

# Recitation01

August 24, 2020

## 1 Recitation 1: setting up our machines!

- Download and install Python (optional)
- Download and install Python IDE (optional)
- Download and install Jupyter Notebooks
- Github account
- Access Jupyter notebook on the LU HPCC

### 1.1 Download and install Python (optional)

**Python 3** is a versatile programming language that is used for just about anything you can think of. We'll use Python for **statistical computation** in Population Health Data Science (PHDS)-I, and if you're interested, in Population Health Data Science (PHDS)-II and PHDS-III.

If you would like to download Python 3 on your own, local, computer, follow the below steps. This is optional. There is no requirement to install Python 3 on your machine. But often students find it easier to work with programming languages like Python on their own machine.

#### 1.1.1 Installation

You have two options for installing Python 3. I recommend installing Anaconda, a suite of the Python language and set of handy packages we'll use throughout the course. The other option is to install just Python 3.

**Anaconda** To install Anaconda, Python 3 and additional useful packages, use a browser to navigate to <https://www.anaconda.com/products/individual>, click "Download", and follow the default installation instructions.



Individual Edition

# Your data science toolkit

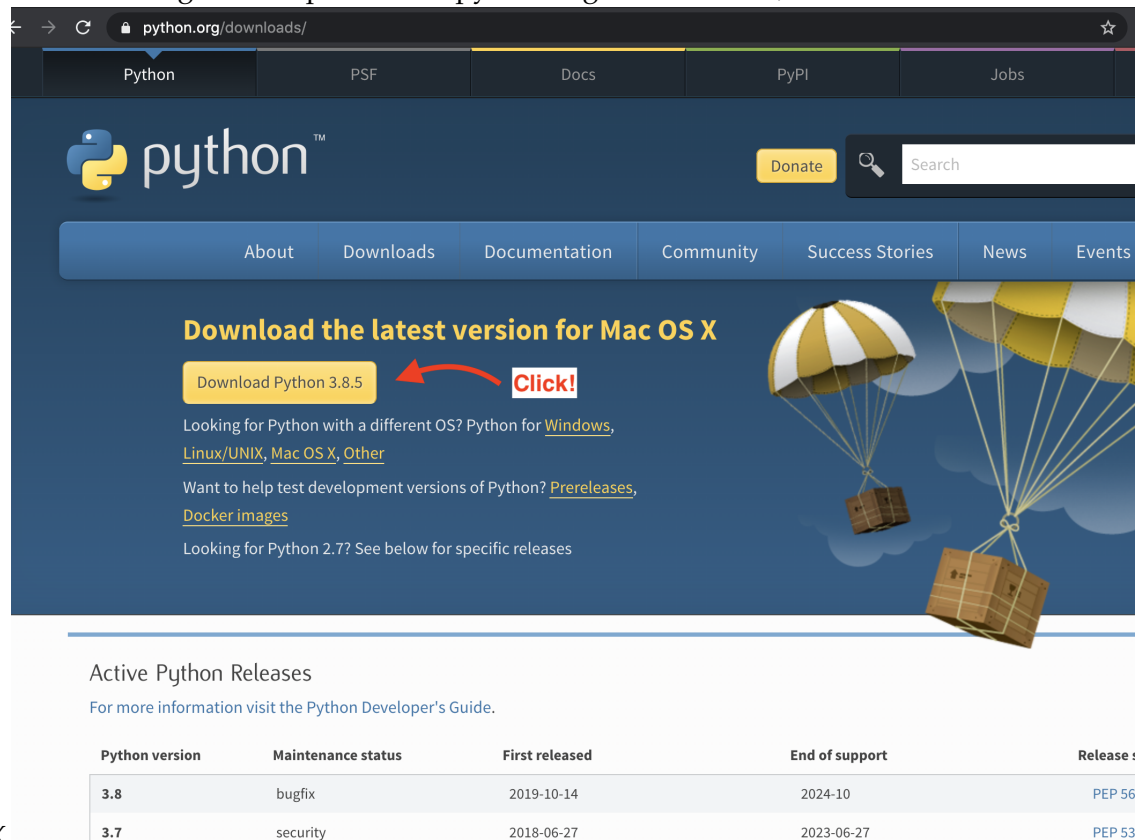
With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.



Click "Download"



**Just Python 3** Use a browser to navigate to <https://www.python.org/downloads/>, and click on



Download Python 3.X.X

Click on the downloaded package and follow the default installation prompts from the installer.

## 1.2 Download and install Python IDE (optional)

An integrated development environment, or IDE for short, is a computer program written to make your work easier. Some IDEs are general purpose software for making a list of programming languages easier to use, others were written for a single language. Python has a few IDEs that have had very good reviews. If you download Python 3 to your local machine, it may be worth checking out the following IDEs:

- [Spyder](#)
- [PyCharm \(the free community edition\)](#)
- [Pyzo](#)

If you try one of these IDEs and love (or hate) using it, let me know!

## 1.3 Download and install Jupyter Notebooks (optional)

A Jupyter notebook is software that integrates the Python programming language with a graphical interface and the ability to include text in between lines of code. Jupyter is a useful way to analyze data and share your results with others (for example your professor).

If you installed Anaconda then the software for Jupyter notebooks is pre-installed. You're pretty much done! First open an Anaconda navigator ([instructions here](#))

<https://docs.anaconda.com/anaconda/user-guide/getting-started/>) and then click Jupyter Notebook and install.

If you decided to manually install Python 3 then you'll need to also install Jupyter Notebooks.

### 1.3.1 For a Mac

Open a "terminal" window and run the following command: `curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py`

### 1.3.2 For Windows

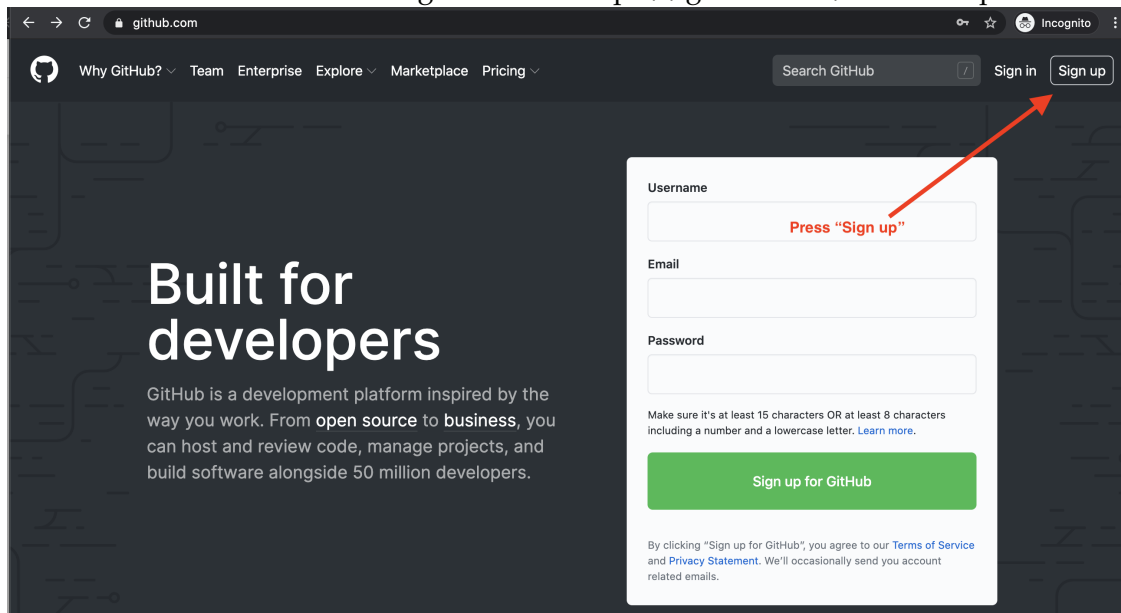
First follow [this tutorial](#) to install PIP and then type the following command into a windows terminal: `python -m pip install jupyter`

## 1.4 Git and GitHub

**Git** is a version control language—a programming language that keeps successive versions of your work to help you rewind to previous code and collaborate with a team. Code, data, visualizations, and any other documents needed for a project are all stored in a **git repository**. One repository should correspond to one project. **GitHub** is a set of servers allowing users to create accounts, their own Git repositories, and most importantly share their projects with others.

### 1.4.1 Creating a GitHub Account

Use a browser to navigate to <https://github.com/> and press "Sign up".

A screenshot of the GitHub homepage in a web browser. The page has a dark background with the GitHub logo and navigation links at the top. A white sign-up form is centered on the right. The form has fields for 'Username', 'Email', and 'Password'. A red arrow points from the 'Sign up' button in the top right corner of the browser window to the 'Sign up' button in the form. The form also includes a green 'Sign up for GitHub' button at the bottom and a link to the 'Terms of Service'.

Clicking "Sign up" takes you to a form where you'll need to enter \*

- \* A username
- \* An email address (use your LU-issued email address)
- \* A password

## Create your account

**Username \***  
ClutchTheMHawk ✓

**Email address \***  
lehighuniv@lehigh.edu ✓

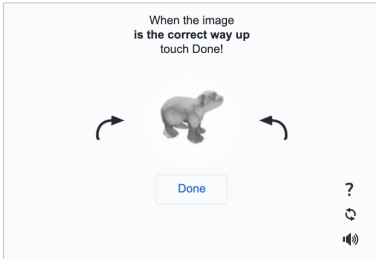
**Password \***  
..... ✓

Make sure it's at least 15 characters OR at least 8 characters including a number and a lowercase letter. [Learn more.](#)

**Email preferences**  
☒ Send me occasional product updates, announcements, and offers.

**Verify your account**

When the image is the correct way up touch Done!



Create account

### 1.4.2 Installing Git

We will work with Git using [GitHubDesktop](#). To download GitHubDesktop, navigate to <https://desktop.github.com/>, click on “Download for XXX” where XXX is Mac, Windows, or Linux, and follow the defaults given by the GitHubDesktop installer.

### 1.5 Access Jupyter notebook on the LU HPCC

If you do not want to download Python 3 and Jupyter to your local computer, we can use resources provided by Lehigh to use Jupyter and Python. To access Jupyter Notebooks and python via Lehigh University’s cluster computer, use a browser to navigate to <https://hpcportal.cc.lehigh.edu/>. You’ll be asked for a username and password. Type in your Lehigh username, password, and press enter.



You can use a Jupyter notebook by clicking “Interactive Apps” and then “Jupyter Notebooks”.

The screenshot shows the hpcportal.cc.lehigh.edu dashboard. The top navigation bar includes links for Open OnDemand, Files, Jobs, Clusters, Interactive Apps (selected), Lehigh Courses, and Help. The 'Interactive Apps' dropdown menu is open, showing categories: Desktops (Desktop), GUIs (ANSYS Workbench, Abaqus/CAE, GNU Octave, MATLAB, Maple, Mathematica, SAS, VMD, Visualization), and Servers (Jupyter Notebooks, RStudio server). A red arrow points to 'Jupyter Notebooks'. The main content area features the 'OPEN OnDemand' logo, a 'Message of the Day' section with a terminal-style ASCII art graphic, and a welcome message: 'Welcome to Sol... run on the head... compilation, an... non-polling da...'. It also lists system details: 'SLURM job su...', 'Files in /hor...', 'Apps in /sha...', and 'User quotas are set to 150GB on /home.' A security notice at the bottom states: 'Use of this system is subject to Lehigh University security and information policies available at http://www.lehigh.edu/security.'

powered by

Interactive Apps

Desktops

Desktop

GUIs

ANSYS Workbench

Abaqus/CAE

GNU Octave

MATLAB

Maple

Mathematica

SAS

VMD

Visualization

Servers

Jupyter Notebooks

RStudio server

Lehigh Courses

Servers

ME450 Fall 2020

Jupyter Notebooks

version: v0.10.0-7-gbf0977e

This app will launch Jupyter Notebook on one Sol node.

Allocation

One or t

Number of cores

2

Maximum number of CPU cores on Sol varies from 16 to 36.

Number of hours

5

Maximum wall time on debug is 1 hour, otherwise 48 hours.

Partition

debug partition, 16 cores, 1 hour, 3.5GB/core

Please select a partition from the drop-down.

Activate the conda environment you want to use (drag text area to enlarge)

- Activate the conda environment you want to use. anaconda/python\_3.6 is loaded by default
  - leave field blank - Use base environment
  - biofluids - source activate biofluids
  - mlidl - source activate mlidl
  - nlp - source activate nlp
  - phonopy - source activate phonopy
  - tensorflow - source activate tensorflow
  - Enter commands to use your custom environment
    - Unload anaconda/python\_3.6 if you are using a different python install.

☐ I would like to receive an email when the session starts

Launch

Next you'll be presented with an allocation screen.

When you use a resource on Lehigh's cluster computer you need to request the exact resources you need. After a request is made your computer is put "in line" until the cluster computer allocates your requested resources for the time you requested. If you request more cores or more time it may take a few minutes for the cluster computer to find the right resources for you. When you specify the number of cores and number of hours requested, click "Launch"



Interactive Apps

Desktops

Desktop

GUIs

ANSYS Workbench

Abaqus/CAE

GNU Octave

MATLAB

Maple

Mathematica

SAS

VMD

Visualization

Servers

Jupyter Notebooks

RStudio server

Lehigh Courses

Servers

ME450 Fall 2020

Jupyter Notebooks version: v0.10.0-7-gbf0977e

This app will launch Jupyter Notebook on one Sol node.

Allocation

Leave blank if using your default allocation.

Number of cores

2

Maximum number of CPU cores on Sol varies from 16 to 36.

Number of hours

5

Maximum wall time on debug is 1 hour, otherwise 48 hours.

Partition

debug partition, 16 cores, 1 hour, 3.5GB/core

Please select a partition from the drop-down.

Activate the conda environment you want to use (drag text area to enlarge)

- Activate the conda environment you want to use. anaconda/python\_3.6 is loaded by default
  - leave field blank - Use base environment
  - biofluids - source activate biofluids
  - mdl - source activate mdl
  - nlp - source activate nlp
  - phonopy - source activate phonopy
  - tensorflow - source activate tensorflow
  - Enter commands to use your custom environment
    - Unload anaconda/python\_3.6 if you are using a different python install.

☐ I would like to receive an email when the session starts

Launch

Session was successfully created.

Home / My Interactive Sessions

Interactive Apps

Desktops

Desktop

GUIs

ANSYS Workbench

Abaqus/CAE

GNU Octave

MATLAB

Maple

Mathematica

SAS

VMD

Visualization

Servers

Jupyter Notebooks

RStudio server

Lehigh Courses

Servers

ME450 Fall 2020

Jupyter Notebooks (4554324)

Created at: 2020-08-23 16:00:04 EDT

Time Requested: 5 hours

Session ID: bfdab071-af66-4617-ac9b-6bedbeb9a5c6

Please be patient as your job currently sits in queue. The wait time depends on the number of cores as time requested.

You should see your requested “job” Queued.

After a few minutes, when the cluster has made space for your requested time and number of cores, you’re screen will change to the following

And finally you will be able to access a Jupyter Notebooks server.