# Windows Installation Guide

ACM Summer School 2020

Please use this guide to install the important packages on your Windows system that would be needed for lab exercises.

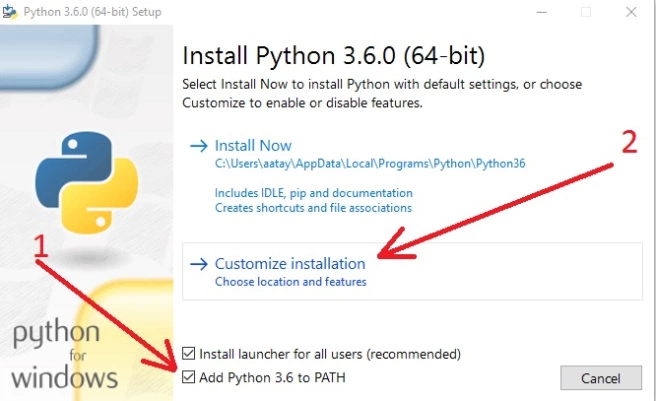
## System-wide Packages

### Python 3

Let’s start with the very first thing, Python!

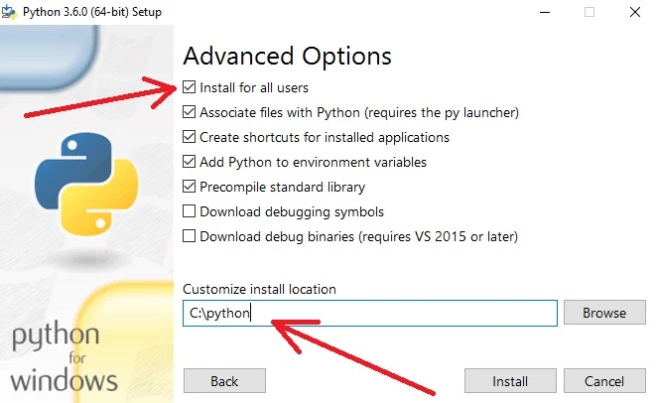
We’re going to use Python 3, specifically 3.6. Python can be installed by running the following [[1]](#bookmark=id.44sinio):

Download the latest Python 3.6 from <https://www.python.org> by going to Downloads->Windows and click on -> Latest Python 3 Release - Python 3.8.3 and at download the executable file acc. To your system configuration and run it. On running the .exe file :



After this, on the next window select all the Optional Features and click next.

After that do the changes as shown in the image.



And click on Install.

In the next screen, click “**Disable path length limit**” and click “Close” button to finish the installation procedure.

### Virtual Environment

Since Python comes in different flavours and versions, it is convenient to use a virtual environment to maintain Python packages independently for each project. You can install *virtualenv*, a tool to have isolated Python environments by following these commands[[2]](#bookmark=id.2jxsxqh):

|  |
| --- |
| $ pip install virtualenv |

Once the *virtualenv* is installed, you can create a project directory as:

|  |
| --- |
| $ mkdir scsummerschool |

Now change the current directory to the new one we just made and initiate a virtual environment with name *summerschoolenv*:

|  |
| --- |
| $ cd scsummerschool $ virtualenv summerschoolenv |

To start using the virtual environment, we activate it using:

|  |
| --- |
| $ summerschoolenv\Scripts\activate.bat |

We install all our packages in this virtual environment.

Once done, we can deactivate the virtual environment by running:

|  |
| --- |
| $ deactivate |

#### Python Packages

We will need a certain Python packages throughout the lab sessions for the Summer School. These packages are listed below, with the commands to install them.

We recommend that you go through rest of this section to install the packages, however, if you don’t want to install them individually, we provide a *win-requirements.txt* file [here](https://drive.google.com/file/d/1NVtyfdugXQ18kLExp82aOPPjCBYjBegm/view?usp=sharing), that can be used to download all the dependencies. Once you download the *win-requirements.txt* file, run the following command, after activating the virtual environment to install the dependencies:

|  |
| --- |
| $ pip install -r win-requirements.txt |

**Preprocessor**

Preprocessor is a preprocessing library for tweet data written in Python[[13]](#bookmark=id.1pxezwc).

Download the zip file from [here](https://github.com/s/preprocessor/files/1539402/tweet-preprocessor-0.4.0.zip).

Run the pip command as :

|  |
| --- |
| $ pip install <path to the file>\tweet-preprocessor-0.4.0.zip |

**Jupyter Notebook**

Jupyter Notebook is widely used in academia and industry for interactable Python development environment. Although, you can work with plain Python scripts as well, we highly recommend that you install it[[3]](#bookmark=id.z337ya).

|  |
| --- |
| $ pip install jupyter |

Once installed, you can run the notebook by:

|  |
| --- |
| $ jupyter notebook |

**Tweepy**

We need a Python wrapper for [Twitter API](https://developer.twitter.com/en/docs.html) in order to write code in Python. There are number of Python wrappers available, but we recommend using *tweepy*[[4]](#bookmark=id.3j2qqm3).

|  |
| --- |
| $ pip install tweepy |

**Matplotlib**

Matplotlib is a popular library used to create visualisations and plots[[5]](#bookmark=id.1y810tw).

|  |
| --- |
| $ pip install matplotlib |

**NetworkX**

NetworkX is used to create, manipulate, and study of the structure and functions of complex networks[[6]](#bookmark=id.4i7ojhp).

|  |
| --- |
| $ pip install networkx |

**NLTK**

Natural Language Toolkit is used for basic NLP tasks[[7]](#bookmark=id.2xcytpi):

|  |
| --- |
| $ pip install nltk |

**NumPy**

NumPy is majorly used to run computation on high-dimensional arrays and has defined methods for high-level mathematical functions[[8]](#bookmark=id.1ci93xb).

|  |
| --- |
| $ pip install numpy |

**Pandas**

Python Data Analysis Library is a common data analysis library[[9]](#bookmark=id.3whwml4).

|  |
| --- |
| $ pip install pandas |

**SpaCy**

SpaCy is another NLP library[[10]](#bookmark=id.2bn6wsx).

|  |
| --- |
| $ pip install spacy |

Troubleshooting: Install spacy by using the binaries provided [here](https://drive.google.com/open?id=1dX44uzBwNCakM1avaaO4I7oMk7vYOFnR).

To install a binary via .whl file, run:

|  |
| --- |
| $ python -m pip install <binary.whl> |

**Gensim**

Gensim is used for topic modelling on text[[11]](#bookmark=id.qsh70q).

|  |
| --- |
| $ pip install gensim |

**Scikit-learn**

It is a machine learning library for Python[[12]](#bookmark=id.3as4poj).

|  |
| --- |
| $ pip install scikit-learn |

**jsonpickle**

jsonpickle is a library used for serialization and deserialization of complex Python objects to and from JSON[[15]](#bookmark=id.2p2csry).

|  |
| --- |
| $ pip install jsonpickle |

**word\_cloud**

A little word cloud generator in Python[[16]](#bookmark=id.147n2zr).

|  |
| --- |
| $ pip install wordcloud |

Troubleshooting: <https://github.com/amueller/word_cloud/issues/134#issuecomment-228208102>

### Gephi

Gephi is the a visualization and exploration software for all kinds of graphs and networks[[14]](#bookmark=id.49x2ik5).

Gephi requires Java version 7 or above install Java by following the instructions [here](https://www.guru99.com/install-java.html).

The software can be downloaded by visiting the following link:

<https://gephi.org/users/download/>

And then run the zip file. Select all in SELECT ADDITIONAL TASKS. And continue.

## Twitter API Generation:

With respect to Twitter data collection, please follow the steps provided in the [link](https://drive.google.com/file/d/1vD2V42SVY_yxnJeay66TUWhIP5wHfdV1/view?usp=sharing).

## Github Account:

If you don’t have a [Github](https://github.com/) account. The complete steps to create Github can be found on this [link](https://www.wikihow.com/Create-an-Account-on-GitHub).

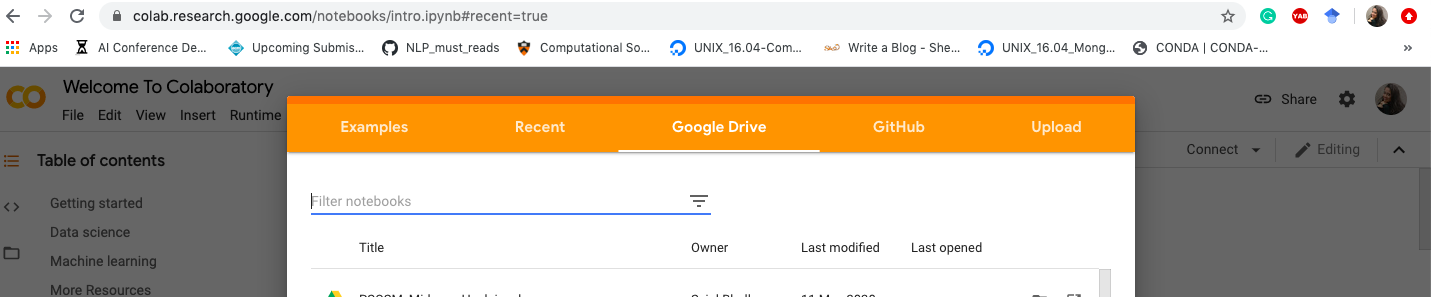
Once your Github account is set up, you can go to the top right corner and click at your profile photo

## 

Mention your Username in the following [Google Form](https://forms.gle/RdrWJWnX6geHEq5t5)

Google Colab

Follow the instructions given for [Google Colab](https://developers.google.com/earth-engine/python_install-colab#open_a_colab_notebook)



## References

[1]<https://aatayyab.wordpress.com/2016/12/25/install-programming-language-python-for-windows-10-64-bit/>

[2] <http://docs.python-guide.org/en/latest/dev/virtualenvs/#lower-level-virtualenv>

[3] <https://jupyter.org/install>

[4] <https://github.com/tweepy/tweepy>

[5] <https://matplotlib.org/users/installing.html>

[6] <https://networkx.github.io/documentation/stable/install.html>

[7] <https://www.nltk.org/install.html>

[8] <https://www.scipy.org/install.html>

[9] <https://pandas.pydata.org/pandas-docs/stable/install.html#installing-from-pypi>

[10] <https://spacy.io/usage/>

[11] <https://radimrehurek.com/gensim/install.html>

[12] <http://scikit-learn.org/stable/install.html>

[13] <https://github.com/s/preprocessor>

[14] <https://gephi.org/>

[15] <https://jsonpickle.github.io/>

[16] <https://github.com/amueller/word_cloud>