Biological algorithms in Python

Kevin Rue-Albrecht
Oxford Biomedical Data Science Training Programme
2020-05-05 (updated: 2020-05-04)

Algorithms

UNIVERSITY OF OXFORD

What is an algorithm?

- A set of rules used to solve a particular problem
- Has a specific input and output

Algorithms

Problem: how to make crepes?

Algorithms

Problem: how to make crepes?

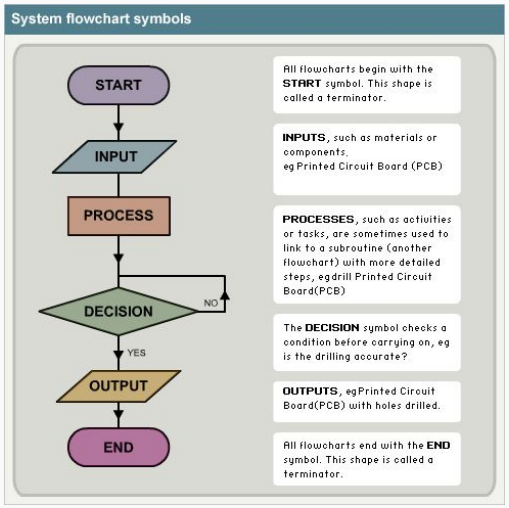
- Inputs:
  - Ingredients
- Algorithm:
  - Preparation
  - Cooking
- Outputs:
  - Crepes

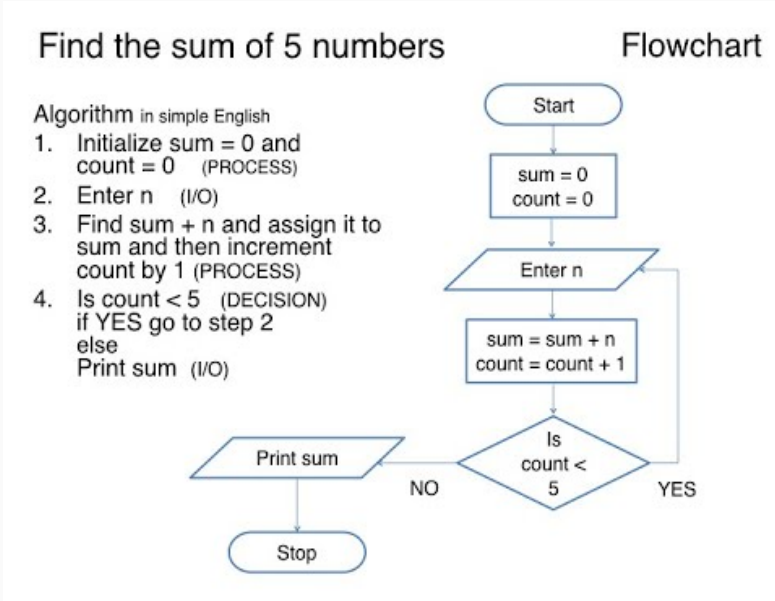
Algorithm Design Tools

Pseudocode

- Steps written in plain English
- Can be written as comments in the script file

Flowcharts





Algorithms

UNIVERSITY OF OXFORD

Computer science

- Finding the maximum number in a list
- Sorting a list of numbers

Biology

- Compute GC content
- Align two sequences (e.g. DNA, protein)

Maximum Number

Find the maximum

- Input: list of numbers.

- Output: Maximum number

Exercise

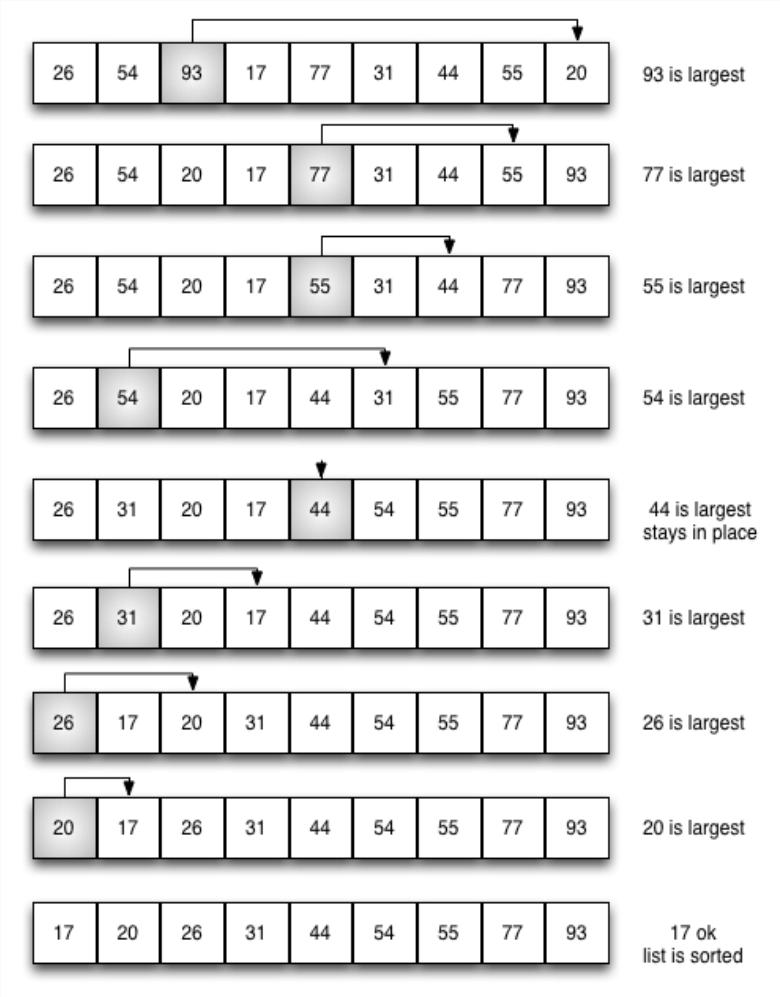
- Let's write an algorithm to do that
- Start with writing pseudocode

Sorting

- Input: list of numbers.

- Output: ?
- Algorithm: ?

Selection sort



Step 1

- Find the maximum across all values
- Swap it with the value at the last position.

Step 2

- Find the maximum across all values but the last one.
- Swap it with the value at the penultimate position.

Step n

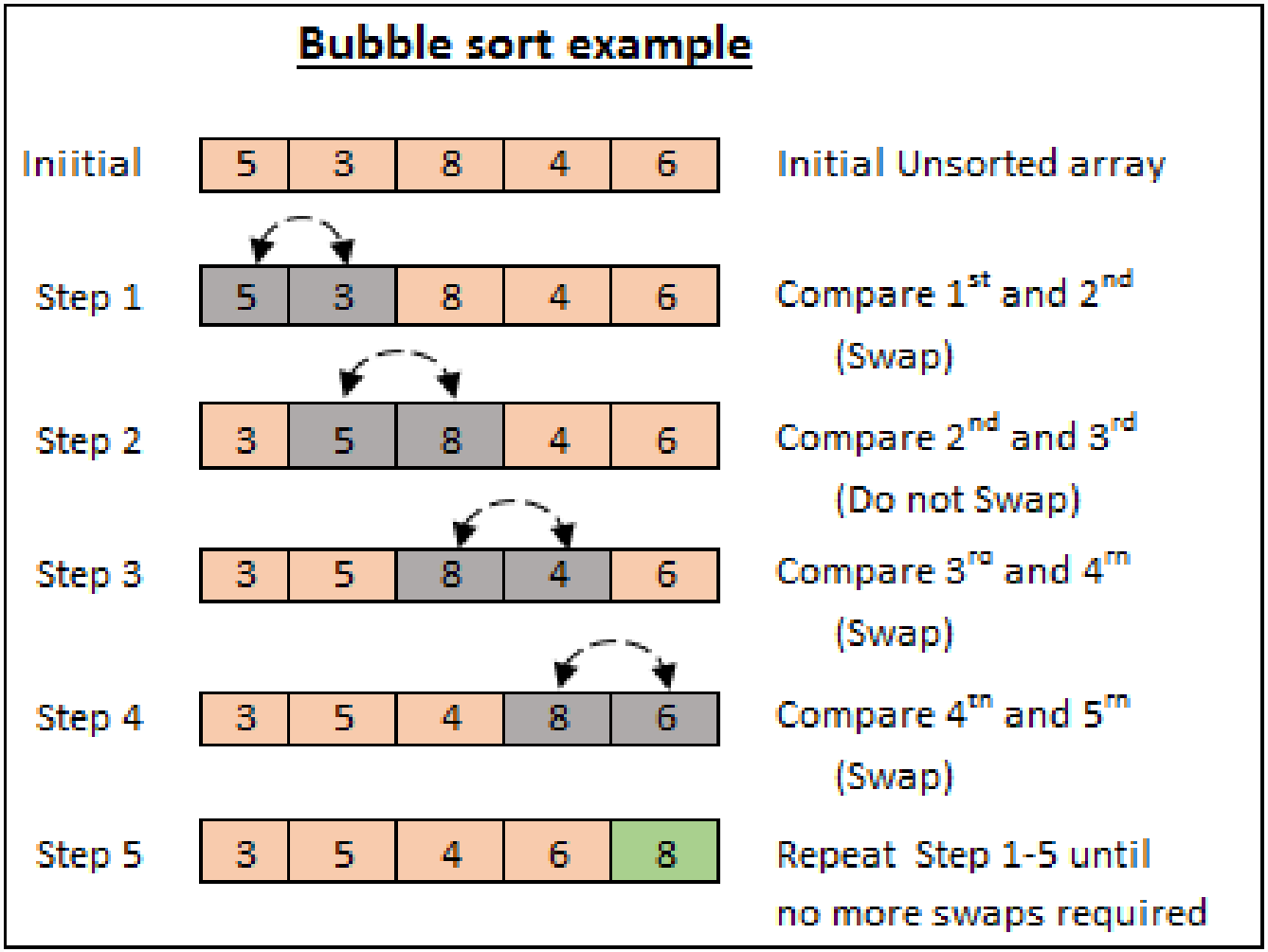
- ...

Exercise

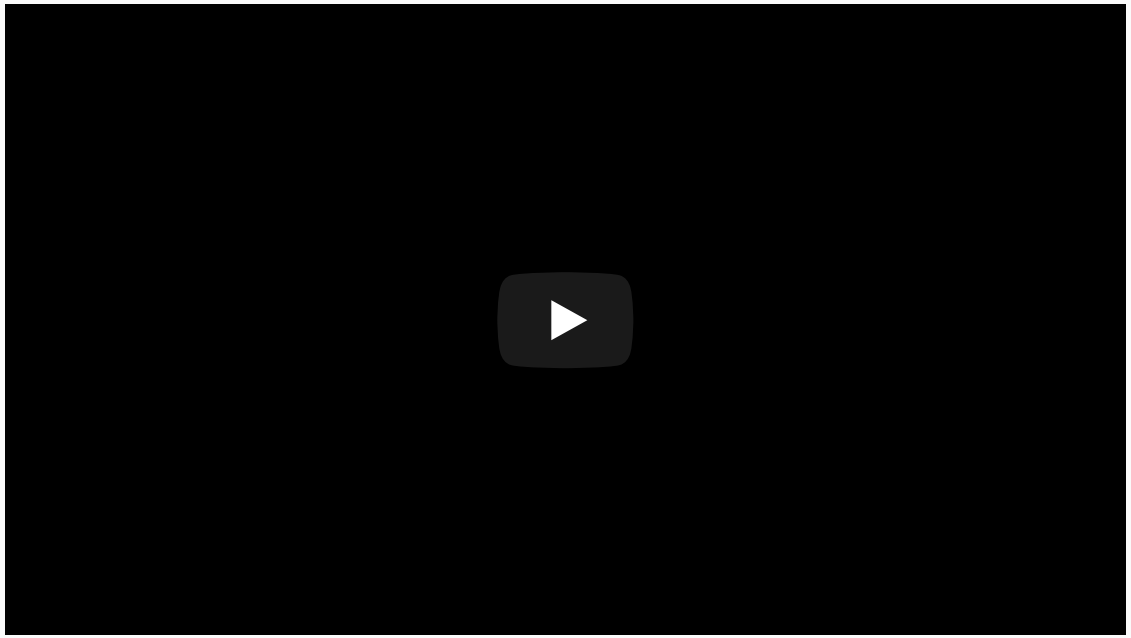
Write the selection sort in Python code.

- Use Spyder

Bubble Sort


Bubble sort example
Iniitial
5 3 8 4 6
Initial Unsorted array
Step 1
5 3 8 4 6
Compare 1st and 2nd
(Swap)
Step 2
3 5 8 4 6
Compare 2nd and 3rd
(Do not Swap)
Step 3
3 5 8 4 6
Compare 3rd and 4th
(Swap)
Step 4
3 5 4 8 6
Compare 4th and 5th
(Swap)
Step 5
3 5 4 6 8
Repeat Step 1-5 until
no more swaps required

Bubble Sort



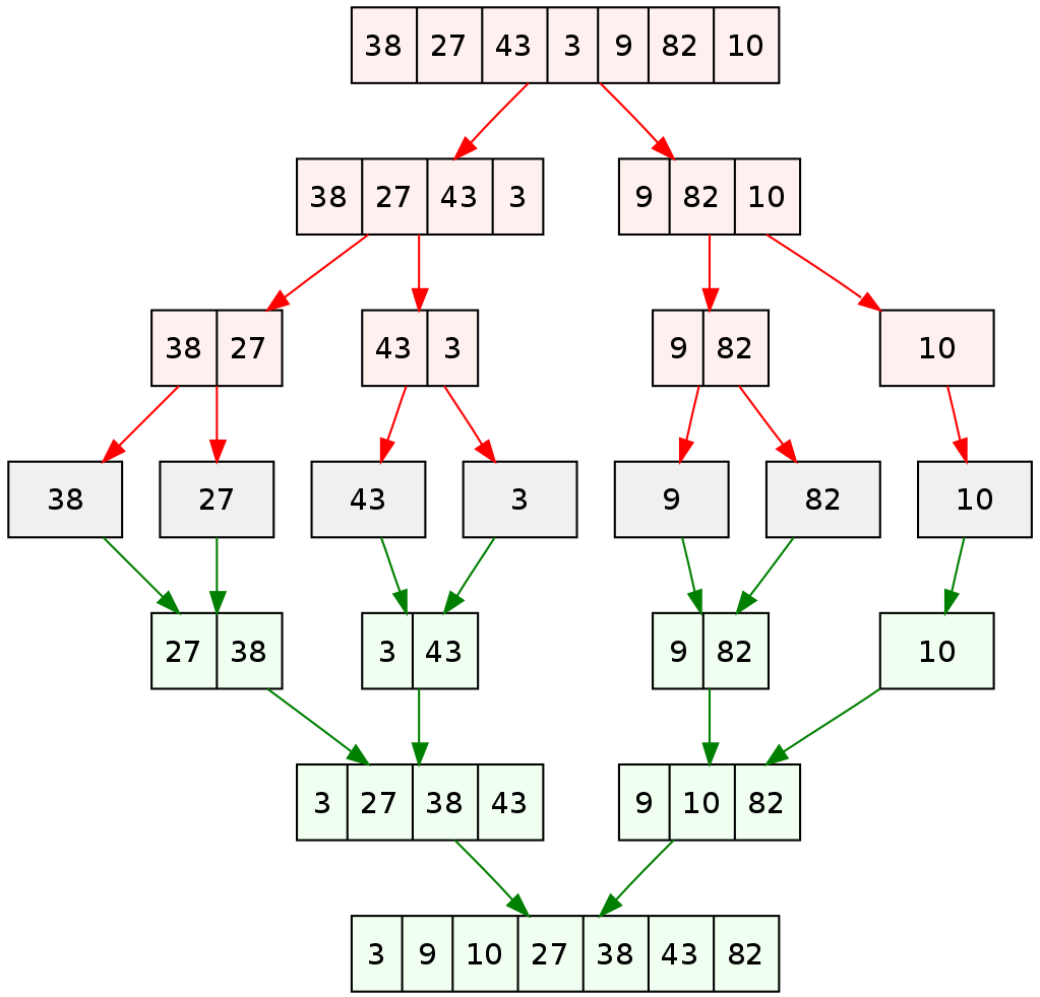
YouTube: https://www.youtube.com/embed/nmhjrI-aW5o

Exercise

Write the bubble sort in Python code.

- Use Spyder

Merge sort


38 27 43 3 9 82 10
38 27 43 3
9 82 10
38 27
43 3
9 82
10
38
27
43
3
9
82
10
27 38
3 43
9 82
10
3 27 38 43
9 10 82
3 9 10 27 38 43 82

Python Standard Library

UNIVERSITY OF OXFORD

- API: Application Programming Interface
  - https://docs.python.org/3
  - https://docs.python.org/3/library/index.html
- Find maximum: `max([6,5,3,1,8])`
- Sorting: `sorted([6,5,3,1,8])`

Exercise

- Write an algorithm to compute GC content
  - https://en.wikipedia.org/wiki/GC-content
  - http://rosalind.info/problems/gc
- Test data:
  - https://www.ensembl.org/Homo_sapiens/Gene/Sequence?db=core;g=ENSG00000172216