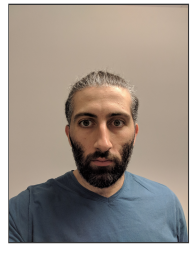
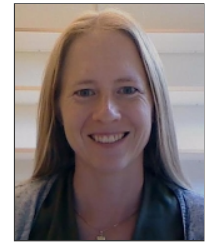
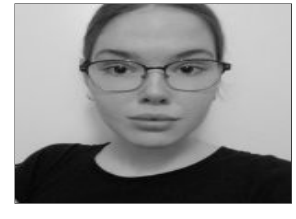
Introduction to

with Application to Bioinformatics

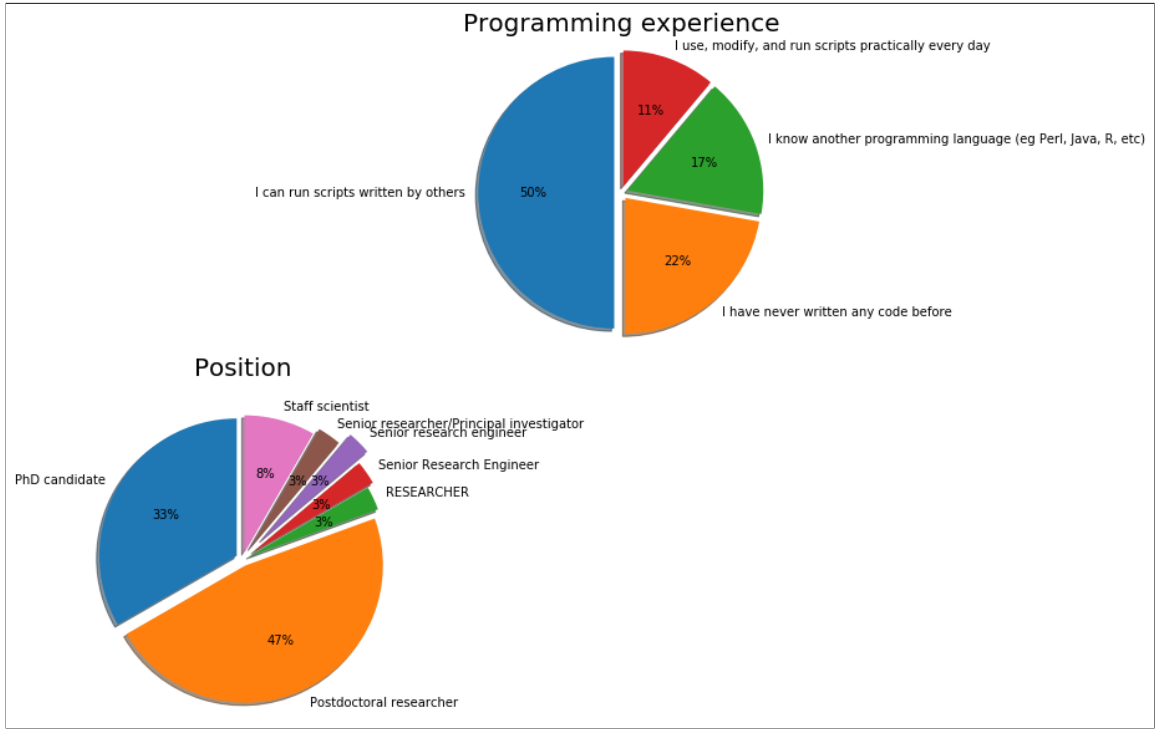
- Day 1

Who we are

Nina	Dimitris	Dan	Jeanette	Ingrid	
					

John	Rui	Kristina	Claudio
			

Who you are


Programming experience
I use, modify, and run scripts practically every day
11%
I know another programming language (eg Perl, Java, R, etc)
17%
I can run scripts written by others
50%
22%
I have never written any code before
Position
Staff scientist
Senior researcher/Principal investigator
Senior research engineer
Senior Research Engineer
RESEARCHER
PhD candidate
8%
3%
3%
3%
3%
33%
47%
Postdoctoral researcher

Practical issues

- Course website: https://nbisweden.github.io/workshop-python/ht20/ (https://nbisweden.github.io/workshop-python/ht20/)
- One main room for lectures
- Same room is used for questions during exercises
- Try to keep your cameras on, but microphone muted
- Breakout rooms are used for discussions in smaller groups, a TA will be assigned to each group
- HackMD used for interaction and questions
- Short lectures with many breaks

Practical issues

- During exercises, TRY TO DISCONNECT FROM ZOOM. You can always connect when you have a question
- Take lots of small breaks also when working with the exercises
- We will try to stick to the schedule, but it's only preliminary until it's happened

If you have any questions during the lecture, feel free to unmute and ask. If you don't want to ask in the Zoom meeting, write the question in the HackMD

To start with

- Write a short presentation of yourself in the HackMD

Schedule


Schedule
Mon
Tue
Wed
Thu
Fri
Time
Intro + Types
ex 1
Operations
ex 2
Discussions
LUNCH
Loops
ex 3
if/else + files
ex 4
Discussions
Project
Review
Review answers
Pseudocode
ex 1
Discussions
LUNCH
Functions/Methods
ex 2
IMDb
Discussions
Project
Review
Review answers
IMDb + sets
Dictionaries
IMDb
Discussions
LUNCH
Functions
ex 1
sys.argv
IMDb
Project
Review
Review exercises
IMDb solution
Keyword args and flow
ex 1
Discussions
LUNCH
Modules and documentation
ex 2
Pandas and plotting
Pandas exercise
Project
Review
Review answers
Recap & Intro to regex
ex 1
Regex in Python
ex 2
Discussions
LUNCH
Sum up
Quiz
Review Quiz
Discussions
Project
Lecture
Exercise
Project
Discussion

Check

- Has everyone managed to install Python?
- Have you managed to run the test script?
- Have you installed notebooks? (optional)

What is programming?

Wikipedia:

"Computer programming is the process of building and designing an executable computer program for accomplishing a specific computing task"

What can we use it for?

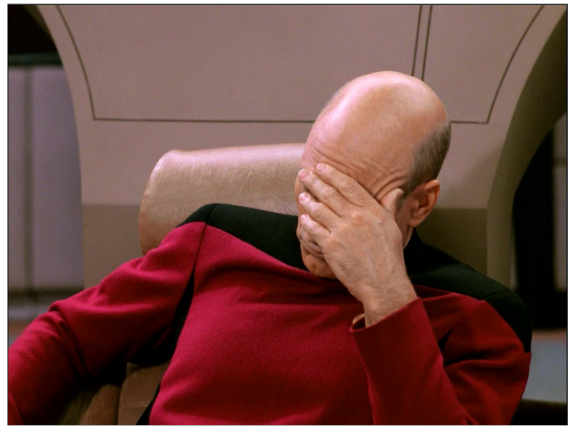
Endless possibilities!

- reverse complement DNA
- custom filtering of VCF files
- plotting of results
- all excel stuff!

Why Python?

Typical workflow

1. Get data
2. Clean, transform data in spreadsheet
3. Copy-paste, copy-paste, copy-paste
4. Run analysis & export results
5. Realise the columns were not sorted correctly
6. Go back to step 2, Repeat



Python versions

Old versions	Python 3
Python 1.0 - January 1994	Python 3.0 - December 3, 2008
Python 1.0 - January 1994	Python 3.1 - June 27, 2009
Python 1.2 - April 10, 1995	Python 3.2 - February 20, 2011
Python 1.3 - October 12, 1995	Python 3.3 - September 29, 2012
Python 1.4 - October 25, 1996	Python 3.4 - March 16, 2014
Python 1.5 - December 31, 1997	Python 3.5 - September 13, 2015
Python 1.6 - September 5, 2000	Python 3.6 - December 23, 2016
Python 2.0 - October 16, 2000	Python 3.7 - June 27, 2018
Python 2.1 - April 17, 2001	Python 3.8 - October 14, 2019
Python 2.2 - December 21, 2001	Python 3.9 - October 5, 2020
Python 2.3 - July 29, 2003	
Python 2.4 - November 30, 2004	
Python 2.5 - September 19, 2006	
Python 2.6 - October 1, 2008	
Python 2.7 - July 3, 2010	

» Course Content

During this course, you will learn about:

- Core concepts about Python syntax: Data types, blocks and indentation, variable scoping, iteration, functions, methods and arguments
- Different ways to control program flow using loops and conditional tests
- Regular expressions and pattern matching
- Writing functions and best-practice ways of making them usable
- Reading from and writing to files
- Code packaging and Python libraries
- How to work with biological data using external libraries (if time allows).

» Learning Outcomes

After this course you should be able to:

- Edit and run Python code
- Write file-processing python programs that produce output to the terminal and/or external files.
- Create stand-alone python programs to process biological data
- Know how to develop your skills in Python after the course (including debugging)

Learning objectives (ie goals for the teachers)

- Increase the student's toolbelt for better quality and performance at work
- Make students understand that there is more to programming than only *knowing* the syntax of a language. This expertise is precisely what NBIS provides.

Some good advice

- 5 days to learn Python is not much
- Amount of information will decrease over days
- Complexity of tasks will increase over days
- Read the error messages!
- Save all your code

How to seek help:

- Google
- Ask your neighbour
- Ask an assistant

Day 1

- Types and variables
- Operations
- Loops
- if/else statements

Example of a simple Python script

In [1]:
```
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is',u)
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

Example of a simple Python script

```
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is '+str(u))
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

Comment

All lines starting with # is interpreted by python as a comment and are not executed. Comments are important for documenting code and considered good practise when doing all types of programming

Example of a simple Python script

```
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is '+str(u))
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

Literals

All literals have a type:

- Strings (str) ‘Hello’ “Hi”
- Integers (int) 5
- Floats (float) 3.14
- Boolean (bool) True or False

Literals define values

In [6]:
```
'this is a string'
"this is also a string"
3       # here we can put a comment so we know that this is an integer
3.14    # this is a float
True    # this is a boolean

type('this is a string')
```

Out[6]:
```
str
```

Collections

In [7]:
```
[3, 5, 7, 4, 99]       # this is a list of integers

('a', 'b', 'c', 'd')   # this is a tuple of strings
{'a', 'b', 'c'}        # this is a set of strings
{'a':3, 'b':5, 'c':7}  # this is a dictionary with strings as keys and integers as values

type([3, 5, 7, 4, 99])
```

Out[7]:
```
list
```

What operations can we do with different values?

That depends on their type:

In [9]:
```
'a string'+' another string'
#2 + 3.4
#'a string ' * 3.2
```

Out[9]:
```
'a string another string'
```

Type	Operations
int	+ - / ** % // ...
float	*+ - / * % // ...*
string	*+*

Example of a simple Python script

```
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is '+str(u))
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

Identifiers

Identifiers are used to identify a program element in the code.

For example:

- Variables
- Functions
- Modules
- Classes

Variables

Used to store values and to assign them a name.

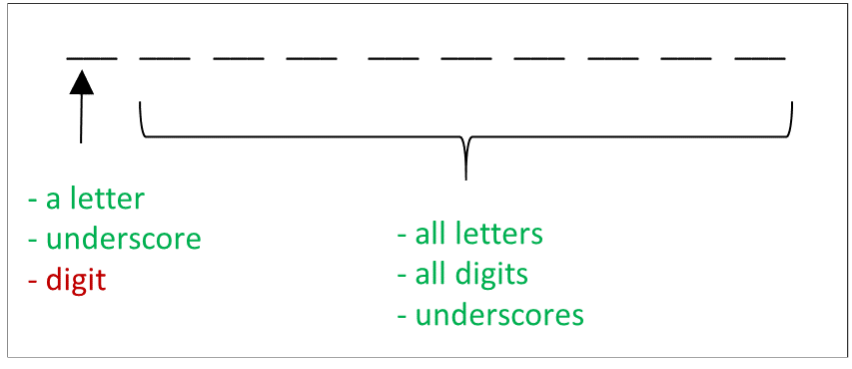
Examples:

- `i       = 0`
- `counter = 5`
- `snpname = 'rs2315487'`
- `snplist = ['rs21354', 'rs214569']`

In [12]:
```
width  = 23564
height = 10

snpname = 'rs56483 '
snplist = ['rs12345','rs458782']

width * height
```

Out[12]:
```
235640
```

How to correctly name a variable



Allowed:

Var_name
_total
aReallyLongName
with_digit_2
dkfsjdsklut *(well, allowed, but NOT recommended)*

Not allowed:

2save
*important
Special%
With spaces

NO special characters:

+ - * $ % ; : , ? ! { } ( ) < > “ ‘ | \ / @

Reserved keywords

False	class	finally	is	return
None	continue	for	lambda	try
True	def	from	nonlocal	while
and	del	global	not	with
as	elif	if	or	yield
assert	else	import	pass	
break	except	in	raise	

These words can not be used as variable names

Summary

- Comment your code!
- Literals define values and can have different types (strings, integers, floats, boolean)
- Values can be collected in lists, tuples, sets, and dictionaries
- The operation that can be performed on a certain value depends on the type
- Variables are identified by a name and are used to store a value or collections of values
- Name your variables using descriptive words without special characters and reserved keywords

→ **Notebook Day_1_Exercise_1 (~30 minutes)**

NOTE!

How to get help?

- Google (https://www.google.com/) and Stack overflow (https://stackoverflow.com/) are your best friends!
- Official python documentation (https://docs.python.org/3/)
- Ask your neighbour
- Ask us

Python standard library

Built-in Functions				
abs( )	delattr( )	hash( )	memoryview( )	set( )
all( )	dict( )	help( )	min( )	setattr( )
any( )	dir( )	hex( )	next( )	slice( )
ascii( )	divmod( )	id( )	object( )	sorted( )
bin( )	enumerate( )	input( )	oct( )	staticmethod( )
bool( )	eval( )	int( )	open( )	str( )
breakpoint( )	exec( )	isinstance( )	ord( )	sum( )
bytearray( )	filter( )	issubclass( )	pow( )	super( )
bytes( )	float( )	iter( )	print( )	tuple( )
callable( )	format( )	len( )	property( )	type( )
chr( )	frozenset( )	list( )	range( )	vars( )
classmethod( )	getattr( )	locals( )	repr( )	zip( )
compile( )	globals( )	map( )	reversed( )	__import__( )
complex( )	hasattr( )	max( )	round( )	

Example `print()` and `str()`

```
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is '+str(u))
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

Note!
Here we format everything to a string before printing it

Python standard library

Built-in Functions				
abs( )	delattr( )	hash( )	memoryview( )	set( )
all( )	dict( )	help( )	min( )	setattr( )
any( )	dir( )	hex( )	next( )	slice( )
ascii( )	divmod( )	id( )	object( )	sorted( )
bin( )	enumerate( )	input( )	oct( )	staticmethod( )
bool( )	eval( )	int( )	open( )	str( )
breakpoint( )	exec( )	isinstance( )	ord( )	sum( )
bytearray( )	filter( )	issubclass( )	pow( )	super( )
bytes( )	float( )	iter( )	print( )	tuple( )
callable( )	format( )	len( )	property( )	type( )
chr( )	frozenset( )	list( )	range( )	vars( )
classmethod( )	getattr( )	locals( )	repr( )	zip( )
compile( )	globals( )	map( )	reversed( )	__import__( )
complex( )	hasattr( )	max( )	round( )	

In [39]:

```
width  = 5
height = 3.6
snps   = ['rs123', 'rs5487']
snp    = 'rs2546'
active = True
nums   = [2,4,6,8,4,5,2]

float(width)
```

Out[39]:

```
5.0
```

More on operations

Operation	Result
x + y	sum of x and y
x - y	difference between x and y
x ** y	x to the power y
....	....
pow(x, y)	x to the power y
float(x)	x converted to float
int(x)	x converted to int!
len(z)	length of z if list
max(z)	maximum in list of z
min(z)	minimum in list of z

In [40]:

```
x = 4
y = 3
z = [2, 3, 6, 3, 9, 23]
pow(x, y)
```

Out[40]: 64

Comparison operators

Operation	Meaning
<	less than
<=	less than or equal
>	greater than
>=	greater than or equal
==	equal
!=	not equal

Can be used on int, float, str, and bool. Outputs a boolean.

In [41]:
```
x = 5
y = 3

#x = 5.14
#y = 3.14

y != x
```

Out[41]:
```
True
```

Logical operators

Operation	Meaning
and	connects two statements, both conditions having to be fulfilled
or	connects two statements, either conditions having to be fulfilled
not	reverses and/or

Membership operators

Operation	Meaning
in	value in object
not in	value not in object

In [42]:

```
x = 2
y = 3

x == 2 and y == 5

x = [2,4,7,3,5,9]
y = ['a','b','c']

2 in x
4 in x and 'd' in y
```

Out[42]:

```
False
```

In [13]:

```
# A simple loop that adds 2 to a number and checks if the number is even
i    = 0
even = [2,4,6,8,10]
while i < 10:
    u = i + 2
    print('u is '+str(u)+'. Is this number even? '+str(u in even))
    i += 1
```

```
u is 2. Is this number even? True
u is 3. Is this number even? False
u is 4. Is this number even? True
u is 5. Is this number even? False
u is 6. Is this number even? True
u is 7. Is this number even? False
u is 8. Is this number even? True
u is 9. Is this number even? False
u is 10. Is this number even? True
u is 11. Is this number even? False
```

In [14]:

```
# A simple loop that adds 2 to a number, check if number is even and below 5
i    = 0
even = [2,4,6,8,10]
while i < 10:
    u = i + 2
    print('u is '+str(u)+'. Is this number even and below 5? '+\
          str(u in even and u < 5))
    i += 1
```

```
u is 2. Is this number even and below 5? True
u is 3. Is this number even and below 5? False
u is 4. Is this number even and below 5? True
u is 5. Is this number even and below 5? False
u is 6. Is this number even and below 5? False
u is 7. Is this number even and below 5? False
u is 8. Is this number even and below 5? False
u is 9. Is this number even and below 5? False
u is 10. Is this number even and below 5? False
u is 11. Is this number even and below 5? False
```

Order of precedence

There is an order of precedence for all operators:

Operators	Descriptions
`**`	exponent
`*`, `/`, `%`	multiplication, division, modulo
`+`, `-`	addition, substraction
`<`, `<=`, `>=`, `>`	comparison operators
`==`, `!=`, in, not in	comparison operators
not	boolean NOT
and	boolean AND
or	boolean OR

Word of caution when using operators

In [43]:
```
x = 5
y = 7
z = 2
(x > 6 and y == 7) or z > 1

x > 6 and (y == 7 or z > 1)

# and binds stronger than or
x > 4 or y == 6 and z > 3
x > 4 or (y == 6 and z > 3)
(x > 4 or y == 6) and z > 3
```

Out[43]: False

In [44]:
```
# BEWARE!
x = 5
y = 8

#xx == 6 or xxx == 6 or x > 2
x > 42 or (y < 8 and someRandomVariable > 1000)
```

Out[44]: False

Python does short-circuit evaluation of operators

More on sequences (For example strings and lists)

Lists (and strings) are an ORDERED collection of elements where every element can be accessed through an index.

Operators	Descriptions
x in s	True if an item in *s* is equal to *x*
s + t	Concatenates *s* and *t*
s * n	Adds *s* to itself *n* times
s[i]	*i*th item of *s*, origin 0
s[i:j]	slice of *s* from *i* to *j*-1
s[i:j:k]	slice of *s* from *i* to *j*-1 with step *k*

In [47]:

```
l = [2,3,4,5,3,7,5,9]
s = 'some longrandomstring'

'o' in s

l[1]
s[0:7]
s[0:8:2]
s[-2]
l[0] = 42
s[0] = 'S'
```

Mutable vs Immutable objects

Mutable objects can be altered after creation, while immutable objects can't.

Immutable objects:

- `int`
- `float`
- `bool`
- `str`
- `tuple`

Mutable objects:

- `list`
- `set`
- `dict`

Operations on mutable sequences

Operation	Result
s[i] = x	item *i* of *s* is replaced by *x*
s[i:j] = t	slice of *s* *from* *i* to *j-1* is replaced by the contents of the iterable t
del s[i:j]	removes element *i* to *j-1*
s[i:j:k] = t	specified element replaced by *t*
s.append(x)	appends *x* to the end of the sequence
s[i:j:k]	slice of *s* from *i* *to* *j-1* with step *k*
s[:] or	creates a copy of *s*
s.copy( )	creates a copy of *s*
s.insert(i, x)	inserts *x* *into* *s* at the index *i*
s.pop([i])	retrieves the item *i* from *s* and also removes it
s.remove(x)	retrieves the first item from *s* where s[i] == x
s.reverse( )	reverses the items of *s* in place

In [48]:
```
s = [0,1,2,3,4,5,6,7,8,9]
s.insert(5,10)
s.reverse()
s
```

Out[48]: [9, 8, 7, 6, 5, 10, 4, 3, 2, 1, 0]

Summary

- The python standard library has many built-in functions regularly used
- Operators are used to carry out computations on different values
- Three types of operators; comparison, logical, and membership
- Order of precedence crucial!
- Mutable object can be changed after creation while immutable objects cannot be changed

→ **Notebook Day_1_Exercise_2 (~30 minutes)**

Loops in Python

In [26]:
```
fruits = ['apple','pear','banana','orange']

print(fruits[0])
print(fruits[1])
print(fruits[2])
print(fruits[3])
```

```
apple
pear
banana
orange
```

In [27]:
```
fruits = ['apple','pear','banana','orange']

for fruit in fruits:
    print(fruit)
#    print('end')
print('end')
```

```
apple
pear
banana
orange
end
```

Always remember to INDENT your loops!

Different types of loops

For loop

In [49]:
```
fruits = ['apple','pear','banana','orange']

for fruit in fruits:
    print(fruit)
print('end')
```

```
apple
pear
banana
orange
end
```

While loop

In [51]:

```
fruits = ['apple','pear','banana','orange']

i = 0
while i < len(fruits):
    print(fruits[i])
    i = i + 1
```

```
apple
pear
banana
orange
```

Different types of loops

`For loop`

Is a control flow statement that performs a fixed operation over a known amount of steps.

`While loop`

Is a control flow statement that allows code to be executed repeatedly based on a given Boolean condition.

Which one to use?

`For` loops better for simple iterations over lists and other iterable objects

`While` loops are more flexible and can iterate an unspecified number of times

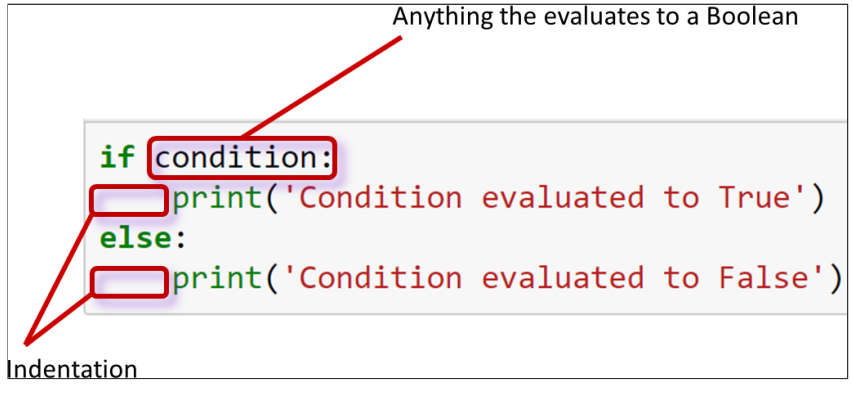
Example of a simple Python script

```python
# A simple loop that adds 2 to a number
i = 0
while i < 10:
    u = i + 2
    print('u is '+str(u))
    i += 1
```

```
u is 2
u is 3
u is 4
u is 5
u is 6
u is 7
u is 8
u is 9
u is 10
u is 11
```

→ **Notebook Day_1_Exercise_3 (~20 minutes)**

Conditional `if/else` statements



In [52]:
```
shopping_list = ['bread', 'egg', 'butter', 'milk']

if len(shopping_list) > 5:
    print('Go shopping!')
else:
    print('Nah! I\'ll do it tomorrow!')
```

```
Nah! I'll do it tomorrow!
```

In [53]:
```
shopping_list = ['bread', 'egg', 'butter', 'milk']
tired         = False

if len(shopping_list) > 5:
    if not tired:
        print('Go shopping!')
    else:
        print('Too tired, I\'ll do it later')
else:
    if not tired:
        print('Better get it over with today anyway')
    else:
        print('Nah! I\'ll do it tomorrow!')
```

```
Better get it over with today anyway
```

This is an example of a nested conditional

Putting everything into a Python script

Any longer pieces of code that have been used and will be re-used SHOULD be saved

Two options:

- Save it as a text file and make it executable
- Save it as a notebook file

Examples

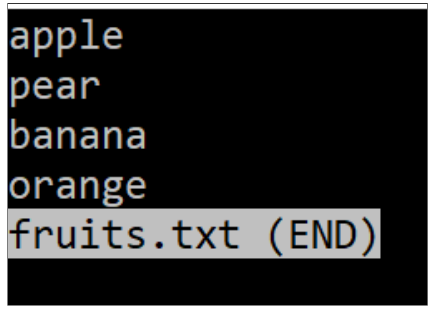
Things to remember when working with scripts

- Put *#!/usr/bin/env python3* in the beginning of the file
- Make the file executable to run with `./script.py`
- Otherwise run script with `python script.py`

Working on files

In [54]:
```
fruits = ['apple','pear','banana','orange']

for fruit in fruits:
    print(fruit)
```

```
apple
pear
banana
orange
```



In [55]:

```
fh = open('../files/fruits.txt', 'r', encoding = 'utf-8')

for line in fh:
    print(line)

fh.close()
```

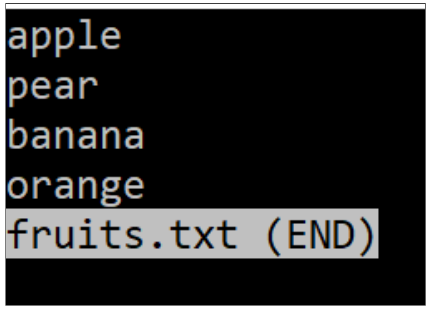
```
apple

pear

banana

orange
```

Aditional useful methods:

`'string'.strip()`	Removes whitespace
`'string'.split()`	Splits on whitespace into list

In [56]:
```
s  = '  an example string to split with whitespace in end   '
sw = s.strip()
sw
#l  = sw.split()
#l
#l  = s.strip().split('\t')
#l
```

Out[56]:
```
'an example string to split with whitespace in end'
```

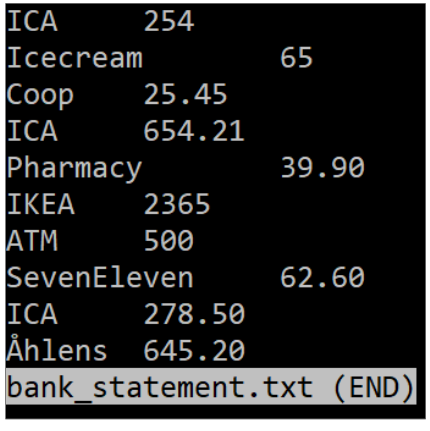


In [36]:

```
fh = open('../files/fruits.txt', 'r', encoding = 'utf-8')

for line in fh:
    print(line.strip())

fh.close()
```

```
apple
pear
banana
orange
```

Another example



How much money is spent on ICA?

In [57]:

```
fh    = open("../files/bank_statement.txt", "r", encoding = "utf-8")

total = 0

for line in fh:
    expenses = line.strip().split()  # split line into list
    store    = expenses[0]           # save what store
    price    = float(expenses[1])    # save the price
    if store == 'ICA':               # only count the price if store is ICA
        total = total + price
fh.close()

print('Total amount spent on ICA is: '+str(total))
```

```
Total amount spent on ICA is: 1186.71
```

Slightly more complex...

```
store   year    month   day     sum
ICA     2018    08      30      254
Icecream        2018    09      05      65
Coop    2018    09      08      25.45
ICA     2018    09      22      654.21
Pharmacy        2018    09      23      39.90
IKEA    2018    09      25      2365
ATM     2018    09      28      500
SevenEleven     2018    09      29      62.60
ICA     2018    09      29      278.50
Åhlens  2018    10      02      645.20
bank_statement_extended.txt (END)
```

How much money is spent on ICA in September?

In [ ]:

```
fh    = open("../files/bank_statement_extended.txt", "r", encoding = "utf-8")

total = 0

for line in fh:
    if not line.startswith('store'):
        expenses = line.strip().split()
        store    = expenses[0]
        year     = expenses[1]
        month    = expenses[2]
        day      = expenses[3]
        price    = float(expenses[4])
        if store == 'ICA' and month == '09':   # store has to be ICA and month september
            total = total + price
fh.close()

out = open("../files/bank_statement_results.txt", "w", encoding = "utf-8")   # open a file for writing the results to
out.write('Total amount spent on ICA in september is: '+str(total))
out.close()
```

Summary

- Python has two types of loops, `For` loops and `While` loops
- Loops can be used on any iterable types and objects
- `If/Else` statement are used when deciding actions depending on a condition that evaluates to a boolean
- Several `If/Else` statements can be nested
- Save code as notebook or text file to be run using python
- The function `open()` can be used to read in text files
- A text file is iterable, meaning it is possible to loop over the lines

→ **Notebook Day_1_Exercise_4**