Basic programming for drug discovery

Python

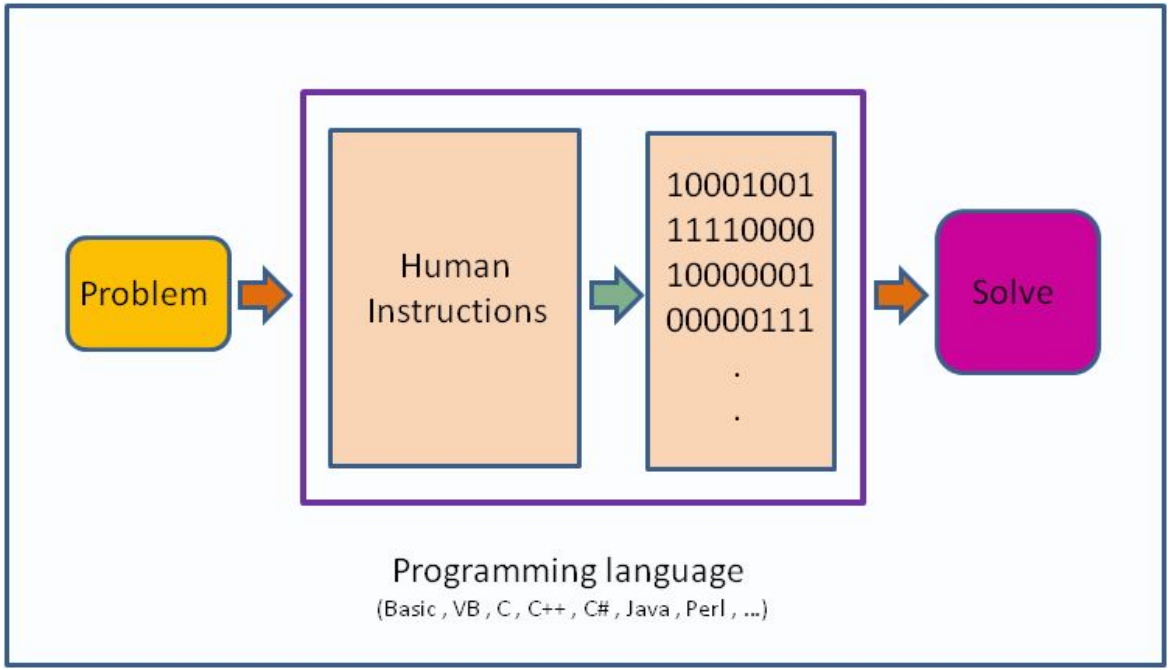
Margriet Palm, Brandon Bongers, and Gerard van Westen

06/11/2018

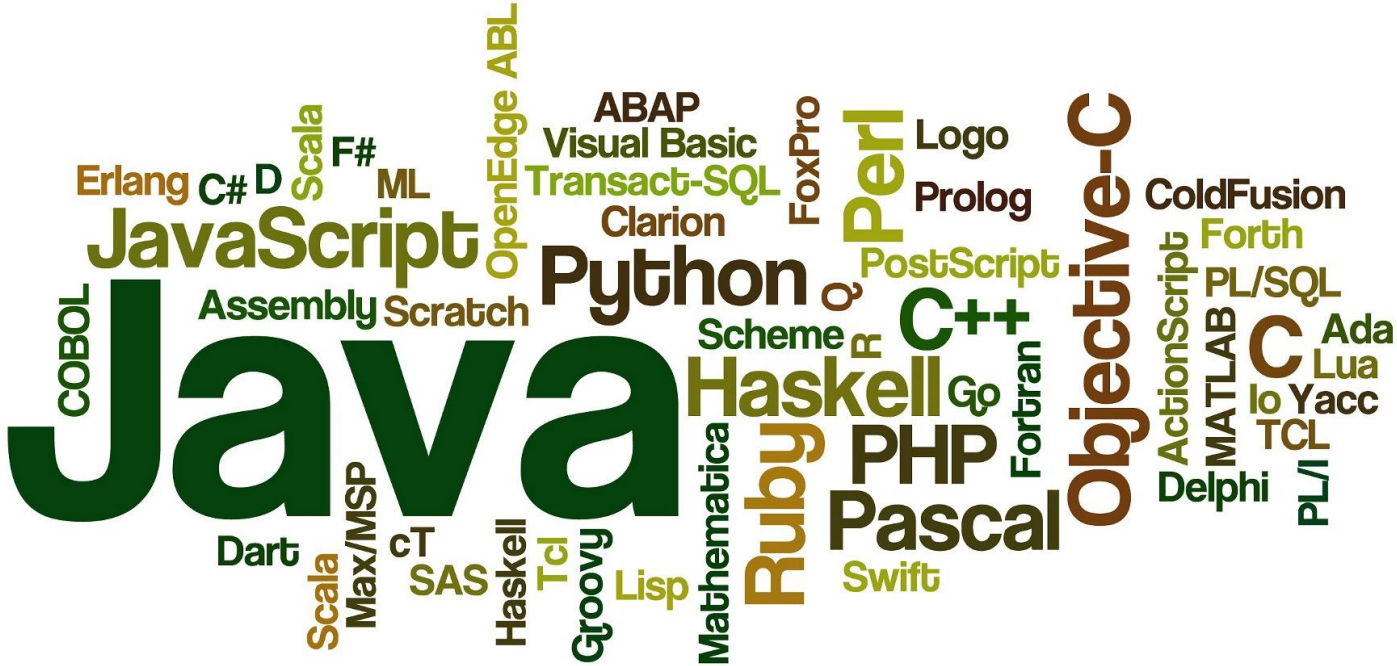
Universiteit
Leiden
The Netherlands

What is programming

Wikipedia: “Computer programming is the process of designing and building an executable computer program for accomplishing a specific computing task.”



Programming languages



A bit of history

...In December 1989, I was looking for a "hobby" programming project that would keep me occupied during the week around Christmas. My office ... would be closed, but I had a home computer, and not much else on my hands. I decided to write an interpreter for the new scripting language I had been thinking about lately: a descendant of ABC that would appeal to Unix/C hackers. I chose Python as a working title for the project, being in a slightly irreverent mood (and a big fan of *Monty Python's Flying Circus*).

— Guido van Rossum

python™

What is python

An interpreted high-level programming language for general-purpose programming.

- Interpreted: not need for *building*, but still quite fast
- General-purpose: used for anything from single line scripts to large commercial projects
- **Syntax focuses on readability**
- Large standard library that can easily be extended with 3rd party packages.
- Large and friendly community of users.

Using Python - IDLE

- IDLE = integrated development environment

```
mpalm@margriet-xps:~$ python
Python 3.6.3 |Anaconda custom (64-bit)| (default, Oct 13 2017, 12:02:49)
[GCC 7.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> print('You can type python code here')
You can type python code here
>>>
```

Using Python - Jupyter notebooks

- Notebook environment for Python and other languages (including R!)
- Mix code, visualization and formatted text
- Runs in the browser and can be accessed remotely

Simple spectral analysis

An illustration of the Discrete Fourier Transform using windowing, to reveal the frequency content of a sound signal.

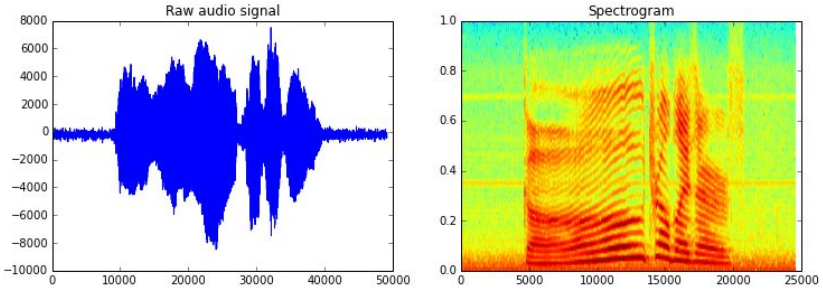
$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \qquad k = 0, \ldots, N-1$$

We begin by loading a datafile using SciPy's audio file support:

In [1]:
```
from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

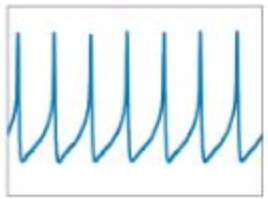
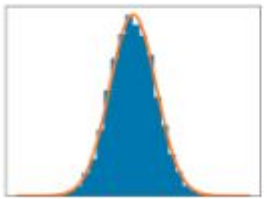
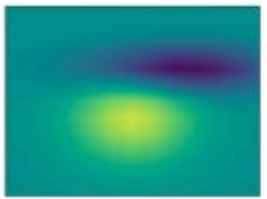
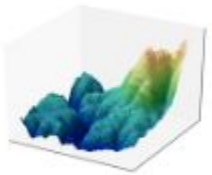
And we can easily view its spectral structure using matplotlib's builtin `specgram` routine:

In [2]:
```
%matplotlib inline
from matplotlib import pyplot as plt
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```



Matplotlib: plotting with python

- Versatile plotting library
- High level of customization for publication/presentation ready plots
- Even better with seaborn for statistical plotting and

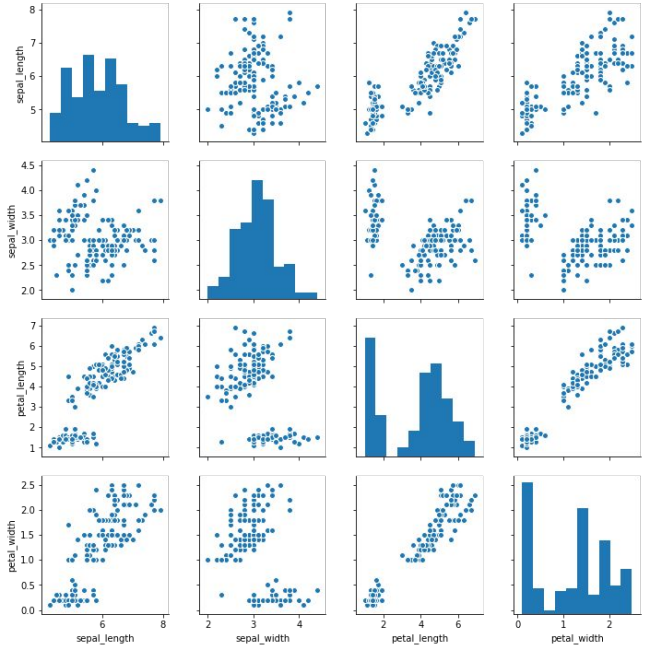
- R's ggplot is also available for Python, but not covered in this course.

Dataframes - Pandas

- Basic Python only has simple data structures
- Dataframes are provided by the pandas package
- Sophisticated plotting with Seaborn library

```
display(iris)
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa



How to actually learn any new programming concept

Essential

Changing Stuff and Seeing What Happens