

For the class we need two things Python and a an editor to write longer scripts.

## Editor

If you do not have any editor installed, please donwload Sublime editor from here:

<https://www.sublimetext.com/>

## Python

### **For those who have some version of python installed:**

Make sure you have the following python packages installed:

- Python 2
- IPython and Jupyter notebooks (also called IPython notebook)
- Matplotlib
- Numpy
- Scipy
- Pandas
- Sklearn
- BeautifulSoup
- NetworkX

The easiest way to do this is with your python package manager (such as pip). E.g., at the command line:

```
pip install numpy networkx matplotlib pandas scipy sklearn beautifulsoup4 ipython[notebook]
```

To be sure your system runs all the required packages please move to the 3. TEST YOUR ENVIRONMENT part.

### **For those who do not have python installed yet:**

I recommend to install the Anaconda environment which was created to handle data science projects with python.

#### **What is Anaconda, and why do we need it?**

Anaconda is an easy-to-install free package manager, environment manager, Python distribution, and collection of over 720 open source packages offering free community. Python similarly to R deals with packages which should be installed and then updated to make sure our codes will run later. Installing and managing packages are relatively easy on OS X and Linux, but on Windows is quite tricky. But Anaconda makes it easier.

# DATA ANALYSIS WITH PYTHON - SYSTEM SETUP

Fall term 2016/2017

Course coordinator: Orsolya Vasarhelyi

website: [https://velf.github.io/Data\\_Analysis\\_with\\_Python](https://velf.github.io/Data_Analysis_with_Python)

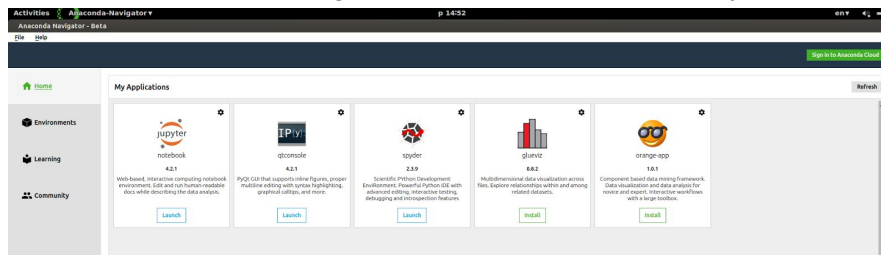
[orsolya.vasarhelyi@gmail.com](mailto:orsolya.vasarhelyi@gmail.com)

## 1. INSTALL PYTHON WITH ANACONDA

Please, go to the official [Anaconda site](https://anaconda.org/) and download Anaconda with **python 2.7**. Simply select the installer appropriate for your platform. Make sure you install the Python 2.7 version. The installer is large, ~200 Mb, so you will need a good connection to download it.

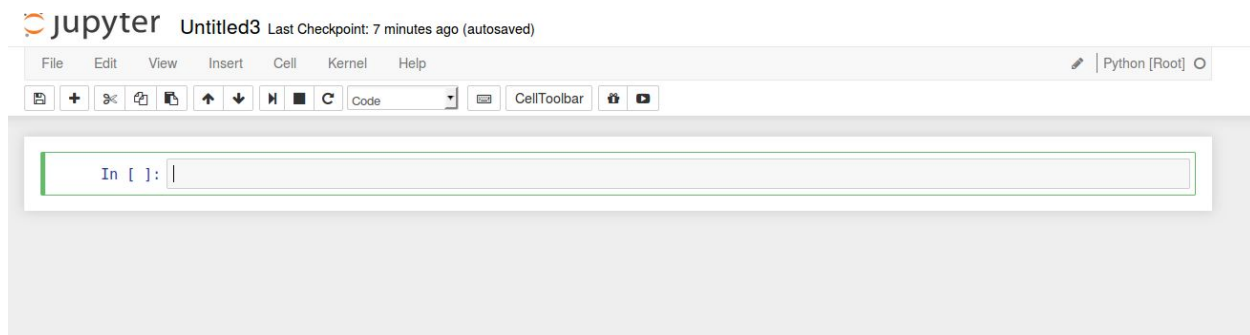
## 2. LAUNCH JUPYTER NOTEBOOK

Open the Anaconda Navigator, and click on Launch Jupyter Notebook.



If you are more advanced in using terminal/command line/shell, after navigating into the folder where you would like to save your notebook simply type “jupyter notebook”.

This should open a window in your default web browser program. In the upper right hand corner, click New, and then “Python 2” under “Notebooks”. You will get a window that looks similar to:



## 3. TEST YOUR ENVIRONMENT \*

Please copy the following code into the first line of your notebook, and then press Shift + Enter to execute:

```
import numpy as np
import pandas as pd
import networkx as nx
from bs4 import BeautifulSoup
import sklearn
import scipy
```

# DATA ANALYSIS WITH PYTHON - SYSTEM SETUP

Fall term 2016/2017

Course coordinator: Orsolya Vasarhelyi

website: [https://velf.github.io/Data\\_Analysis\\_with\\_Python](https://velf.github.io/Data_Analysis_with_Python)

[orsolya.vasarhelyi@gmail.com](mailto:orsolya.vasarhelyi@gmail.com)

```
from scipy import special
import matplotlib.pyplot as plt
%matplotlib inline
```

It should run without any errors! Then copy and paste the following code and press shift+enter:

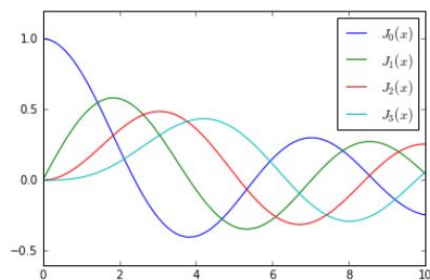
```
x = np.linspace(0, 10, 100)
fig, ax = plt.subplots()
for n in range(4):
    ax.plot(x, scipy.special.jn(n, x),
            label=r"$J_{%d}(x)$" % n)
ax.legend()
*Source: Roberta Sinatra
```

Your notebook should look like this:

```
In [10]: import numpy as np
import pandas as pd
import networkx as nx
from bs4 import BeautifulSoup
import sklearn
import scipy
from scipy import special
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [6]: x = np.linspace(0, 10, 100)
fig, ax = plt.subplots()
for n in range(4):
    ax.plot(x, scipy.special.jn(n, x),
            label=r"$J_{%d}(x)$" % n)
ax.legend()
```

Out[6]: <matplotlib.legend.Legend at 0x7f181d9499d0>



In [ ]: |

Congratulations! Your system is ready! :)