1_IntroToPythonAndJupyter

October 23, 2019

1 Introduction to python and Jupyter notebook

Andrea Volkamer

This notebook is an adaption from a notebook on Samo Turks's GitHub page.

1.1 Python

Python is widely used general-purpose high-level programming language.

Its design philosophy emphasizes code readability. It is very popular in science.

1.2 Jupyter

The **Jupyter notebook** is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.

- Evolved from IPython notebook
- In addition to python it supports many other programming languages (Julija, R, Haskell, etc..)
- http://jupyter.org/

Can be easily installed using **Anaconda/Conda**

- https://www.continuum.io/downloads
- This notebook uses python 3.6

1.3 The notebook: cell types - markdown and code

This is **Markdown** cell.

[1]:
```
print('This is cell with code')
```

```
This is cell with code
```

1.4 Some simple python lines…

1.4.1 Variables, lists and dictionaries

[1]:

```
# integer
var1 = 1
# string
my_string = "This is a string"
```

[2]:

```
var1
```

[2]:

```
1
```

[3]:

```
var1+var1
```

[3]:

```
2
```

[4]:

```
print(my_string)
```

```
This is a string
```

[5]:

```
# lists
my_list = [1, 2, 3, 'x', 'y']
my_list
```

[5]:

```
[1, 2, 3, 'x', 'y']
```

[6]:

```
# note we always start counting at 0
my_list[0]
```

[6]:

```
1
```

[7]:

```
# range
my_list[1:3]
```

[7]:

```
[2, 3]
```

[8]:

```
# dictionaries (key:value)
salaries = {'Mike':2000, 'Ann':3000}
```

[9]:

```
salaries['Mike']
```

[9]:

```
2000
```

[10]:

```
# add a new entry
salaries['Jake'] = 2500
```

[11]:

```
salaries
```

[11]:

```
{'Mike': 2000, 'Ann': 3000, 'Jake': 2500}
```

1.4.2 Strings

[12]:
```
# newline is indicated by special character '\n'
long_string = 'This is a string \n Second line of the string'
```

[13]:
```
print(long_string)
```

```
This is a string
 Second line of the string
```

[14]:
```
long_string.split(" ")
```

[14]:
```
['This', 'is', 'a', 'string', '\n', 'Second', 'line', 'of', 'the', 'string']
```

[15]:
```
long_string.split("\n")
```

[15]:
```
['This is a string ', ' Second line of the string']
```

[16]:
```
long_string.count('s') # case sensitive!
```

[16]:
```
4
```

[17]:
```
long_string.upper()
```

[17]:
```
'THIS IS A STRING \n SECOND LINE OF THE STRING'
```

1.4.3 Conditionals

[18]:
```
# if -else clause
if long_string.startswith('X'):
    print('It starts with X')
elif long_string.startswith('T'):
    print('It starts with T')
else:
    print('No')
```

```
It starts with T
```

1.4.4 Loops

[19]:
```
for line in long_string.split('\n'):
    print (line)
```

```
This is a string
 Second line of the string
```

[20]:
```
c = 0
while c < 10:
    c += 2
    print (c)
```

```
2
4
6
8
10
```

1.4.5 File operations

[1]:
```python
with open('./data/EGFR-course.csv', 'r') as f:
    content = f.read()
```

[2]:
```python
print(content)
```

```
COc1cc2ncnc(Nc3ccc(F)c(Cl)c3)c2cc1OCCCN1CCOCC1,Gefitinib
C#Cc1cccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1,Erlotinib
CS(=O)(=O)CCNCc1ccc(-c2ccc3ncnc(Nc4ccc(OCc5cccc(F)c5)c(Cl)c4)c3c2)o1,Lapatinib
CN(C)C/C=C/C(=O)Nc1cc2c(Nc3ccc(F)c(Cl)c3)ncnc2cc1O[C@H]1CCOC1,Afatinib
C=CC(=O)Nc1cc(Nc2nccc(-c3cn(C)c4ccccc34)n2)c(OC)cc1N(C)CCN(C)C,Osimertinib
```

1.4.6 Functions

[23]:
```python
def average(numbers):
    return float(sum(numbers)/len(numbers))
```

[24]:
```python
my_numbers = [1,2,2,2.5,3,]
average(my_numbers)
```

[24]:
```
2.1
```

1.4.7 Python libraries

Library is a collection of resources. These include pre-written code, subroutines, classes, etc.

[25]:
```python
from math import exp
```

[26]:
```python
exp(2) #shift tab to access documentation
```

[26]:
```
7.38905609893065
```

[27]:
```python
import math
```

[28]:
```python
math.exp(10)
```

[28]:
```
22026.465794806718
```

1.4.8 Packages we might need during the course

[29]:
```python
import numpy as np # Numpy - package for scientifc computing
```

[30]:
```python
import pandas as pd # Pandas - package for working with data frames (tables)
```

[31]:
```
import sklearn # Scikit-learn - package for machine larning
```

[32]:
```
from rdkit import Chem # RDKit - chemoinformatics library
```

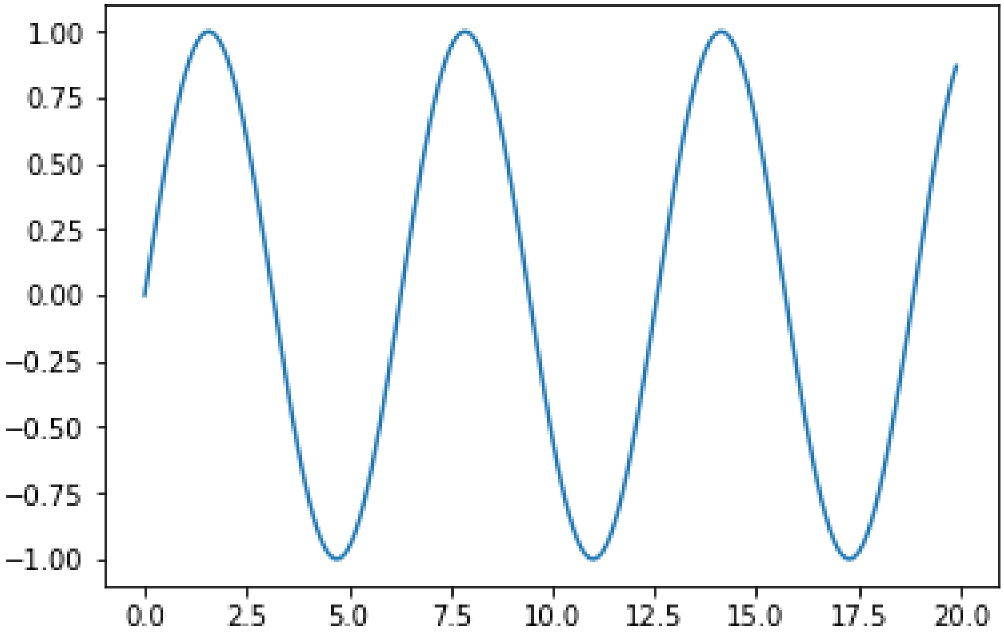
1.4.9 Plotting

[33]:
```
%matplotlib inline
```

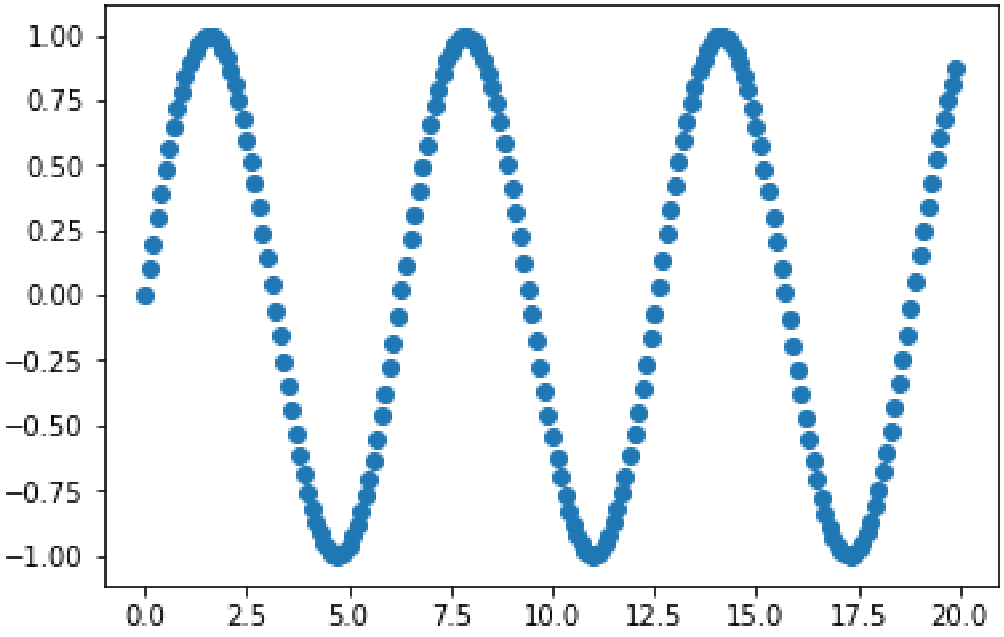
[34]:
```
import matplotlib.pyplot as plt
```

[35]:
```
x_values = np.arange(0, 20, 0.1)
y_values = [math.sin(x) for x in x_values]
```

[36]:
```
plt.plot(x_values, y_values);
```

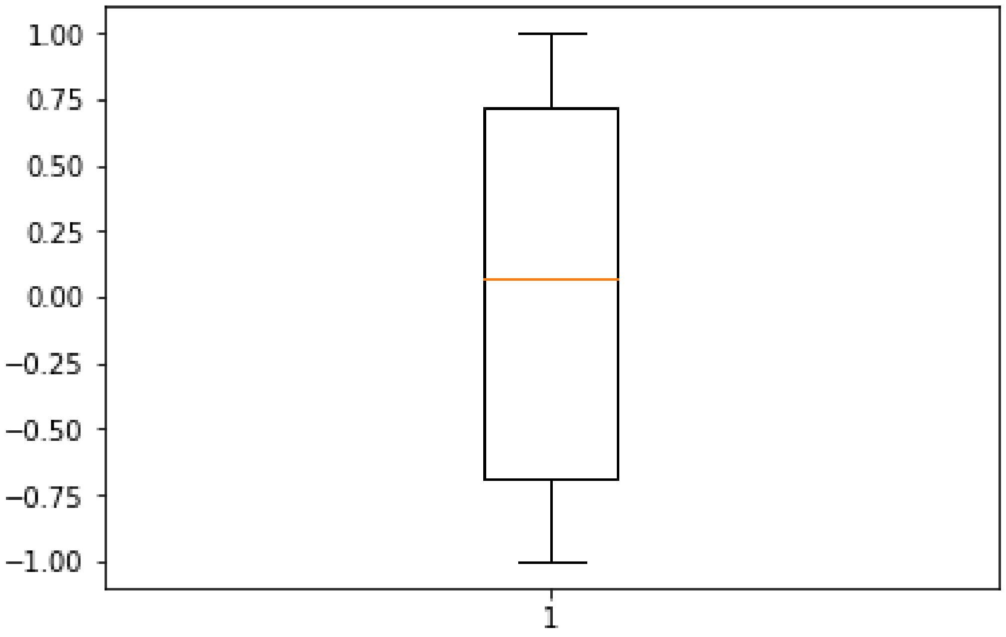


[37]:
```
plt.scatter(x_values, y_values);
```



[38]:
```
plt.boxplot(y_values)
```

[38]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7f91bbb742b0>,
  <matplotlib.lines.Line2D at 0x7f91bbb747f0>],
 'caps': [<matplotlib.lines.Line2D at 0x7f91bbb74c50>,
  <matplotlib.lines.Line2D at 0x7f91bbb830f0>],
 'boxes': [<matplotlib.lines.Line2D at 0x7f91bbb74128>],
 'medians': [<matplotlib.lines.Line2D at 0x7f91bbb83550>],
 'fliers': [<matplotlib.lines.Line2D at 0x7f91bbb839b0>],
 'means': []}
```


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
1