

# Python

## Bibliotecas para Ciencia de Datos



Rogelio Ferreira Escutia

Profesor / Investigador  
Tecnológico Nacional de México  
Campus Morelia



# Virtualización

# Virtualización

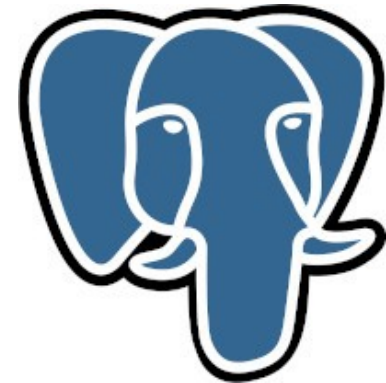
- **Aplicaciones:**
  - **VirtualBox**



# Almacenamiento de Información

# Almacenamiento de Información

- **Manejadores de Bases de Datos SQL**
  - SQLite
  - MariaDB (antes MySQL)
  - PostgreSQL



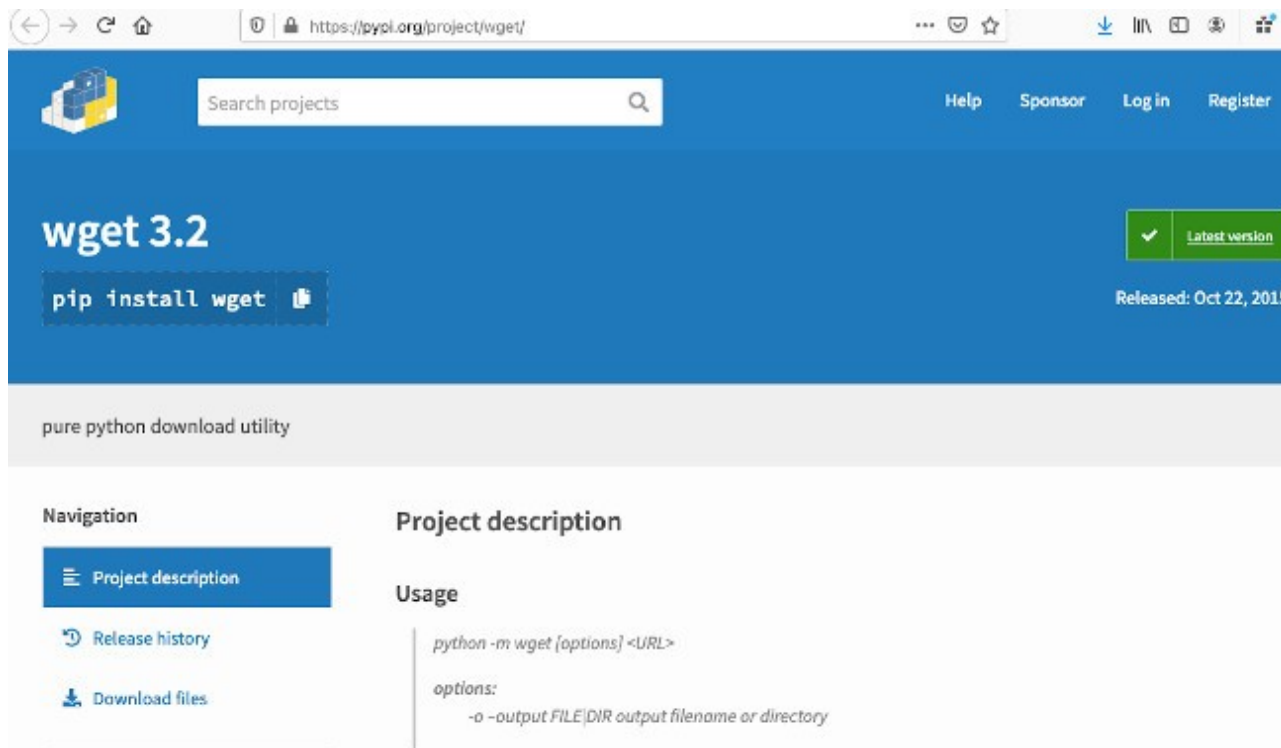
- **Sistemas de Bases de datos NoSQL**
  - MongoDB



# Manejo de imágenes

# wget

- Es una biblioteca para el descargar imágenes que se encuentren en internet.
- Fué realizada por anatoly techtonik bajo la licencia Public Domain



# Procesamiento de imágenes



# Procesamiento de Imágenes

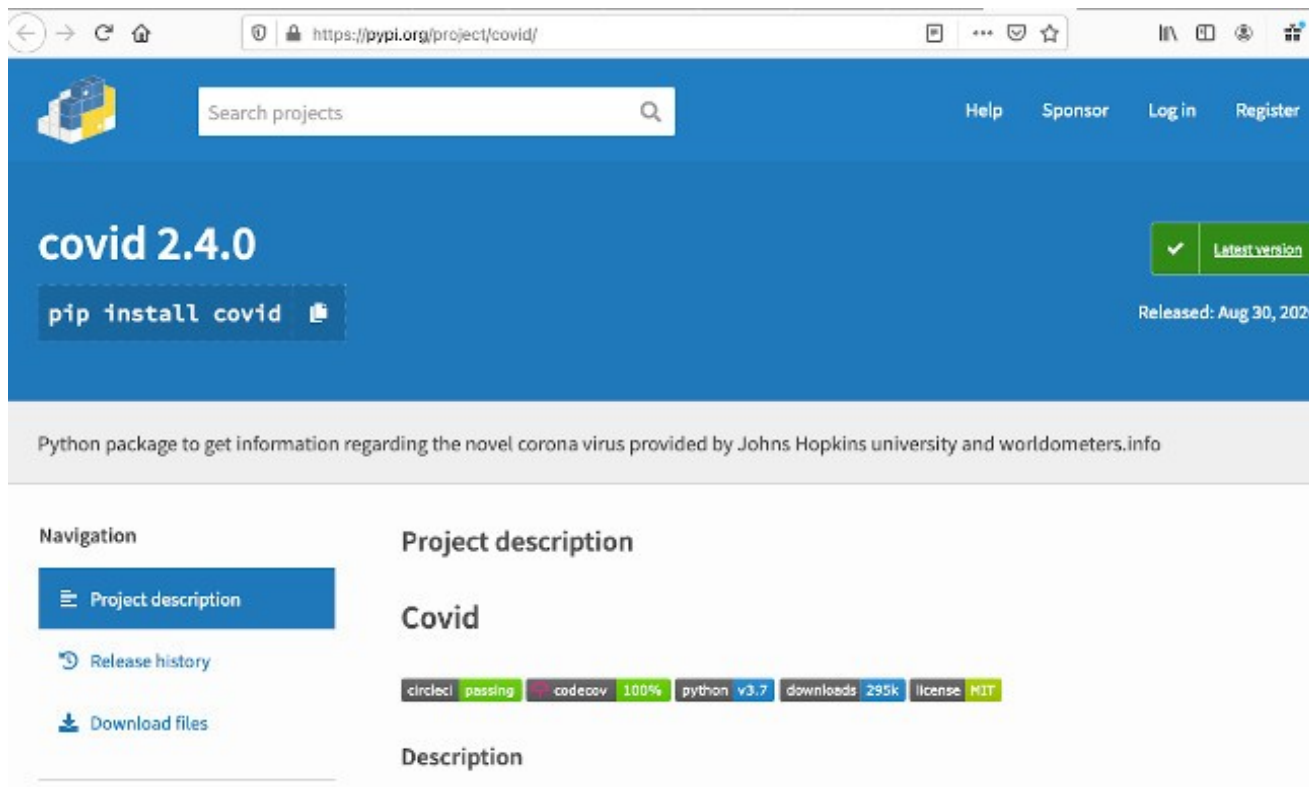
- **Bibliotecas:**
  - **OpenCV**



# Datasets

# Covid

- **Es una biblioteca para el manejo de datos con acceso de datasets con datos de Covid.**
- **Fu  realizada por Ahmed Nafies bajo la licencia del MIT**

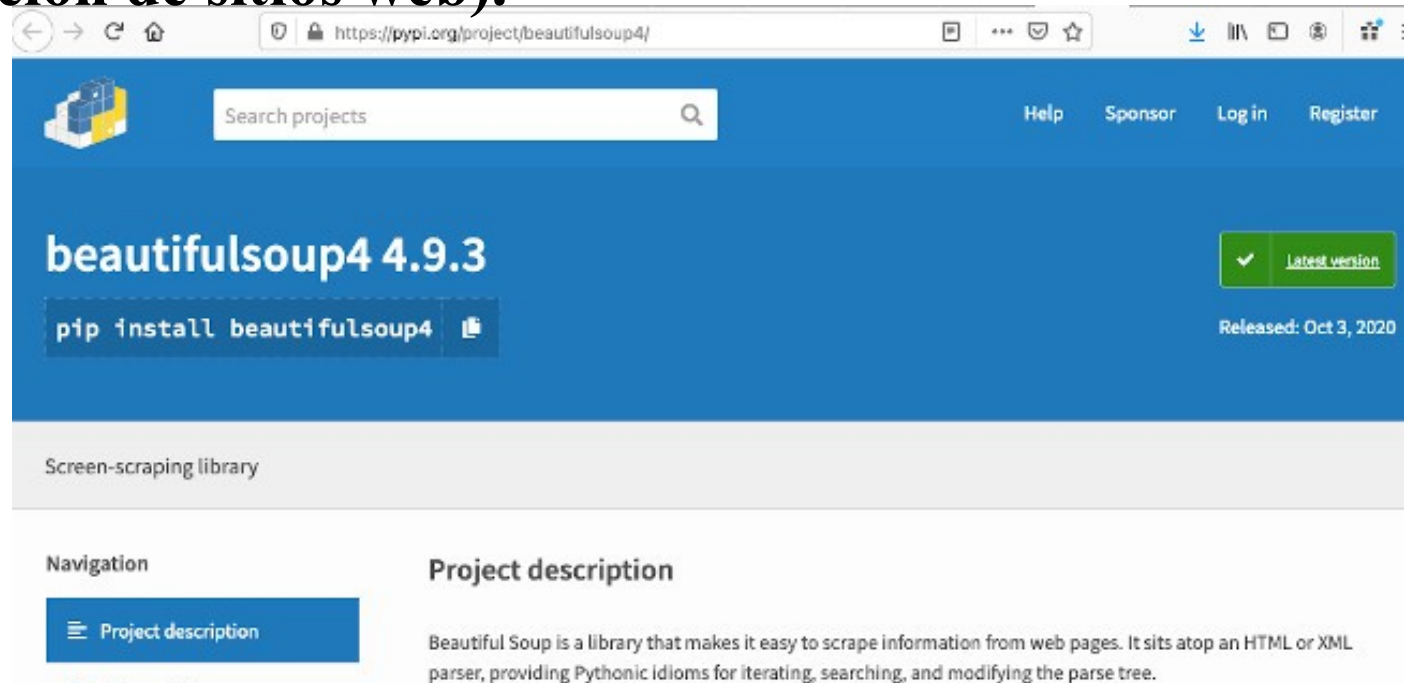


The image shows a screenshot of the PyPI project page for 'covid'. The browser address bar shows 'https://pypi.org/project/covid/'. The page features a blue header with a search bar, navigation links (Help, Sponsor, Log in, Register), and the project name 'covid 2.4.0'. A green button indicates it is the 'Latest version', and the release date is 'Released: Aug 30, 2020'. Below the header, there is a description: 'Python package to get information regarding the novel corona virus provided by Johns Hopkins university and worldometers.info'. The page is divided into two main sections: 'Navigation' on the left with links for 'Project description', 'Release history', and 'Download files'; and 'Project description' on the right, which includes a 'Covid' title and a row of badges for 'circleci passing', 'codecov 100%', 'python v3.7', 'downloads 295k', and 'license MIT'. The 'Description' section is partially visible at the bottom.

# Web Scraping

# Beautifulsoup

- Beautiful Soup es una biblioteca de Python para analizar documentos HTML (incluyendo los que tienen un marcado incorrecto).
- Esta biblioteca crea un árbol con todos los elementos del documento y puede ser utilizado para extraer información. Por lo tanto, esta biblioteca es útil para realizar web scraping (extraer información de sitios web).



The screenshot shows the PyPI project page for BeautifulSoup4 4.9.3. The page features a blue header with a search bar and navigation links (Help, Sponsor, Log in, Register). The main content area displays the project name 'beautifulsoup4 4.9.3' in large white text, with a green 'Latest version' badge. Below this is a dark blue button with the command 'pip install beautifulsoup4' and a package icon. The page is categorized as a 'Screen-scraping library'. A navigation menu on the left includes 'Project description' and 'Release history'. The 'Project description' section contains the text: 'Beautiful Soup is a library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.'

# Cálculo Numérico

# Numpy

- **NumPy proporciona una estructura de datos universal que posibilita el análisis de datos y el intercambio de datos entre distintos algoritmos.**
- **Las estructuras de datos que implementa son vectores multidimensionales y matrices con capacidad para gran cantidad de datos.**
- **Además, esta librería proporciona funciones matemáticas de alto nivel que operan en estas estructuras de datos.**



The fundamental package for scientific computing with Python

GET STARTED

## NumPy v1.20.0

Type annotation support - Performance improvements through multi-platform SIMD

### POWERFUL N-DIMENSIONAL ARRAYS

Fast and versatile, the NumPy vectorization, indexing, and broadcasting concepts are the de-facto standards of array computing today.

### NUMERICAL COMPUTING TOOLS

NumPy offers comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, and more.

### INTEROPERABLE

NumPy supports a wide range of hardware and computing platforms, and plays well with distributed, GPU, and sparse array libraries.

### PERFORMANT

The core of NumPy is well-optimized C code. Enjoy the flexibility of Python with the speed of compiled code.

### EASY TO USE

NumPy's high level syntax makes it accessible and productive for programmers from any background or experience level.

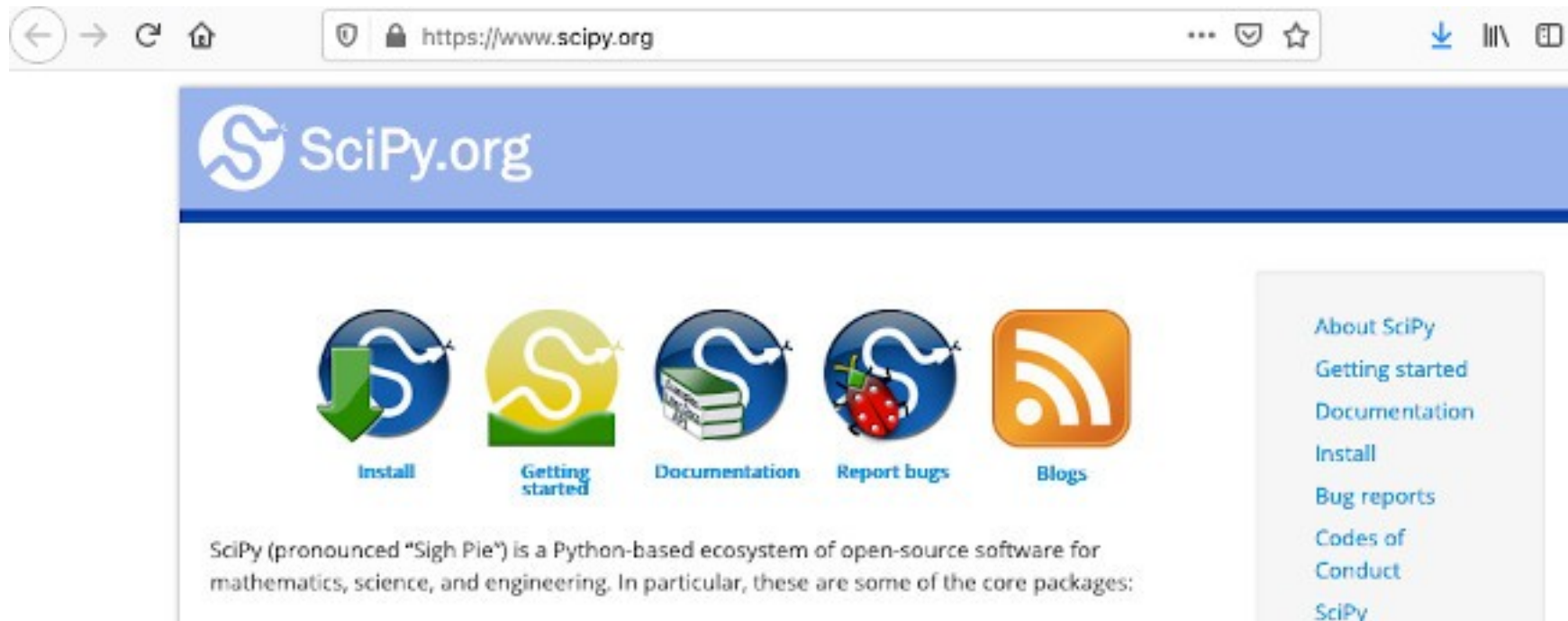
### OPEN SOURCE

Distributed under a liberal BSD license, NumPy is developed and maintained publicly on GitHub by a vibrant, responsive, and diverse community.



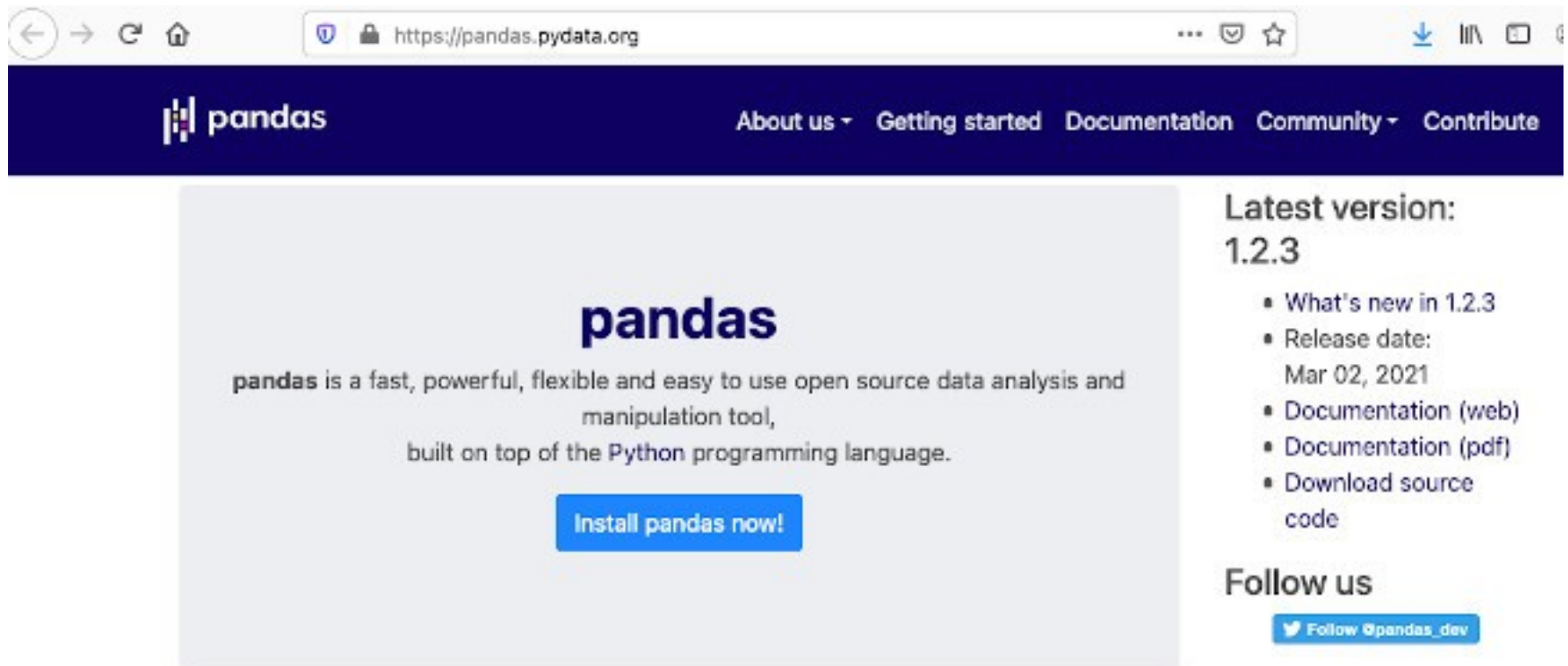


- **SciPy es una biblioteca libre y de código abierto para Python. Se compone de herramientas y algoritmos matemáticos.**
- **Se creó a partir de la colección original de Travis Oliphant, que se componía de módulos de extensión para Python y fue lanzada en 1999 bajo el nombre de Multipack.**



# Pandas

- Es una biblioteca escrita para Python para el análisis y manipulación de datos.
- Está orientado especialmente para el manejo numérico de tablas y series de tiempos.



The image shows a screenshot of the pandas website homepage. The browser address bar displays "https://pandas.pydata.org". The website has a dark blue header with the pandas logo and navigation links: "About us", "Getting started", "Documentation", "Community", and "Contribute". The main content area features the pandas logo, a description: "pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.", and a blue button that says "Install pandas now!". On the right side, there is a section for the "Latest version: 1.2.3" with a list of links: "What's new in 1.2.3", "Release date: Mar 02, 2021", "Documentation (web)", "Documentation (pdf)", and "Download source code". Below this is a "Follow us" section with a button to "Follow @pandas\_dev" on Twitter.

# Machine Learning

# scikit-learn

- **Scikit-learn es una biblioteca para aprendizaje automático de software libre para el lenguaje de programación Python.**
- **Incluye varios algoritmos de clasificación, regresión y análisis de grupos entre los cuales están máquinas de vectores de soporte, bosques aleatorios, Gradient boosting, K-means y DBSCAN.**
- **Está diseñada para interoperar con las bibliotecas numéricas y científicas NumPy y SciPy.**

# scikit-learn

Machine Learning in Python

[Getting Started](#)
[Release Highlights for 0.24](#)
[GitHub](#)

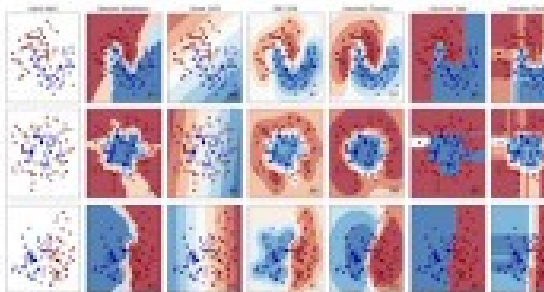
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

## Classification

Identifying which category an object belongs to.

**Applications:** Spam detection, image recognition.

**Algorithms:** SVM, nearest neighbors, random forest, and more...

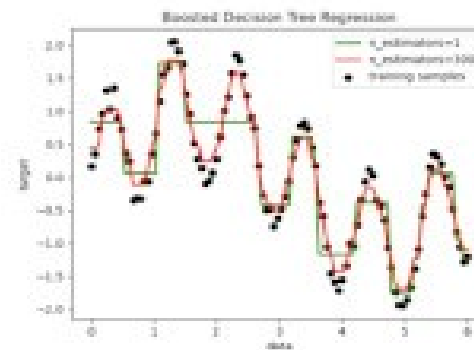


## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** SVR, nearest neighbors, random forest, and more...



## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** k-Means, spectral clustering, mean-shift, and more...



# Deep Learning

# TensorFlow

- **TensorFlow es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos.**
- **Actualmente es utilizado tanto en la investigación como en los productos de Google.**

Google se compromete a impulsar la igualdad racial para las comunidades afrodescendientes. [Obtén información al respecto.](#)

# Una plataforma de extremo a extremo de código abierto para el aprendizaje automático

TensorFlow

Para JavaScript

Para dispositivos móviles y de IoT

Para produ...

La principal biblioteca de código abierto para enseñarte a desarrollar y entrenar modelos de AA. Comienza enseguida y ejecuta notebooks de Colab directamente en tu navegador.

Comienza a usar TensorFlow





# Lenguaje Natural

# NLTK

- **El kit de herramientas de lenguaje natural, o más comúnmente NLTK, es un conjunto de bibliotecas y programas para el procesamiento del lenguaje natural (PLN) simbólico y estadísticos para el lenguaje de programación Python.**
- **NLTK incluye demostraciones gráficas y datos de muestra. Se acompaña de un libro que explica los conceptos subyacentes a las tareas de procesamiento del lenguaje compatibles el toolkit, además de programas de ejemplo.**

# NLTK 3.5 documentation

[NEXT](#) | [MODULES](#) | [INDEX](#)

## Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](#) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active [discussion forum](#).

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called "a wonderful tool for teaching, and working in, computational linguistics using Python," and "an amazing library to play with natural language."

### TABLE OF CONTENTS

[NLTK News](#)

[Installing NLTK](#)

[Installing NLTK Data](#)

[Contribute to NLTK](#)

[FAQ](#)

[Wiki](#)

[API](#)

[HOWTO](#)

### SEARCH

---

# spaCy

- spaCy es una librería open source para procesamiento de lenguaje natural escrita para Python y Cython.
- La biblioteca es publicada bajo la licencia del MIT y desarrollada por Matthew Honnibal y Ines Montani.

The screenshot shows the spaCy documentation website. The header includes the spaCy logo, a notification for 'Out now: spaCy v3.0', and navigation links for 'USAGE', 'MODELS', 'API', and 'UNIVERSE'. A search bar is located in the top right corner. The main content area is titled 'Install spaCy' and features a configuration panel with the following options:

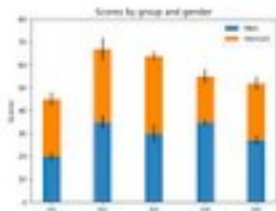
- Operating system: macOS / OSX (selected), Windows, Linux
- Package manager: pip (selected), conda, from source
- Hardware: CPU (selected), GPU
- Configuration:  virtual env,  train models
- Trained pipelines:  Chinese,  Danish,  Dutch,  English,  French,  German

A dark blue callout box on the right side of the page contains the text: 'LOOKING FOR THE OLD DOCS? To help you make the transition from v2.x to v3.0, we've uploaded the old website to [v2.spacy.io](https://v2.spacy.io). To see what's changed and how to migrate, see the [v3.0 guide](#).'

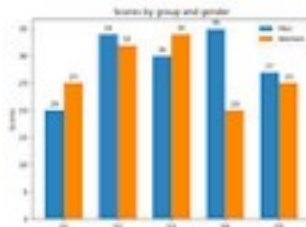
# Visualización de Datos

# Matplotlib

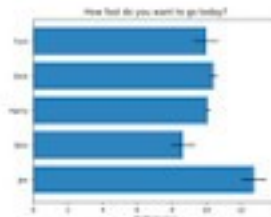
- **Matplotlib es la librería gráfica de python estándar y la más conocida.**
- **Puedes usar matplotlib para generar gráficos de calidad necesaria para publicarlas tanto en papel como digitalmente.**
- **Con matplotlib puedes crear muchos tipos de gráficos: series temporales, histogramas, espectros de potencia, diagramas de barras, diagramas de errores, etc.**



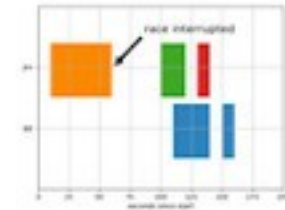
Stacked Bar Graph



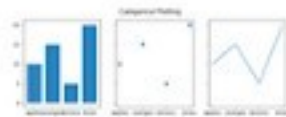
Grouped bar chart with labels



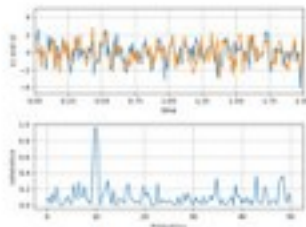
Horizontal bar chart



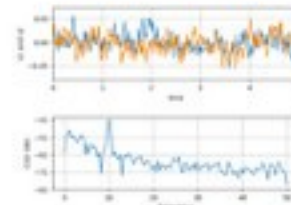
Broken Barh



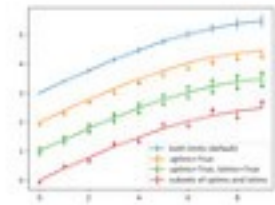
Plotting categorical variables



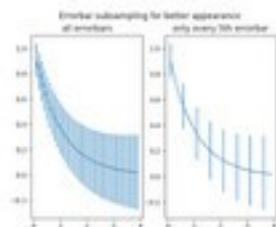
Plotting the coherence of two signals



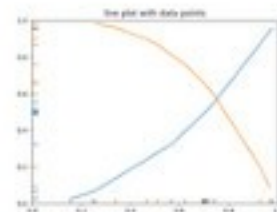
CSD Demo



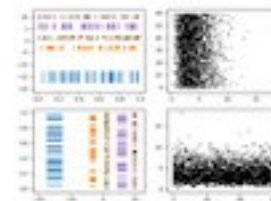
Errorbar limit selection



Errorbar Subsample



EventCollection Demo



Eventplot Demo



Filled polygon

# matplotlib

Version 3.1.1

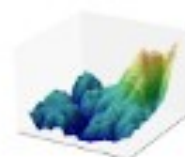
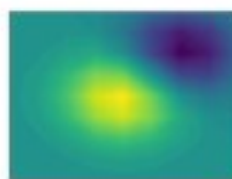
Fork me on GitHub

[Installation](#) | [Documentation](#) | [Examples](#) | [Tutorials](#) | [Contributing](#)

[home](#) | [contents](#) »

[modules](#) | [index](#)

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.



Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For examples, see the [sample plots](#) and [thumbnail gallery](#).

For simple plotting the `pyp1ot` module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Quick search

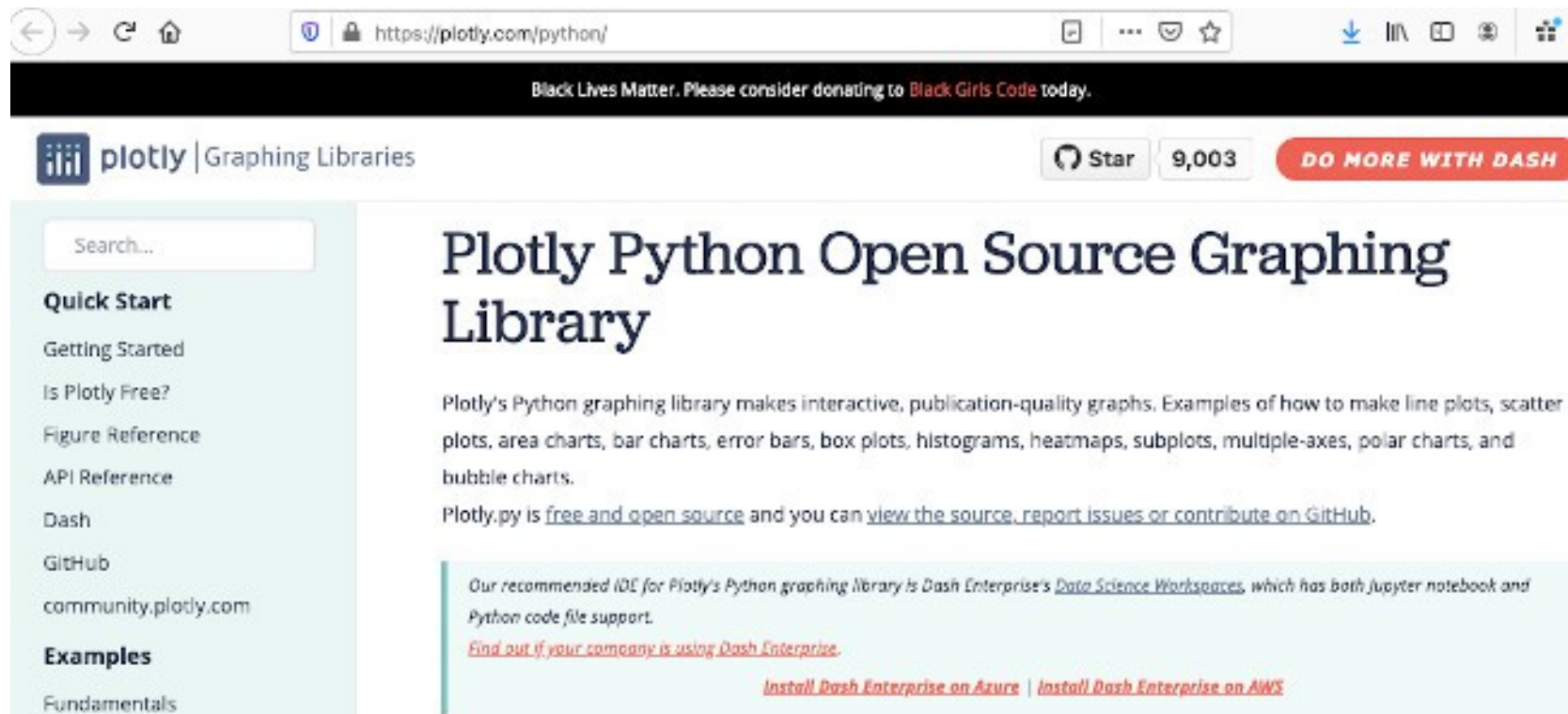
Matplotlib 3.0 is Python 3 only.  
For Python 2 support, Matplotlib 2.2.x will be continued as a LTS release and updated with bugfixes until January 1, 2020.

Support Matplotlib



# Plotly

- **Plotly es una compañía de Montreal Canadá que desarrolla herramienta para análisis y visualización de datos.**
- **Desarrollan bibliotecas para análisis de datos en Python, R, MATLAB, Perl, Julia, Arduino y REST.**



The screenshot shows the Plotly website for Python. The browser address bar displays "https://plotly.com/python/". A black banner at the top contains the text "Black Lives Matter. Please consider donating to Black Girls Code today." Below this, the Plotly logo and "Graphing Libraries" are visible, along with a "Star" button showing "9,003" and a red button that says "DO MORE WITH DASH". The main heading is "Plotly Python Open Source Graphing Library". The introductory text states: "Plotly's Python graphing library makes interactive, publication-quality graphs. Examples of how to make line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heatmaps, subplots, multiple-axes, polar charts, and bubble charts." Below this, it says "Plotly.py is [free and open source](#) and you can [view the source](#), [report issues](#) or [contribute on GitHub](#)." A light blue box contains a recommendation: "Our recommended IDE for Plotly's Python graphing library is Dash Enterprise's [Data Science Workspaces](#), which has both Jupyter notebook and Python code file support." Below this box are two links: "Find out if your company is using Dash Enterprise" and "Install Dash Enterprise on Azure | Install Dash Enterprise on AWS". On the left side, there is a navigation menu with a search bar and links for "Quick Start", "Getting Started", "Is Plotly Free?", "Figure Reference", "API Reference", "Dash", "GitHub", "community.plotly.com", "Examples", and "Fundamentals".

# Páginas Web

- **Lenguajes y Herramientas:**
  - HTML
  - CSS
  - JS
  - PHP

**HTML**





## Rogelio Ferreira Escutia

Profesor / Investigador  
Tecnológico Nacional de México  
Campus Morelia



[rogelio.fe@morelia.tecnm.mx](mailto:rogelio.fe@morelia.tecnm.mx)



[rogeplus@gmail.com](mailto:rogeplus@gmail.com)



[xumarhu.net](http://xumarhu.net)



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/  
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

