

Python and Machine learning: An Introduction

ISSAA & RCAA 21

Day #1

Vishal Upendran
IUCAA

Intended Takeaways

1. Open data files and visualize them.
2. Perform vectorized operations on this data.
3. ML-Astro: Have a basic understanding of techniques used.
 - a. Be able to critique data preprocessing, choice of algorithm, etc. in research papers.
 - b. Able to phrase the correct problem to use ML.
 - c. Have a basic understanding of the effectiveness of different techniques when applied to a given task.

Menu for today

1. What *is* python?
2. Where do I code?
3. Some data types and their usage.
4. What are packages?
5. Simple examples?
6. What does astro data look like?
7. Exercise!

What is python?

Python is an **interpreted** high-level general-purpose programming language. Python's design philosophy emphasizes **code readability** with its notable use of **significant indentation**.

– *Wikipedia*

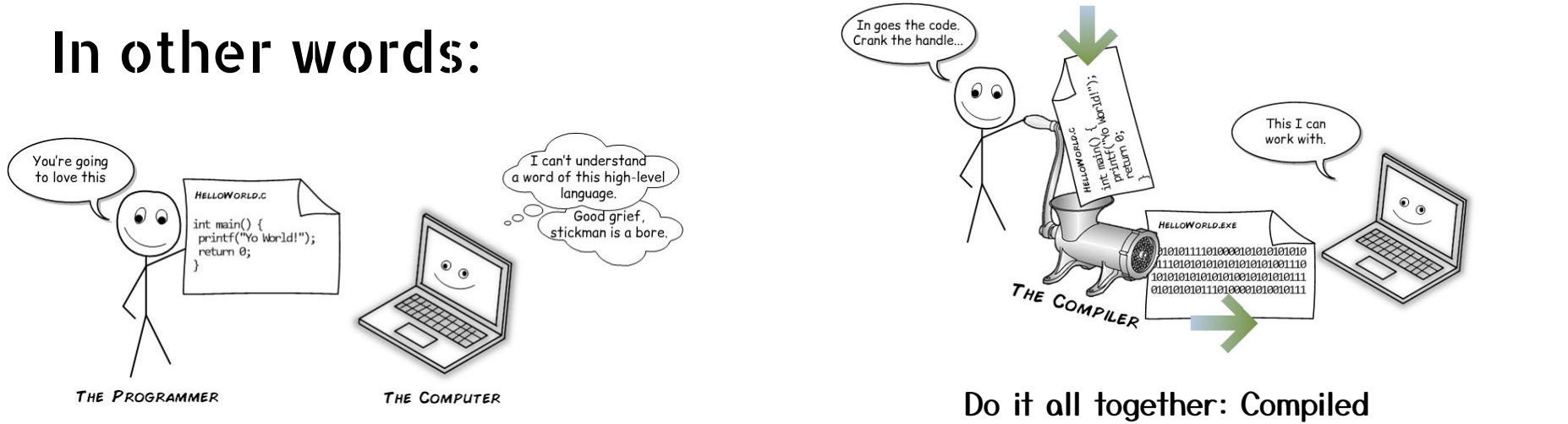
- Interpreted language.
- Code readability.
- Indentation.

Why python?

- Open source.
- Support community.

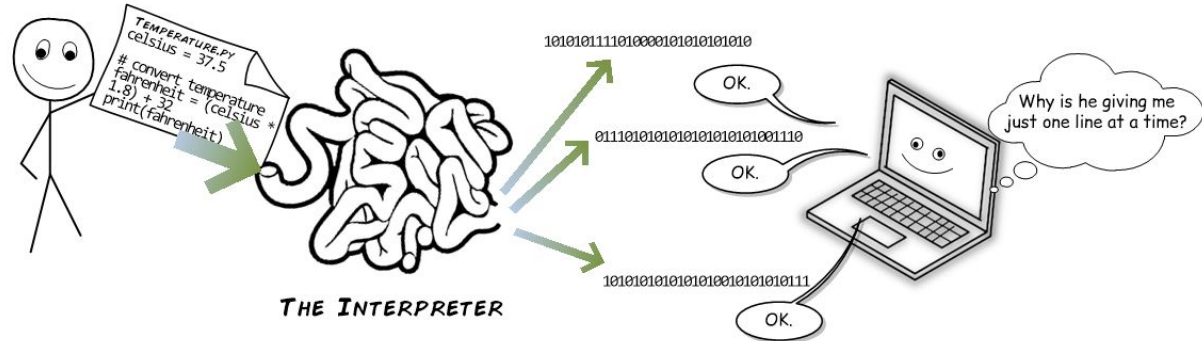
Source: Young.Coder The Difference Between Compiled and Interpreted Languages

In other words:



Your task: Show “Yo World” on monitor

Do it line by line: Interpreted



Isn't that bad?

- Line by line: easily understandable code.
- Execute blocks of code together.
 - Easy to debug.
 - Easy to prototype.
 - Easy to teach!
- Interpreting on the fly: compatibility issues reduced.

An example?



```
# include <iostream>
void main()
{
    cout<<"Hello world!";
}
```



```
print("Hello world!")
```

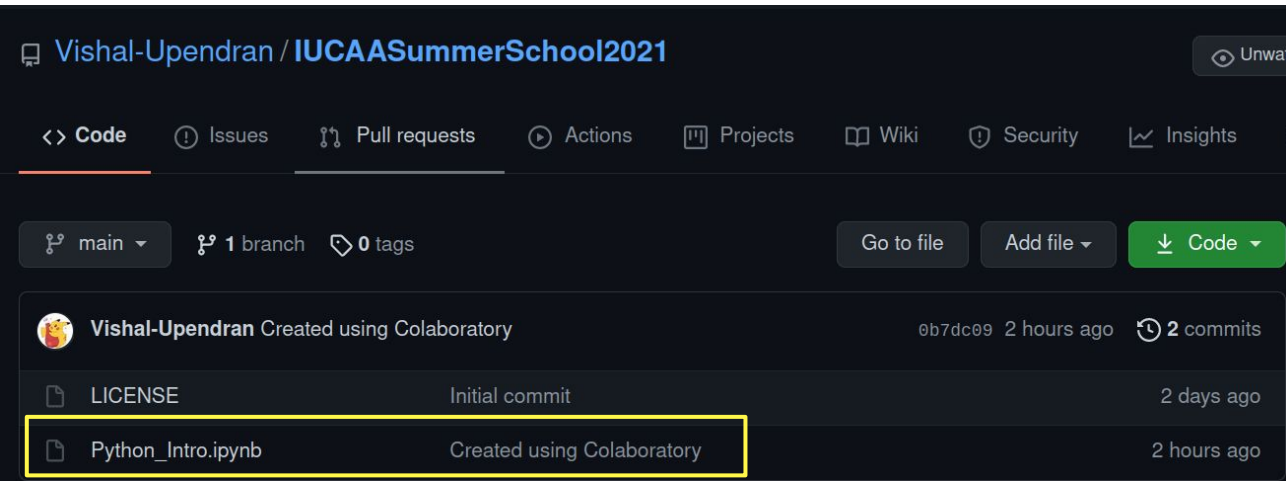
Where to write code?

Jupyter Notebook and Google Colab



This section has:

1. Some basic intro to python, and OOP.
2. Data types - what is nan?
3. Lists, dictionaries, dataframes.



Let us move on to Jupyter →

Some references and important packages

1. [Jupyter](#): Today's codes were written in Jupyter notebook.
2. [Numpy](#): For numerical computation, vector operations; bread & butter.
3. [Matplotlib](#): Plotting and visualization library.
4. [Scipy](#): Scientific-python - curve fitting, special functions, statistics, etc.
5. [Pandas](#): Data loading and manipulation package. If you use GPU, check out [Dask](#).
6. [Skimage](#): Image processing in python.
7. [Astropy](#): Core astronomy packages, coordinate systems, FITS generation. Very important for astronomers.
8. [Sunpy](#): Associated packages for image astro manipulation, plotting, analysis

Always checkout the documentation of the packages if needed.
Stackoverflow is your best friend to venture into python.