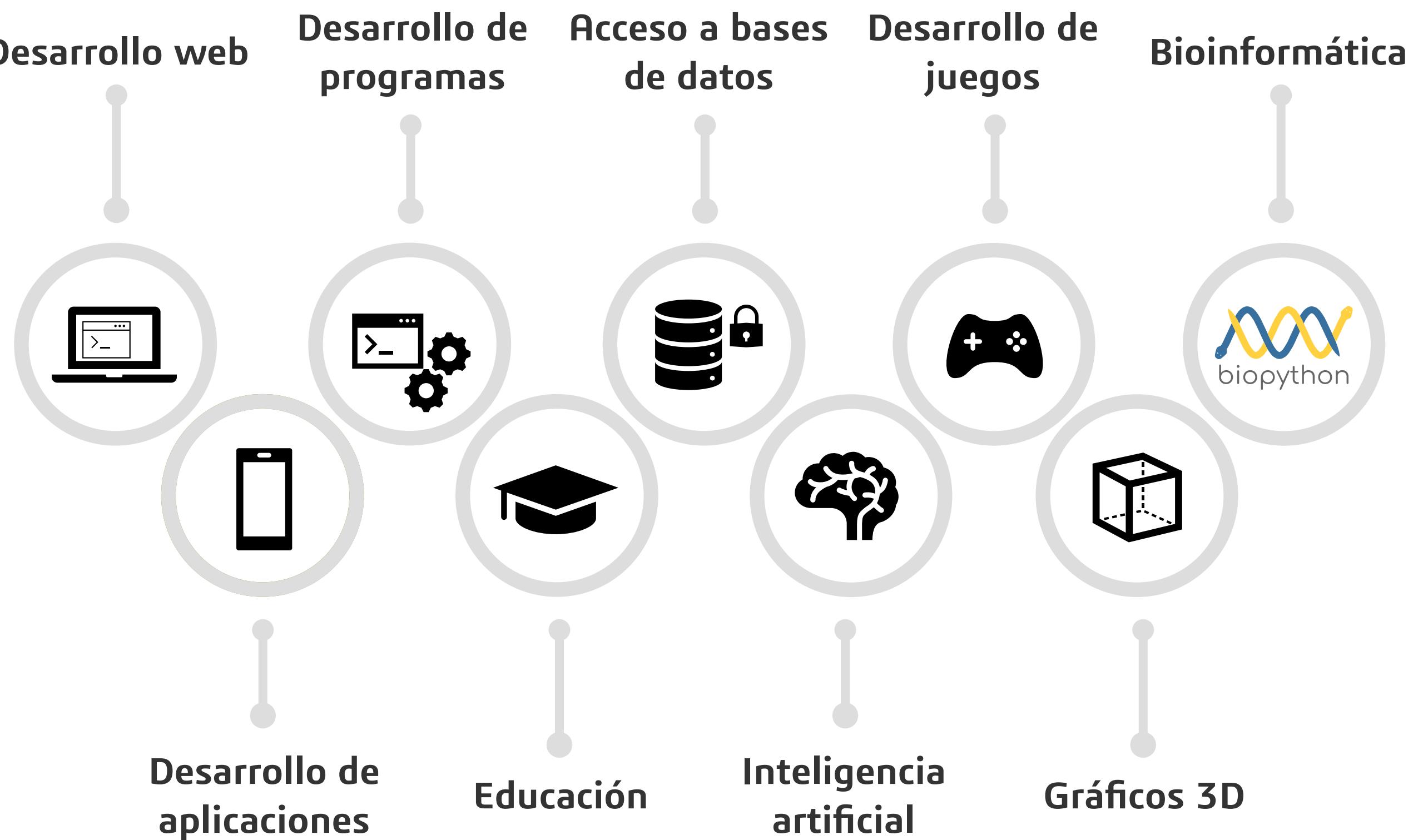
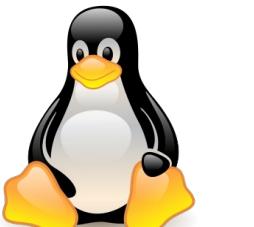
R

php



# Python

Propuesta general.  
Lenguaje de alto nivel.  
Tipado dinámico.  
Sintaxis sencilla.  
Altamente interactivo.  
Gran cantidad de bibliotecas.  
Código abierto.  
Aplicaciones científicas.  
Tiene una gran comunidad.



# Python

Journal List > PLoS Comput Biol > v.12(6); 2016 Jun > PMC4896647

**PLOS COMPUTATIONAL BIOLOGY**

PLoS Comput Biol. 2016 Jun; 12(6): e1004867.  
Published online 2016 Jun 7. doi: [10.1371/journal.pcbi.1004867](https://doi.org/10.1371/journal.pcbi.1004867)

**An Introduction to Programming for Bioscientists: A Python-Based Primer**

Berk Ekmekci,<sup>#</sup> Charles E. McAnany,<sup>#</sup> and Cameron Mura<sup>\*</sup>

Francis Ouellette, Editor

► Author information ► Copyright and License information Disclaimer

This article has been cited by other articles in PMC.

---

**Associated Data**

► Supplementary Materials

RESEARCH ARTICLE | JANUARY 01 2021

**Introducing Python Programming into Undergraduate Biology** FREE

Andrew A. David

*The American Biology Teacher* (2021) 83 (1): 33–41.  
<https://doi.org/10.1525/abt.2021.83.1.33>

Split-Screen Views PDF Share Tools

The rise of “big data” within the biological sciences has resulted in an urgent demand for coding skills in the next generation of scientists. To address this issue, several institutions and departments across the country have incorporated coding into their curricula. I describe a coding module developed and deployed in an undergraduate parasitology course, with the overarching goal of familiarizing students

Journal List > Front Genet > v.7; 2016 > PMC4843855



Front Genet. 2016; 7: 66.  
Published online 2016 Apr 26. doi: [10.3389/fgene.2016.00066](https://doi.org/10.3389/fgene.2016.00066)

**Book Review: Python Programming for Biology**

Reviewed by Alberto Marin-Sanguino\*

T. J. Stevens, W. Boucher, editors.  
Python Programming for Biology: Bioinformatics and Beyond.  
University Press. ISBN: 978-0-521-72009-0.

Article notes ► Copyright and License information Disclaimer



## ¿Qué podría hacer con Python?

- Análisis estadístico.
- Visualizaciones estáticas e interactivas.
- Edición de texto.
- Flujos de trabajo, por ejemplo protocolos de investigación, desarrollo de proyectos.
- Reportes.
- Procesamiento, análisis e interpretación de resultados de tesis.
- Apoyo académico.

# Python: sintaxis

Hace referencia al conjunto de reglas que definen cómo se tiene que escribir el código.



## Tipos de errores:

- errores de sintaxis,
- errores semánticos,
- errores de ejecución.

```
1 print 'sintaxis correcta'  
File "<ipython-input-28-273812d1371f>", line 1  
    print 'sintaxis correcta'  
^
```

**SyntaxError**: Missing parentheses in call to 'print'. Did you mean print('sintaxis correcta')?

```
1 print('sintaxis correcta')
```

sintaxis correcta

```
1 if 'A' == 'C':  
2     print('iguales')
```

```
File "<ipython-input-37-df137936708c>", line 2  
    print('iguales')  
^
```

**IndentationError**: expected an indented block

# Estructuras de datos

Números (*int* o *float*) : 1, 2, 3, 4, 5 ó 0.1, 3.5, 11.2

Cadenas de caracteres (*str*) : "ACTGACAGTCTCA"

Listas (*list*) : ["ACG", "CCT", "ATT", "GCT"]

Tuplas (*tuple*) : ("ACG", "CCT", "ATT", "GCT")

Diccionarios (*dict*) : {"A":123, "C":"GGG", "G":"--", "T":5}

# Estructuras de datos

## Arreglos (*array*)

```
>>> a[0,3:5]
```

```
array( [3,4] )
```

```
>>> a[4:, 4:]
```

```
array( [ 28, 29],  
      [ 34, 35] ])
```

```
>>> a[ :, 2]
```

```
array( [2, 8, 14, 20, 26, 32] )
```

```
>>> a[2 :: 2, :: 2]
```

```
array( [ 12, 14, 16],  
      [ 24, 26, 28] ])
```

Column names

## DataFrames

## Conjuntos (*set*)

{1, 2, 3}



set([3, 1, 2, 1])

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0	6-8	235.0	LSU	1170960.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0	6-2	190.0	Louisville	1824360.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN	6-9	260.0	Ohio State	2569260.0
6	Evan Turner	Boston Celtics	11.0	SG	27.0	6-7	220.0	Ohio State	3425510.0

# Jupyter Notebook

The logo for Jupyter Notebook features the word "jupyter" in a lowercase, sans-serif font. Above the letter "j", there is a thick orange semi-circle. Below the letter "p", there is a thin orange semi-circle. There are also two small dark gray circles, one above the "j" and one below the "p".

# Jupyter Notebook

Aplicación web interactiva para ciencia de datos y computación científica.

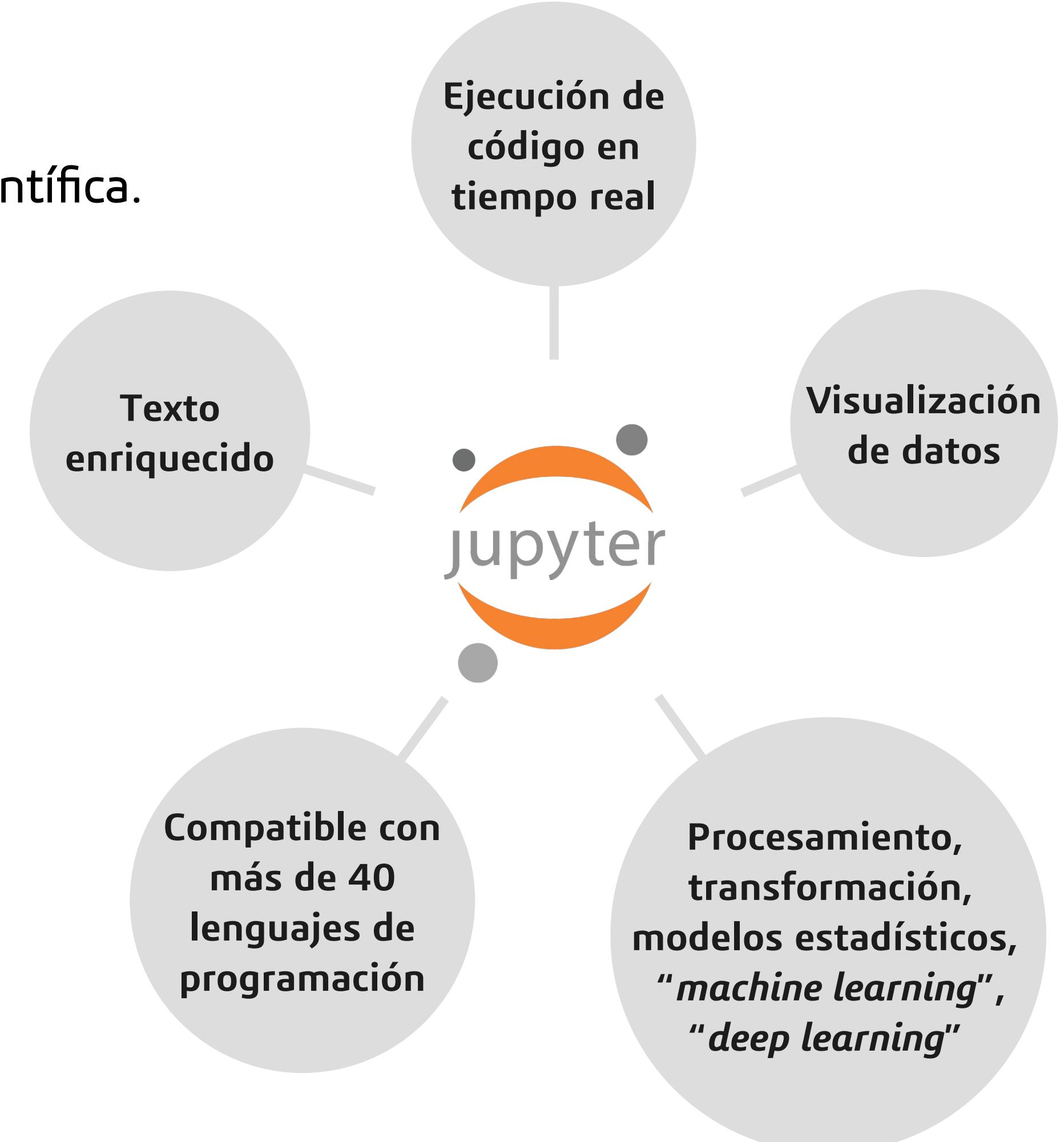
Es una versión de python enriquecida.

Integra bloques de código, texto, gráficos o imágenes.

Multipropósito.

**Julia + Python + R = Jupyter**

Google



# Jupyter Notebook

## Educación y ciencia

Aplicado en programas de enseñanza.

Permite el análisis de grandes volúmenes de datos, apoya la reproducibilidad y permite visualizaciones interactivas.

Los documentos ejecutables ofrecen una manera transparente y reproducible de comunicar los resultados de una investigación.

Sesión 1 (02-08-2021)

INTRODUCCIÓN A BIOPYTHON Y BIOSERVICES

## communications physics

Explore content ▾ Journal information ▾ Publish with us ▾

nature > communications physics > comment > article

Comment | Open Access | Published: 19 August 2020

### Creating an executable paper is a journey through Open Science

Jana Lasser 

Commu

5286 A

PLOS COMPUTATIONAL BIOLOGY



[PLoS Comput Biol.](#) 2020 Nov; 16(11): e1008326.

Published online 2020 Nov 5. doi: [10.1371/journal.pcbi.1008326](https://doi.org/10.1371/journal.pcbi.1008326)

PMCID: PMC7643937

PMID: 33151926

### Using interactive digital notebooks for bioscience and informatics education

[Alan Davies](#), \* [Frances Hooley](#), [Peter Causey-Freeman](#), [Iliada Eleftheriou](#), and [Georgina Moulton](#)

Francis Ouellette, Editor

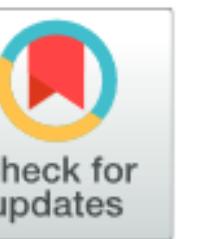
► Author information ► Copyright and License information [Disclaimer](#)

#### Abstract

Interactive digital notebooks provide an opportunity for researchers and educators to carry out data analysis and report the results in a single digital format. Further to just being digital, the format allows for rich content to be created in order to interact with the code and data contained in such a

## Retos actuales

- Implementación de documentos ejecutables en programas académicos.
- Inclusión de bases bioinformáticas en licenciaturas del área biológica.
- Dominio de al menos un lenguaje de programación (Python, R, Java, etc.).
- Nuevas tecnologías, algoritmos y enfoques informáticos.
- Producción masiva de datos.



### OPEN ACCESS

**Citation:** Armenta-Medina D, Díaz de León-Castañeda C, Valderrama-Blanco B (2020) Bioinformatics in Mexico: A diagnostic from the academic perspective and recommendations for a public policy. PLoS ONE 15(12): e0243531. <https://doi.org/10.1371/journal.pone.0243531>

**Editor:** Mariagrazia Benassi, University of Bologna, ITALY

**Received:** November 14, 2019

**Accepted:** November 24, 2020

**Published:** December 15, 2020

### RESEARCH ARTICLE

## Bioinformatics in Mexico: A diagnostic from the academic perspective and recommendations for a public policy

Dagoberto Armenta-Medina<sup>1,2</sup>\*, Christian Díaz de León-Castañeda<sup>1,2</sup>\*, Brenda Valderrama-Blanco<sup>3</sup>

<sup>1</sup> Dirección de Cátedras, CONACYT Consejo Nacional de Ciencia y Tecnología, Ciudad de México, México,

<sup>2</sup> INFOTEC Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación, Aguascalientes, México, <sup>3</sup> Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

\* These authors contributed equally to this work.

✉ Current address: Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

\* [dagoberto.armenta@infotec.mx](mailto:dagoberto.armenta@infotec.mx) (DAM); [cddeleon@conacyt.mx](mailto:cddeleon@conacyt.mx) (CDLC)

### Abstract

In this work, we present a diagnostic analysis of strengths, weaknesses, opportunities and threats (SWOT) of the current state of Bioinformatics in Mexico. We conducted semi-structured interviews among researchers and academics with key expertise in this field, identified by bibliometric analyses and qualitative sampling techniques. Additionally, an online survey was conducted reaching a higher number of respondents. Among the relevant findings of our study, the lack of specialized human resources and technological infrastructure stood out, along with deficiencies in the number and quality of academic programs, scarce public investment and a weak relationship between public and private institutions. However, there are great opportunities for developing a national Bioinformatics to support different economic sectors. In our opinion, this work could be useful to favor a comprehensive network among Mexican researchers, in order to lay the foundations of a national strategy towards a well designed public policy.

# Jupyter Notebook

Sesión 1 (02-08-2021)

INTRODUCCIÓN A BIOPYTHON Y BIOSERVICES

Google

Microsoft

IBM

Bloomberg

O'REILLY®

ANACONDA®

rackspace®  
the #1 managed cloud company

SOUND CLOUD

Quantopian

NetApp®

software  
carpentry

hhmi janelia  
Research Campus

<CODE NEURO>

N-Site LLC

COCALC

BRYN  
MAWR  
COLLEGE

CAL POLY  
SAN LUIS OBISPO

Berkeley  
UNIVERSITY OF CALIFORNIA

The  
University  
Of  
Sheffield.

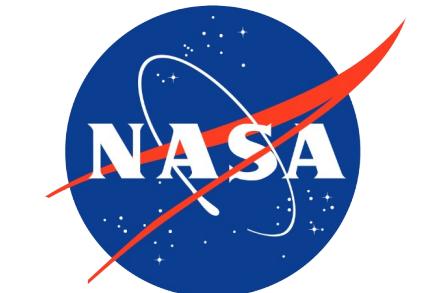
THE GEORGE  
WASHINGTON  
UNIVERSITY  
WASHINGTON, DC

CLEMSON  
UNIVERSITY

MICHIGAN STATE  
UNIVERSITY

Northwestern  
University

NYU



AYASDI

The Data Incubator

QuantStack  
Scientific Computing

# Jupyter Notebook



**Estos documentos ejecutables registran el desarrollo de todo un proceso y pueden ser compartidos.**

```

1 import pandas as pd
2 import numpy as np
3 import networkx as nx
4 import matplotlib.pyplot as plt

1 dbs = pd.read_csv('databases.csv', sep = '\t')

1 dbs.head(3)

```

Category	Dbs
0 Proteins	Protein Sequences
1 Proteins	Protein Domains and Families
2 Proteins	Protein Structures

```

1 F = nx.Graph()
2 F.add_edges_from([(i, j) for i, j in zip(dbs.Category, dbs.Dbs)])

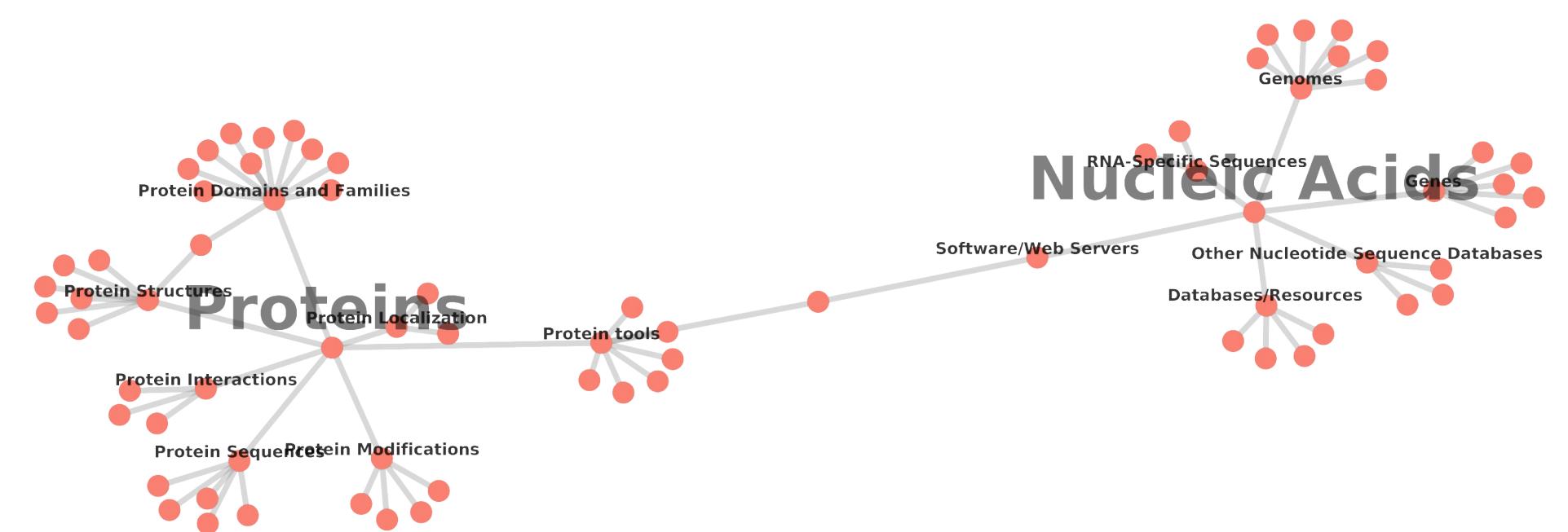
1 POSICIONES = nx.spring_layout(F, seed = 12)
2 POSICIONES = nx.spring_layout(F, k = 0.1, iterations=100, threshold=0.001
3
4 POS = {}
5 POS_LAB = {}
6 aumento_lab_y = 0
7 for i in F.nodes:
8     POS[i] = POSICIONES[i]
9     POS_LAB[i] = POSICIONES[i]

```

```

1 fig,ax = plt.subplots(figsize=(5, 15))
2 nx.draw_networkx_nodes(F, POS, node_color= 'salmon',
3                         alpha= 1,node_size = 100,
4                         zorder = 2)
5 nx.draw_networkx_edges(F, POS, width = 3,alpha= 0.3,edge_color= 'gray',
6 for i in POS_LAB:
7     if i in ['Protein Sequences','Protein Domains and Families','Protein
8         'Modifications','Protein Interactions','Protein Loca
9             'Protein tools',
10            'Genes','RNA-Specific Sequences','Software/Web Servers','Gen
11            'Other Nucleotide Sequence Databases','Databases/Resources']
12            ax.annotate(i, xy=POS_LAB[i], xycoords='data', rotation=0, zorde
13                        alpha = 0.8, fontweight='bold',
14                        xytext=POS_LAB[i], textcoords='data', ha="left", va =
15                            size = 8)
16        if i in ['Proteins', 'Nucleic Acids']:
17            ax.annotate(i, xy=POS_LAB[i], xycoords='data', rotation=0, zorde
18                        alpha = 0.5, fontweight='bold',
19                        xytext=POS_LAB[i], textcoords='data', ha="left", va =
20                            size = 30)
21    ax.axis('off')
22    plt.show()

```



# Jupyter Notebook

The screenshot shows the Jupyter Notebook interface with the following components:

- Toolbar:** Includes icons for File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a language selector set to Python 3.
- Menu Bar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Python 3.
- Toolbar Icons:** Save, New, Cut, Copy, Paste, Up, Down, Run, Cell Type, Cell Kernel, Code Cell, and Cell Kernel.
- Header:** jupyter Notebook Last Checkpoint: Yesterday at 2:14 AM (unsaved changes), Python logo, Logout.
- Cell Content:** Five input cells labeled In [ ]: 1.
- Code Cell Context Menu:** A tooltip titled "jupyter-notebook command group" is displayed, listing various cell modification commands. A red arrow points from the "Cell" menu icon in the toolbar to the start of this list.
- Text Overlay:** The word "Entorno" is overlaid in large black font across the bottom of the notebook area.

Action	Description	Mode
automatically indent selection		
change cell to code		(command mode) Y
change cell to heading 1		(command mode) 1
change cell to heading 2		(command mode) 2
change cell to heading 3		(command mode) 3
change cell to heading 4		(command mode) 4
change cell to heading 5		(command mode) 5
change cell to heading 6		(command mode) 6
change cell to markdown		(command mode) M
change cell to raw		(command mode) R
clear all cells output		
clear cell output		

# Instalaciones

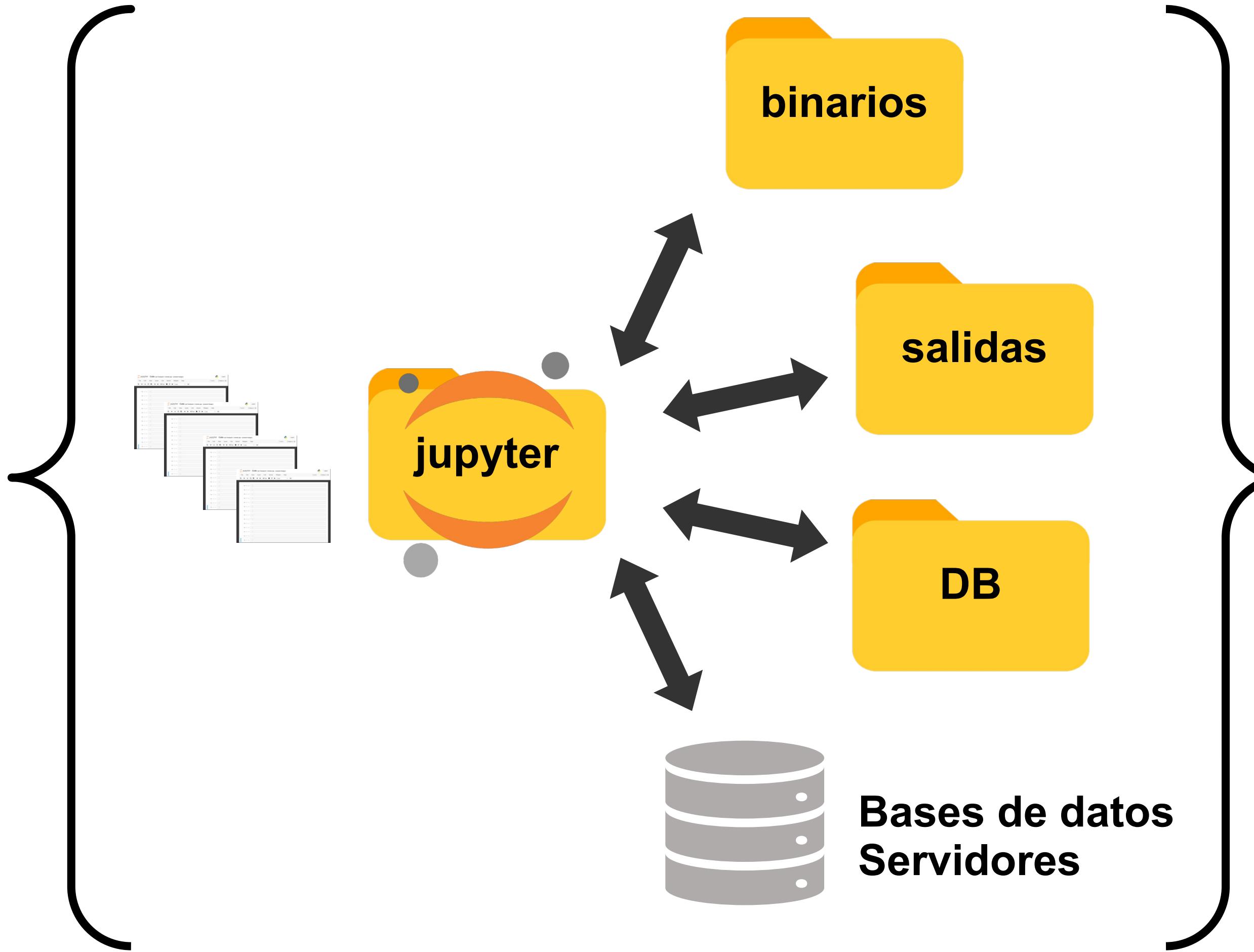
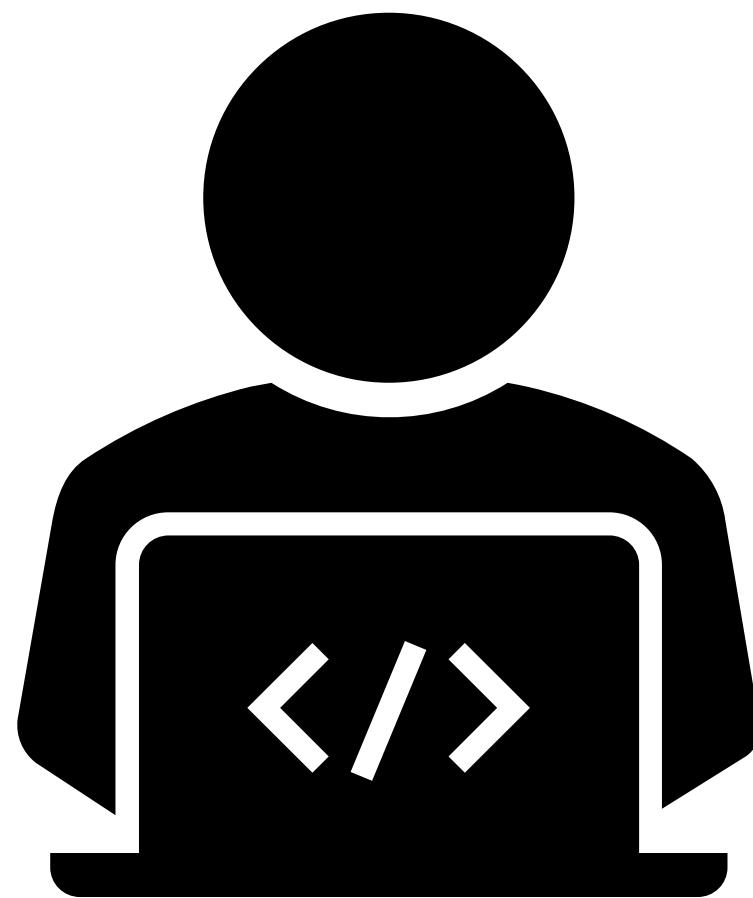
# Instalación de Python v3.6.7 en Windows

Descargar python-3.6.7-amd64

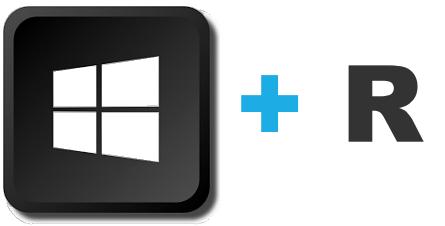
Ejecutar como administrador



# Ambiente de trabajo



# Instalación de Jupyter Notebook v6.0.3



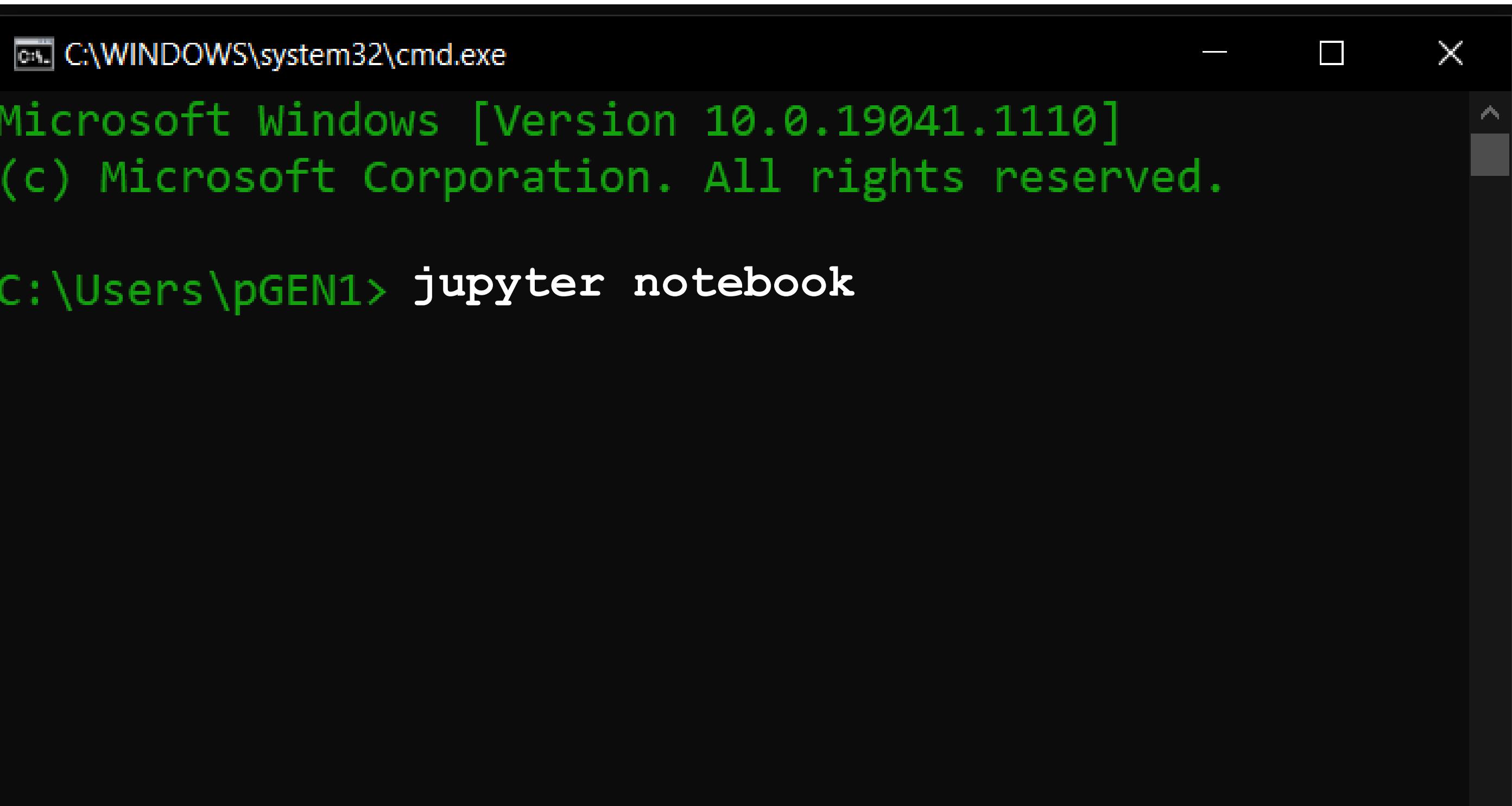
**cmd**

A screenshot of a Windows Command Prompt window titled 'C:\WINDOWS\system32\cmd.exe'. The window displays the following text:  
Microsoft Windows [Version 10.0.19041.1110]  
(c) Microsoft Corporation. All rights reserved.  
C:\Users\pGEN1> python -m pip install jupyter

# Ejecutar Jupyter Notebook

Windows icon + R

cmd



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.19041.1110]
(c) Microsoft Corporation. All rights reserved.

C:\Users\pGEN1> jupyter notebook
```