

# Data Science For All

## Lecture 03

July 12, 2023, UC Irvine

Prof. Chen Li

# Agenda today

- Python overview (using a Python UDF in Texera)
  - Example workflow: <https://kiwi1.ics.uci.edu/workflow/348>

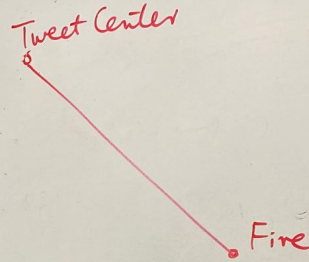
1) Library: reusability / abstraction

2) Variables, constants, assignment

3) Class: <sup>(OO)</sup> object-oriented programming - Class / Object

4) Function

5) Key-value pairs



tuple-:

key	"lat_center":	123,	value
	"long_center":	789,	
	"speed-id":	743952,	
	:		
	:		
	"distance":	(?)	

Python Script

```
4 import haversine as hs # An external package
5 fire_center_coord = (34.235, -118.70128) # Coordinate for a specific
6
7
8
9 class ProcessTupleOperator(URFOperatorV2):
10     @overrides
11     def process_tuple(self, tuple_: Tuple, port: int) -> Iterator[Opti
12         the library takes in a coordinate -> Comments
13         tweet_center_coord = (tuple_["lat_center"], tuple_["long_center"]
14         tweet_center_coord["distance"] = hs.haversine(
15             fire_center_coord, tweet_center_coord, unit=hs.Unit.MILES)
16         yield tuple_
17
```

Handwritten annotations on the screen:

- variable (pointing to `fire_center_coord`)
- Assignment (pointing to `=`)
- constants (pointing to `34.235, -118.70128`)
- Interface (pointing to `URFOperatorV2`)
- Input (pointing to `tuple_`)
- Parameter/Argument (pointing to `tuple_`)
- Output (pointing to `-> Iterator`)
- Function (pointing to `def process_tuple`)
- Comments (pointing to `the library takes in a coordinate`)
- value (pointing to `tuple_["lat_center"]`)

Handwritten notes on the left side of the screen:

- Fire
- value
- 123,
- 789,
- 743952.

Handwritten notes at the bottom left:

- ?