“

most of the good programmers do
programming not because they expect
to get paid or get adulation by the public,
but because it is fun to program.

“

Linus Torvalds

Open Source Computing

2024-10-10

Leibniz Institute for Neurobiology
Magdeburg

Aryo Zare

Open Source Software
(Free Software)

Open Source Software (Free Software)

- Free in free software does not mean free as in having no price, but rather free as in “liberty”.

1. The freedom to run the program, for any purpose.
2. The freedom to study how the program works, and adapt it to your needs (access to the source code).
3. Freedom to redistribute copies.
4. The freedom to improve the program, and release improvements to the public, so that the whole community benefits.

(github fork & pull-request)

- *The freedom to play with the software code.*
- *The freedom to nurture your curiosity.*

back to 30 years ago :
Why was open source software
not popular ?

Open Source Initiative

- Internet
  - Github
    - Time needed to develop full-featured libraries.

How Open-Source Software is Eating the World

1983

Richard Stallman announces GNU Project aimed at creating free operating system.

1985

Free Software Foundation created to support open source movement and enforce license requirements.

Laying the Foundation

1991

Linux kernel is released under GNU General Public License, explicitly granting end-users the right to run, modify and share software.

1998

IBM announces WebSphere, an early enterprise web server based on Apache open-source software.

1999

Red Hat goes public in one of the first IPOs for an open source company.

Apache Software Foundation created in partnership with IBM and leading technology firms to enable developers to better collaborate on open-source projects.

2001

IBM invests $1 billion into Linux, helping launch it into the commercial mainstream.

IBM helps establish the Eclipse Foundation, with the Eclipse IDE Framework quickly becoming the world's most popular Java development environment.

Scaling the Technology

2002

MediaWiki engine launches, laying foundation for Wikipedia.

2003

Mozilla Project releases Firefox, first major alternative to proprietary Windows Explorer.

2005

Linus Torvalds releases Git distributed revision control system.

2007

The Linux Foundation is established through a consortium of organizations including IBM.

Entering the Commercial Mainstream

2008

Google releases Android operating system, now the world's most used mobile operating system.

2009

MongoDB launches, quickly becoming among the most popular NoSQL open-source databases in the world.

2010

In collaboration with 25 technology companies including IBM, Rackspace and NASA launch OpenStack, an open-source platform for cloud computing.

2013

OpenPOWER foundation formed as an independent organization to foster innovation around Power chips.

2015

TensorFlow released under Apache 2.0 open-source license, helping spur machine learning development.

2016

IBM helps Linux Foundation establish Hyperledger in one of first initiatives to build an open-source enterprise blockchain platform.

IBM helps launch Cloud Native Computing Foundation to support open governance around Kubernetes.

2017

In one of the first publicly collaborative quantum computing projects, IBM open sources the Qiskit developer toolkit.

Cloudera debuts publicly in highly-anticipated IPO of an open source company.

2018

IBM announces plans to purchase Red Hat, which will accelerate the impact of open-source development in the cloud.

Since the advent of modern computing, the ability of passionate programmers to collaborate on open source code has led to some of the most important software breakthroughs. Today, open-source software is more important than ever. It powers our web browsers, databases, operating systems and smartphones. Peek behind the curtain of any new innovation, and chances are you'll find a foundation built on decades of open-source contributions.

Developmental milestones

https://www.ibm.com/topics/open-source

How Open-Source Software is Eating the World

1983

Richard Stallman announces GNU Project aimed at creating free operating system.

Laying the Foundation

1985

Free Software Foundation created to support open source movement and enforce license requirements.

1991

Linus Torvalds

Linux kernel is released under GNU General Public License, explicitly granting end-users the right to run, modify and share software.

1998

IBM announces WebSphere, an early enterprise web server based on Apache open-source software.

1999

Red Hat goes public in one of the first IPOs for an open source company.

initial public offering

Apache Software Foundation created in partnership with IBM and leading technology firms to enable developers to better collaborate on open-source projects.

2001

IBM invests $1 billion into Linux, helping launch it into the commercial mainstream.

IBM helps establish the Eclipse Foundation, with the Eclipse IDE Framework quickly becoming the world's most popular Java development environment.

Scaling the Technology

2002

MediaWiki engine launches, laying foundation for Wikipedia.

2003

Mozilla Project releases Firefox, first major alternative to proprietary Windows Explorer.

2005

Linus Torvalds releases Git distributed revision control system.

2007

The Linux Foundation is established through a consortium of organizations including IBM.

Entering the Commercial Mainstream

2008

Google releases Android operating system, now the world's most used mobile operating system.

2009

MongoDB launches, quickly becoming among the most popular NoSQL open-source databases in the world.

2010

openstack.

In collaboration with 25 technology companies including IBM, Rackspace and NASA launch OpenStack, an open-source platform for cloud computing.

2013

OpenPOWER foundation formed as an independent organization to foster innovation around Power chips.

2015

TensorFlow released under Apache 2.0 open-source license, helping spur machine learning development.

2016

IBM helps Linux Foundation establish Hyperledger in one of first initiatives to build an open-source enterprise blockchain platform.

IBM helps launch Cloud Native Computing Foundation to support open governance around Kubernetes.

2017

In one of the first publicly collaborative quantum computing projects, IBM open sources the Qiskit developer toolkit.

Cloudera debuts publicly in highly-anticipated IPO of an open source company.

cloudera

2018

IBM announces plans to purchase Red Hat, which will accelerate the impact of open-source development in the cloud.

Since the advent of modern computing, the ability of passionate programmers to collaborate on open source code has led to some of the most important software breakthroughs. Today, open-source software is more important than ever. It powers our web browsers, databases, operating systems and smartphones. Peek behind the curtain of any new innovation, and chances are you'll find a foundation built on decades of open-source contributions.

The 10 top-ranked supercomputers on the planet as of November 2023 and their operating systems

	Name	Country	OS
1	Frontier	USA	Linux
2	Aurora	USA	Linux
3	Fugaku	Japan	Linux
4	Sunway TaihuLight	China	Linux
5	Tianhe-2A	China	Linux
6	Summit	USA	Linux
7	Sierra	USA	Linux
8	Selene	USA	Linux
9	Dammam-7	Saudi Arabia	Linux
10	SuperMUC-NG	Germany	Linux

Most supercomputers run Linux for several key reasons:

- **Customisation** :
  for specialized configurations.
- **Scalability**:
  scalability for parallel processing.
- **Stability and Reliability**:
  Lots of eyes watch upto the last bit of source code.
- **Community and Support**:
  The large volunteer & passionate community.
- **Cost-Effectiveness**:

SuperMUC-NG

MAJOR LINUX DISTRIBUTIONS

The red hat Linux

UBUNTU

debian

The OS in the server
in this institute.

The most influential open source projects, along with their estimated market share in their respective domains.

	Domain	Estimated Market Share
Linux	Operating Systems	~70% of web servers[1]
Kubernetes	Container Orchestration	~85% of containerized apps[2]
TensorFlow	Machine Learning	~65% of ML frameworks[3]
React [1]	Web Development	~40% of front-end frameworks[4]
Apache Hadoop [2]	Big Data Processing	~50% of big data solutions
MySQL	Databases	~30% of database systems
Git	Version Control	~90% of version control systems
WordPress	Content Management Systems	~40% of websites

These projects have significantly shaped their respective fields and continue to be widely adopted and supported by the developer community.

1 : JavaScript library
2 : Cluster computing

The ten economically most important open source projects.

	Project	Leading company	Market Value
1	Linux	Red Hat	$16 billion
2	Git	GitHub	$2 billion
3	MySQL	Oracle	$1.87 billion
4	Node.js	NodeSource	?
5	Docker	Docker	$1 billion
6	Hadoop	Cloudera	$3 billion
7	Elasticsearch	Elastic	$700 million
8	Spark	Databricks	$513 million
9	MongoDB	MongoDB	$1.57 billion
10	Selenium	Sauce Labs	$470 million

Ranking

The market value of a company refers to the total value of all its outstanding shares.

Wikipedia : open source.

How can open-source organizations earn money ?

Subscriptions and services: They offer enterprise versions of their software with additional features, support, and services, which businesses pay for.

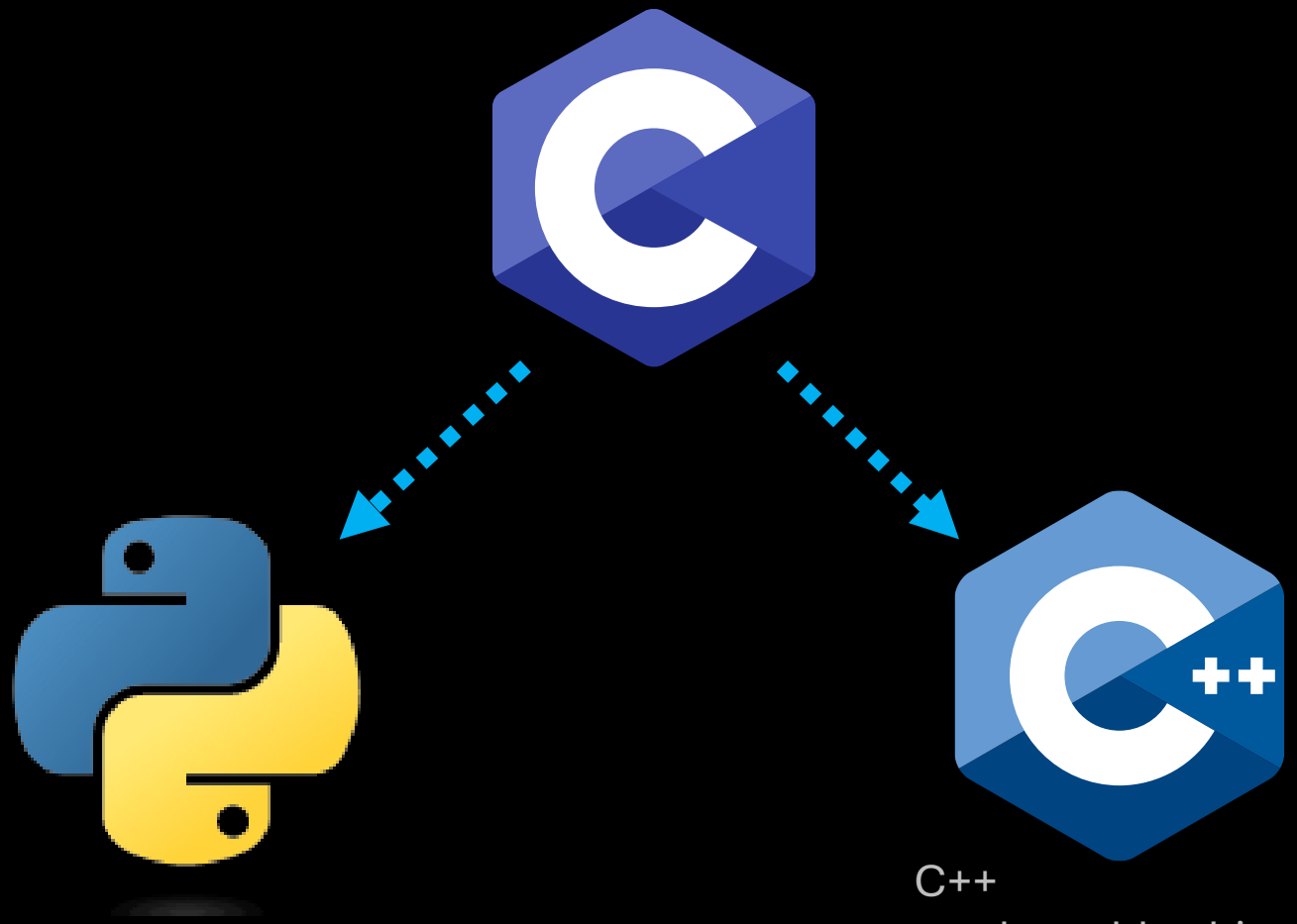
But the intention of creating open-source software is not making money !

More info on open-soure software :

https://www.ibm.com/topics/open-source

Python

Origin of Python



Python
- Interpretor for C
- Adds objects.
- Garbage collection (automatic memory management)

C++
- also adds objects to C
- Empowers low-level access to hardware.
- Manual memory management.

	C	C++	Python
Development History	Dennis Ritchie & Ken Thompson between 1969 and 1973 at AT&T Bell Labs.	Bjarne Stroustrup in 1979 at AT&T Bell Labs.	Guido van Rossum , released in 1991, Amsterdam.
Common Uses	System programming (operating systems), embedded systems	System programming, game development, high-performance computing	Web development, data analysis, scripting, automation , AI
Typing (Variable Declaration)	Statically typed (explicitly declared.)	Statically typed	Dynamically typed
Line Length			3-5 times shorter than equivalent C/C++ programs.
Compiled/Interpreted	Compiled	Compiled	Interpreted
Programming Paradigms	Procedural programming.	Hybrid (supports both procedural and object-oriented programming).	Supports multiple paradigms: procedural, object-oriented, and functional programming.
Inheritance	Does not support inheritance (not objec-oriented).	Supports both single and multiple inheritance.	Supports all 5 types of inheritance (single, multiple, multilevel, hierarchical, hybrid).
Memory Allocation	Provides malloc() and calloc() for dynamic memory allocation, and free() for deallocation.	Provides new operator for memory allocation and delete operator for deallocation.	Automatic memory allocation and deallocation (garbage collection).
Performance	top-notch	top-notch	much slower than C-C++ numerical analysis can be accelerated by just-in-time coompilation ... .
Exception Handling	Direct support for exception handling is not available.	supported.	supported.
Header Files (organization)	Header files for function prototypes and data types.	Header files for function prototypes and data types.	Modules for defining functions, classes, and variables.
Learning Curve	intermedate	most difficult	easiest to learn

OOP

- Object : combines data & functions
- principles
  - Abstraction
  - Encapsulation
  - Inheritance
  - polymorphism

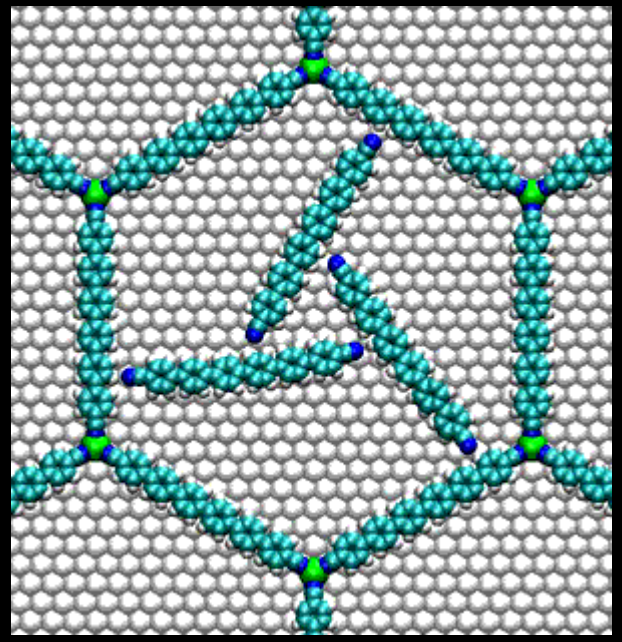
Each chess piece has :

- Attributes : Coordinates
- Methods : Legal moves

- Interaction between pieces

OOP main characteristics

- **Abstraction** : beneficial for the user :
  - the methods & attributes create a user-friendly interface to interact with the program.
- **Encapsulation** : beneficial for the programmer :
  - Organizes & classifies an otherwise entangled web of code.
- **Inheritance** : subclasses inherit from superclasses.
- **Polymorphism** : how subclasses belonging to the same superclass behave differently.



OOP is good for simulations.

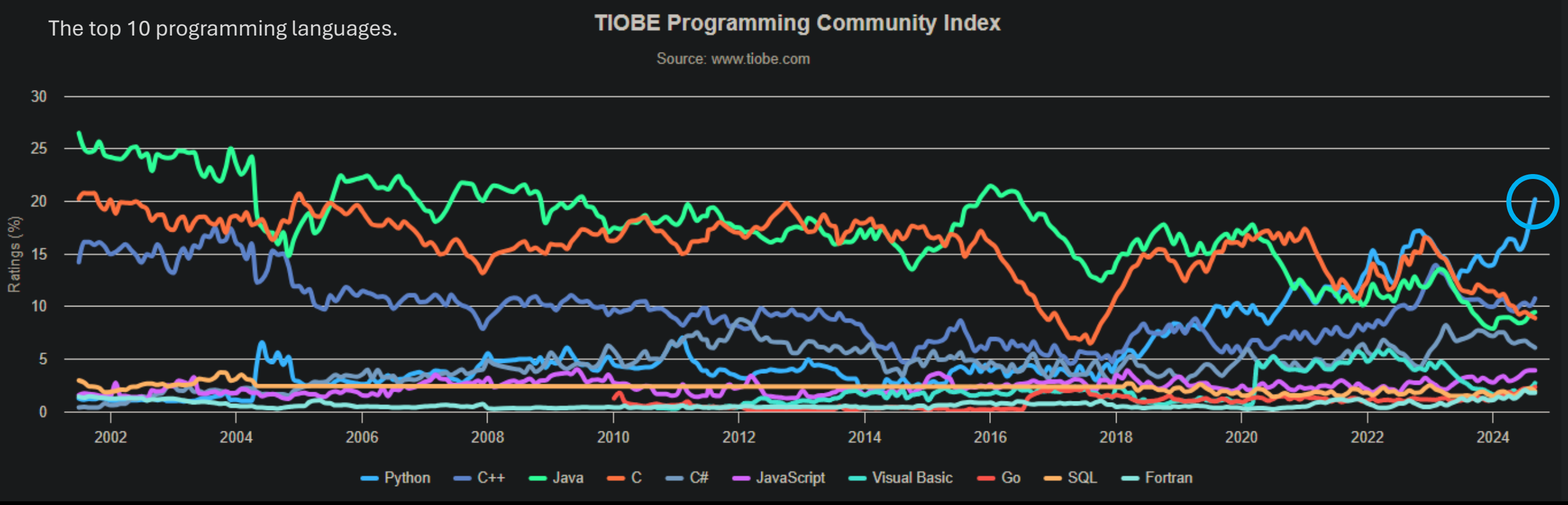
1 nanosecond movie of dynamics simulation of a synthetic molecule in a nanopore (outer diameter 6.7 nm).

One can define each molecule or structure as an object , assign attribute & methods to it , & finally create interaction between them.

Wikipedia : Molecular dynamics

Computer games are also simulations
of real-world scenarios.

The top 10 programming languages.



- The TIOBE Programming Community index is an indicator of the popularity of programming languages.
  - The ratings are based on the number of skilled engineers world-wide, courses and third party vendors.
- TIOBE index is not about the *best* programming language or the language in which *most lines of code* have been written.
- The index can be used to
  - check whether your programming skills are still up to date
  - make a strategic decision about what programming language should be adopted when starting to build a new software system.

Support

- Standard library
- third-party Python software

Numpy

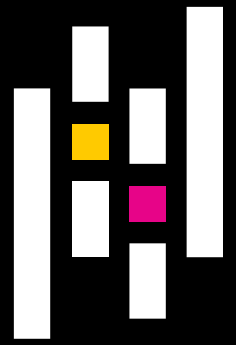
The fundamental package for scientific computing with Python.

	NumPy Arrays	Python Lists
Purpose	Primarily used for scientific computing and numerical operations.	General-purpose data structures for storing collections of items.
Homogeneity	Homogeneous: All elements must be of the same data type.	Heterogeneous: Can hold different data types within the same list.
Memory Allocation	Contiguous (elements are stored next to each other)	Fragmented (randomized allocation of elements)
Memory Efficiency	Efficient memory usage due to contiguous storage of elements.	Additional overhead for each element (type and reference count).
Element-wise Ops	Supports element-wise operations (e.g., addition, multiplication) efficiently (vectorized operations).	Not optimized for numerical computations; slower due to Python's interpretation overhead.
Performance	Faster for numerical operations (vectorized operations)	Slower for numerical operations
Dimensionality	Supports N-dimensional arrays (e.g., 1D, 2D, 3D, etc.).	By default, 1-dimensional, but can create N-dimensional lists (nested list)(still 1D lists storing other 1D lists).
Common Uses	Numerical computations, linear algebra, signal processing	General purpose data storage, storing mixed data types
Advantages	faster and more memory-efficient data generation and manipulation.	• more flexibility and versatility in storing and operating on different types of data. • Supports nested and dynamic data structures.

https://numpy.org/

Pandas

tables containing various data-types.
Feasible for economic or statistical computations.



	pandas	R	SQL
components	DataFrame	data.frame	table
	Series	vector	column
	Index	row.names	primary key
functionality	MultiIndex	nested list	composite key
	GroupBy	dplyr	GROUP BY
	pivot_table	tidyr	PIVOT
	merge	merge	JOIN
	concat	rbind, cbind	UNION, UNION ALL
	Merge() function	?	Relational databases (foreign key)

Group-by : should be accompanied by an aggregation function.

Daa tranformation

- Pivot : long to wide
- Melt : wide to long

https://pandas.pydata.org/

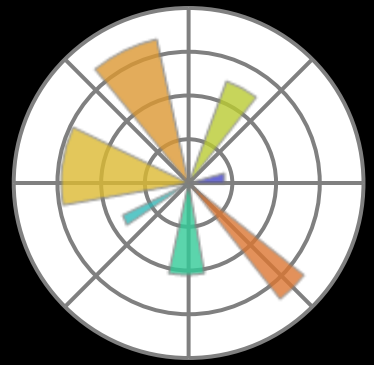
SQL ALCHEMY

SCIPY

Technical and engineering algorithms for

- optimization
- integration
- algebraic equations
- differential equations
- Signal processing
- statistics
- ...

Matplotlib _ Seaborn



Two most important plotting libraries.
Seaborn is based on matplotlib functions.

Other plotting libraries also exist.

numba

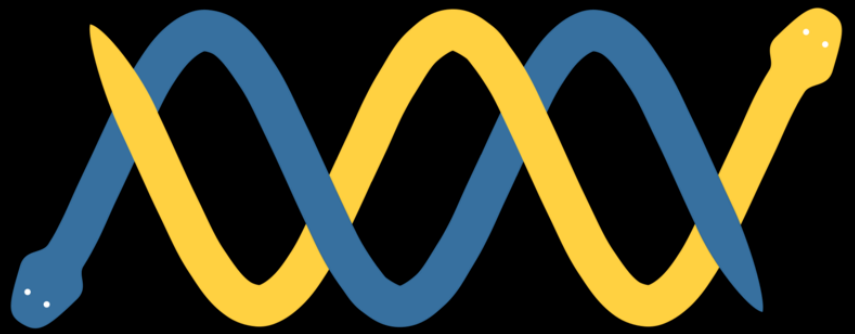
jit

Can reach speeds 1000 times conventional python.

It’s as easy as adding a decorator to your function :

```
@njit(parallel=True)
def function(x1 , x2) :
    ...
```

Biopython



- sequence alignment
- protein structure
- population genetics
- Phylogenetics
- …

Omics data

OOP in AI :
The most important advantage of Python to other languages.

- Machine learning
- Deep learning

Hardware compatibility

Various hardware types **:**

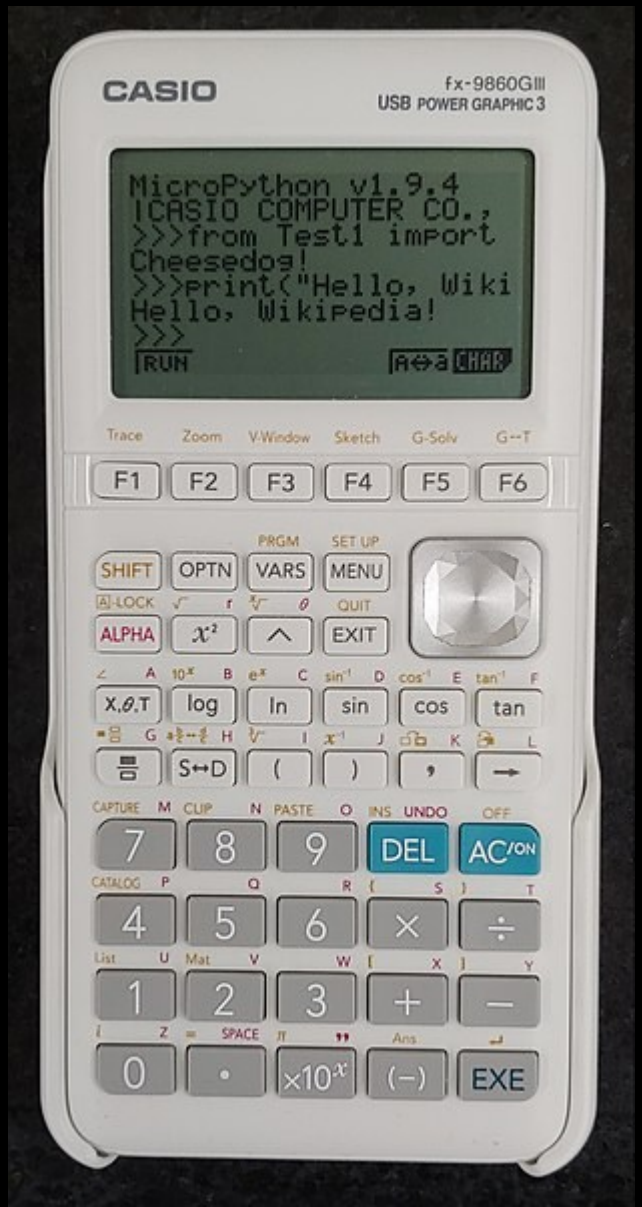
A **single-board computer** (SBC) is a complete computer built on a single circuit board, with microprocessor(s), memory, input/output (I/O) and other features required of a functional computer.

- Mainly made for demonstration, development or educational systems.
- Example : Raspberry Pi

Embedded systems are commonly found in various devices.

most embedded systems use C or C++ to program !

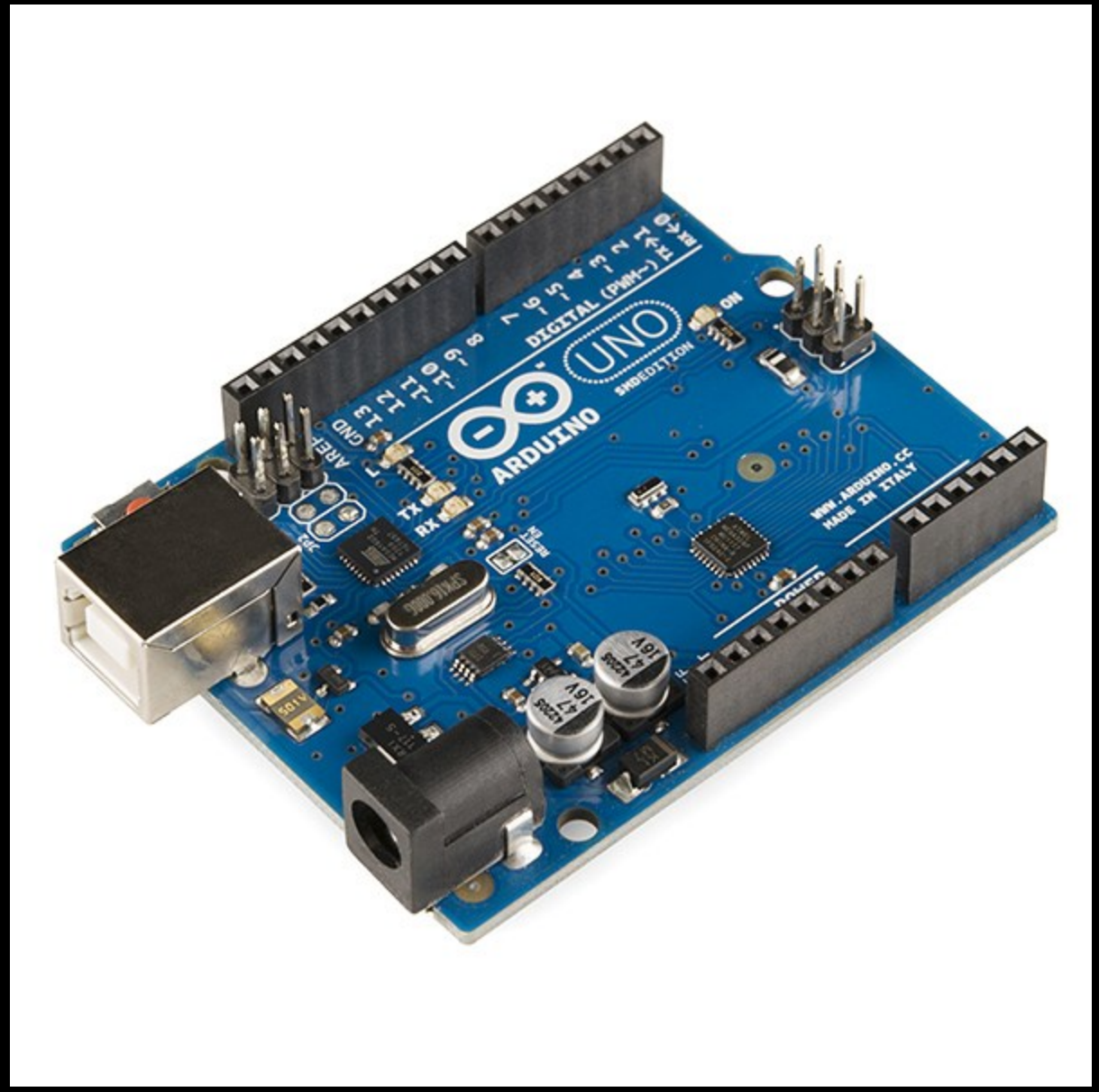
Micropython



- May not be considered a package !
- MicroPython is a software implementation of a programming language largely compatible with Python, written in C, that is optimized to run on a microcontroller.

https://www.casio.com/us/calculators/education/python/

Wikipedia : embedded systems



Single-board microcontroller

Default : no operating system

Arduino is adding the Python language as an additional option for programming microcontrollers. The platform of choice is **MicroPython**.

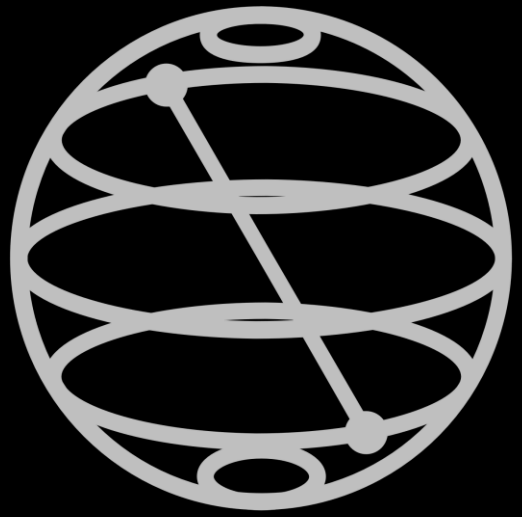
Raspberry Pi

The suffix ‘pi’ is a reference to the python programming language.

Single-board computer
Default OS : LINUX

Raspberry Pi 1, Pi 5, Pi 400, Zero 2 and Pico SBCs (wikipedia)

NASA's *Open Source Rover* powered by a Raspberry Pi 3 (wikipedia)



Qiskit

Qiskit is an open-source software development kit (SDK) for working with quantum computers at the level of circuits, pulses, and algorithms.

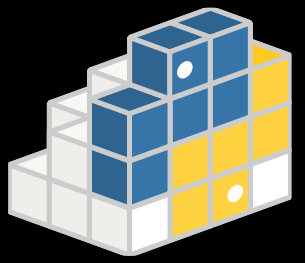
The primary version of Qiskit uses the Python programming language.

wikipedia

Repositories & Package managers.

Create environments to avoid dependency conflicts.

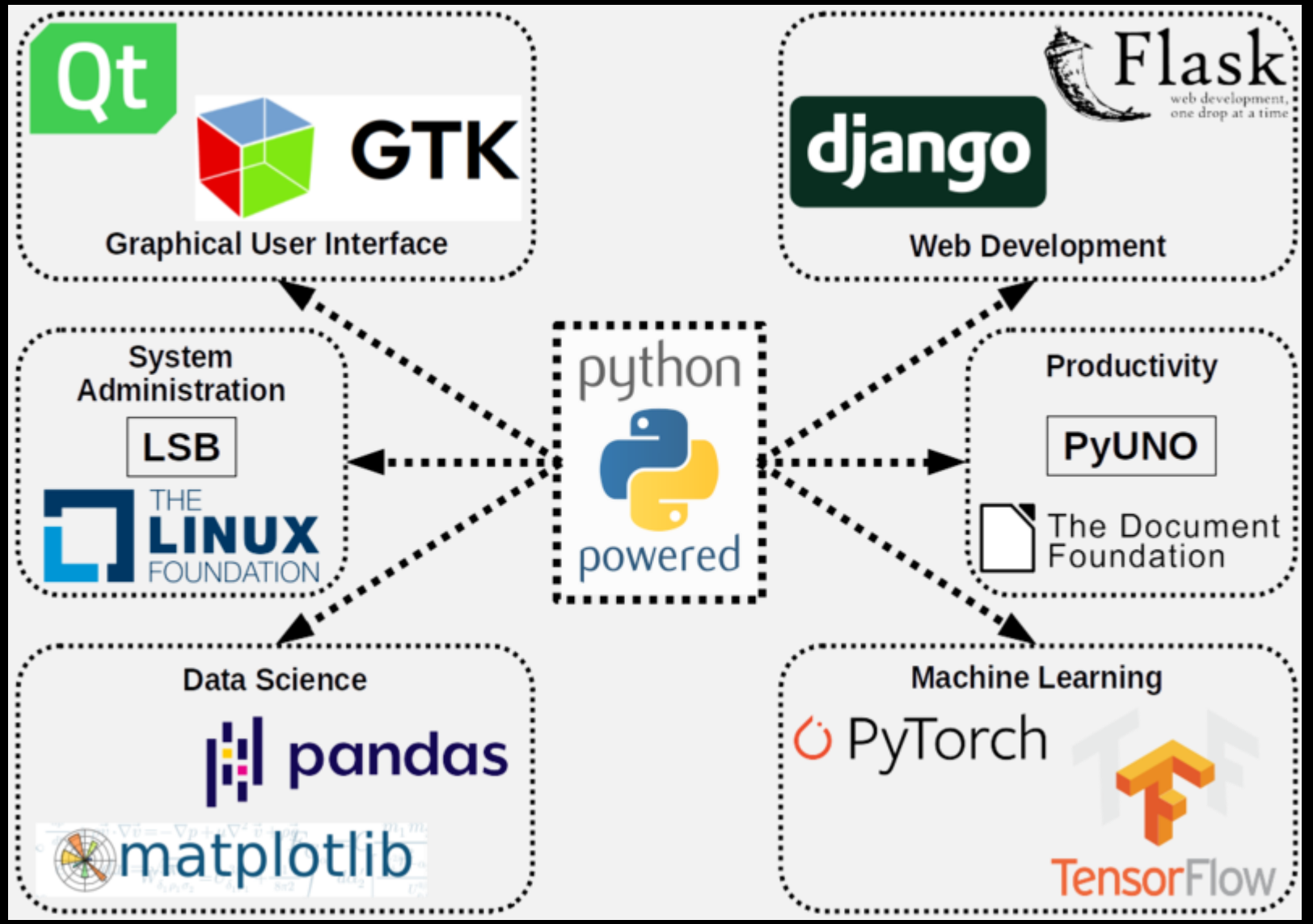
ANACONDA



pypi

As of 14 November 2022, the Python Package Index (PyPI), the official repository for third-party Python software, contains over 415,000[121] packages with a wide range of functionality, including:

- Automation
- Data analytics
- Databases
- Documentation
- Graphical user interfaces
- Image processing
- Machine learning
- Mobile apps
- Multimedia
- Computer networking
- Scientific computing
- System administration
- Test frameworks
- Web frameworks



Wikipedia : python

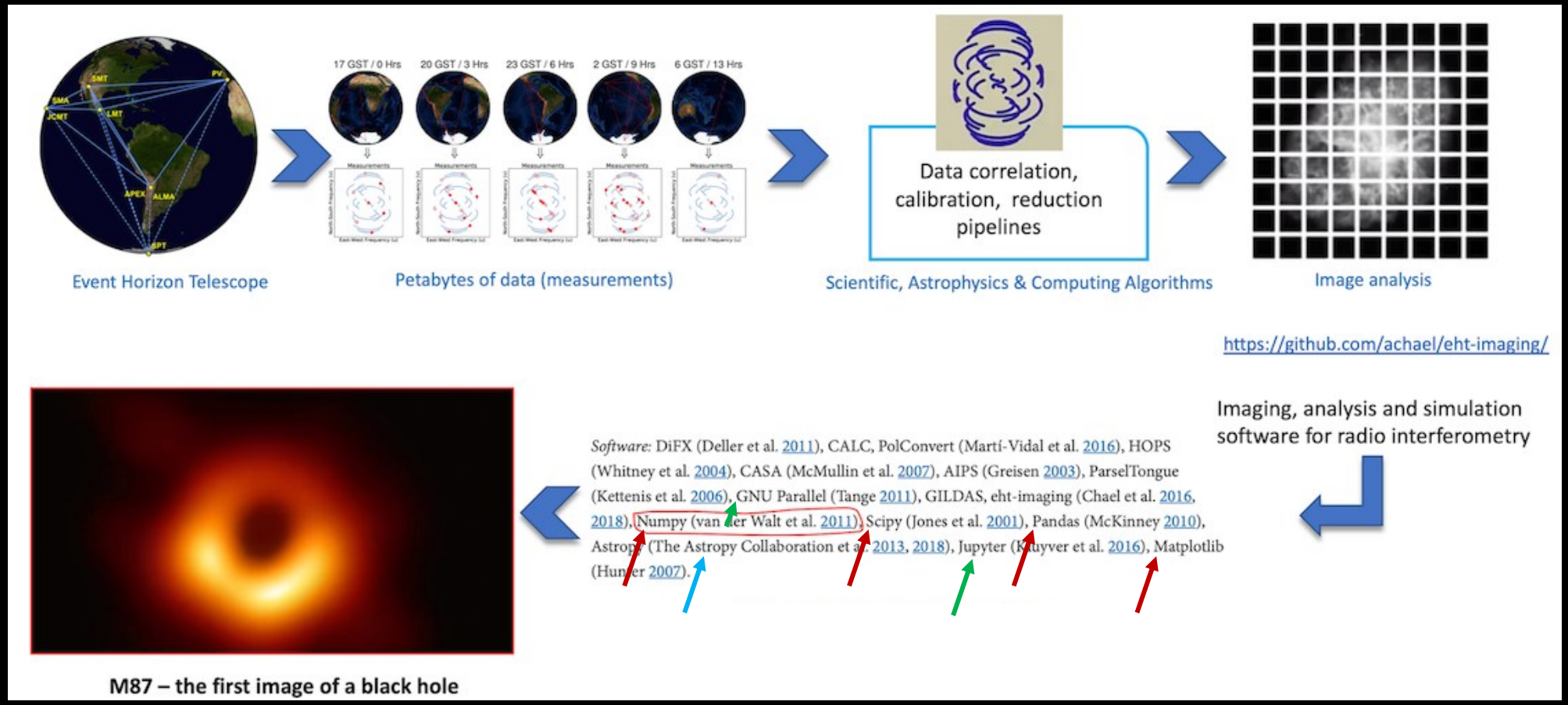
- **FULLY LOADED**
  - Without packages, Python would not be useful.

@ceesjan_van_der_ende (instagram)

Use Case

Event Horizon Telescope

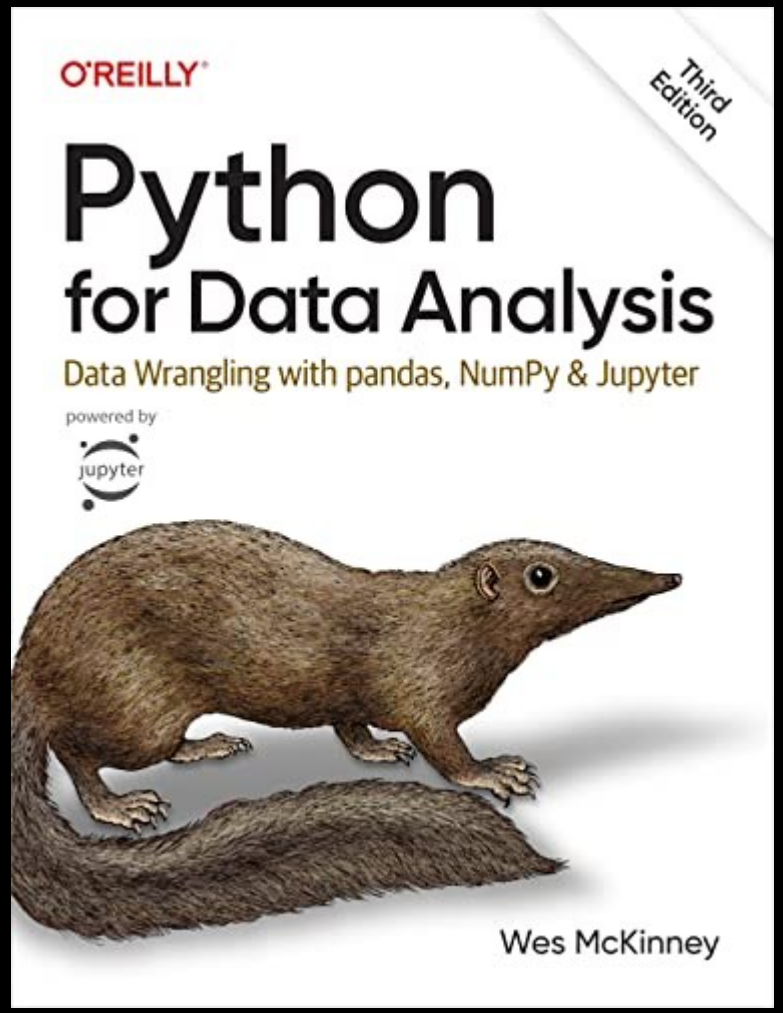

GLT
SMT
Kitt Peak
JCMT
SMA
LMT
NOEMA
30-M
ALMA
APEX
SPT



use case

https://numpy.org/case-studies/blackhole-image/

Resources



@coreyms
Corey Schafer

YouTube

https://scipy-lectures.org/index.html

Gemini

stackoverflow


spyder
jupyterhub

Google Colab

More Topics

- SQL (databases)
- Parallel computing (exploiting GPUs)
- Git _ Github
- Linux OS

- C++
- Rust ?
- Julia

- Libre-office
- ...

- Organizers ?
- Future presenters ?