Unstructured Data Analysis

Setting up Anaconda Python and Jupyter

Fall 2017

Written by Yoonjung Kim with some small edits by George Chen

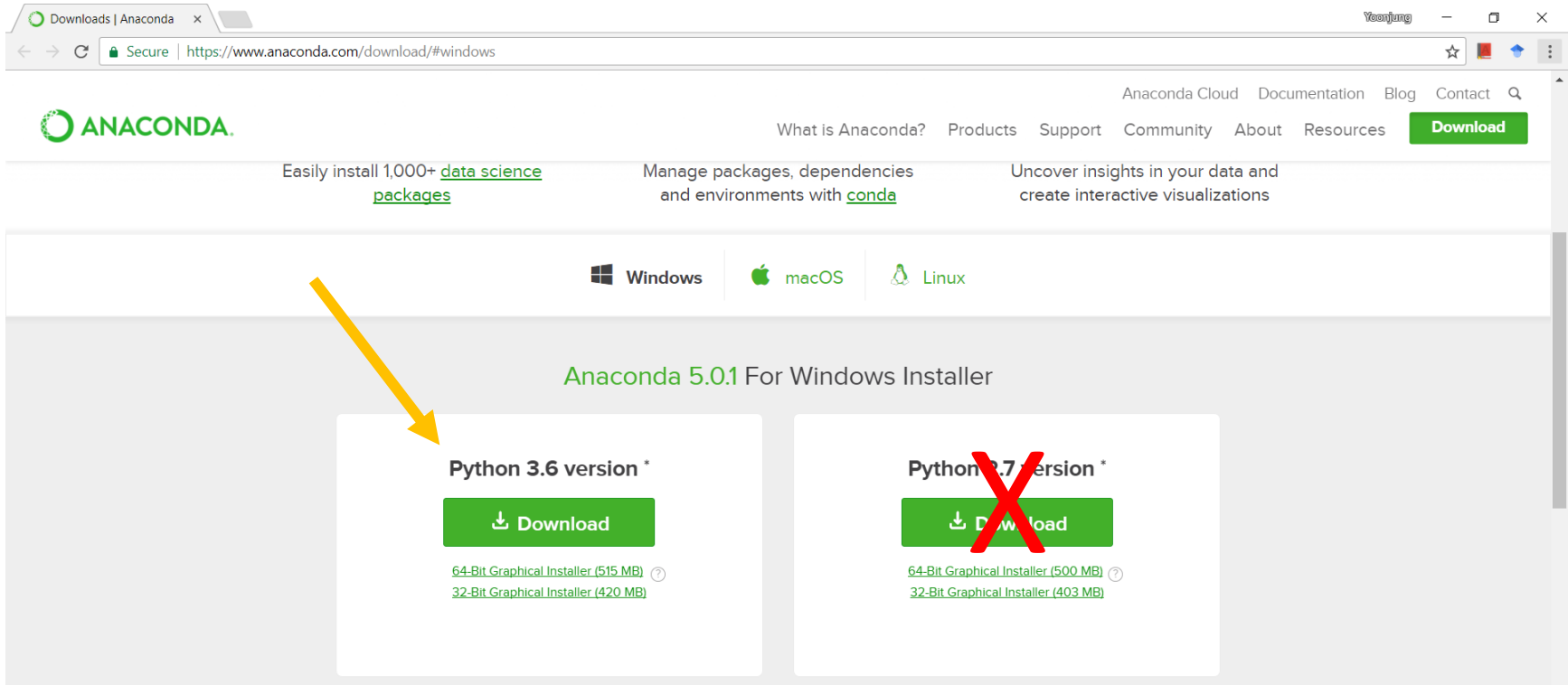
Preface

Anaconda Python can be installed on your local machine as well as on an Amazon AWS instance. You may work on your local machine for the homework assignments and final exam. However, it is your responsibility to make sure your code runs in a Linux environment (on Windows, there are a few ways to do this such as using “Bash on Ubuntu on Windows”, cygwin, or running a Linux virtual machine within Windows).

HW1 will not require an enormous amount of computational power. However, this will change as we progress to HW2 and HW3 when you will very likely want to switch to using AWS.

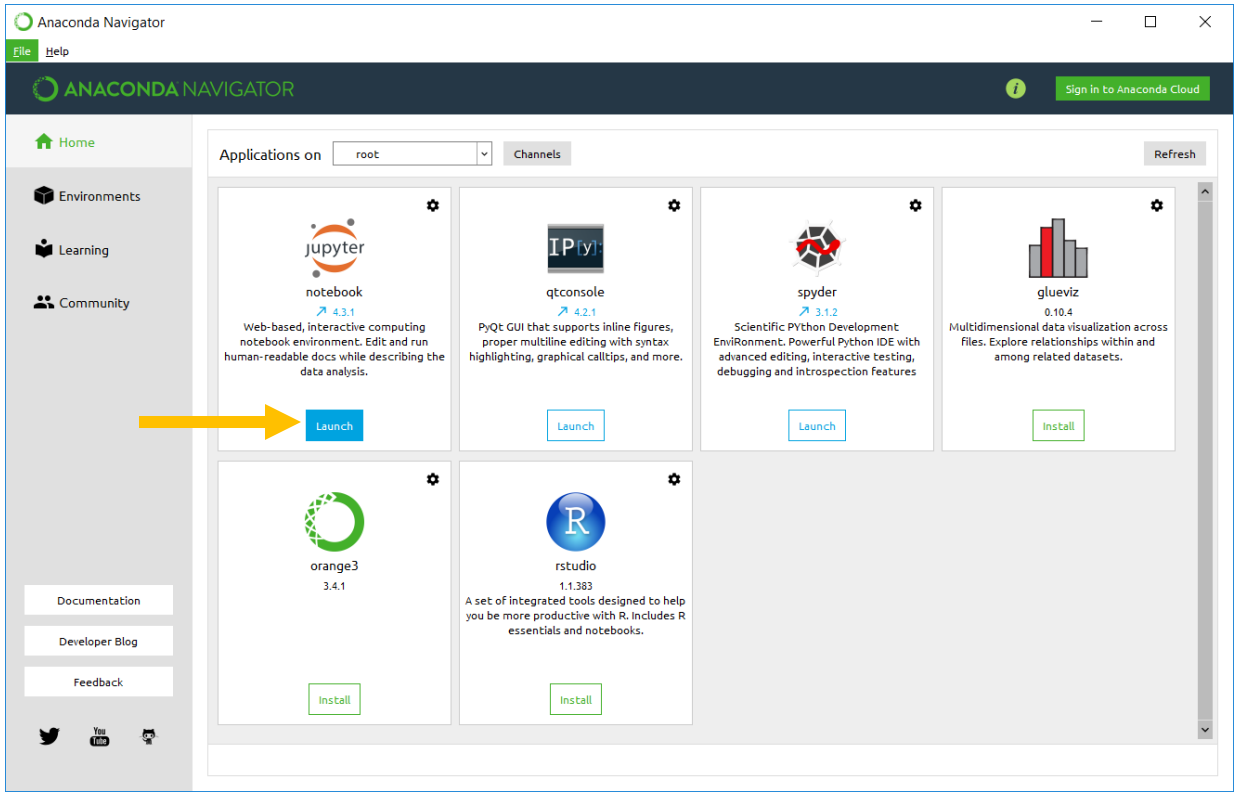
Set up your local machine

- Go to https://www.anaconda.com/download/
- Download & install Anaconda for Python 3.6 version that fits your OS



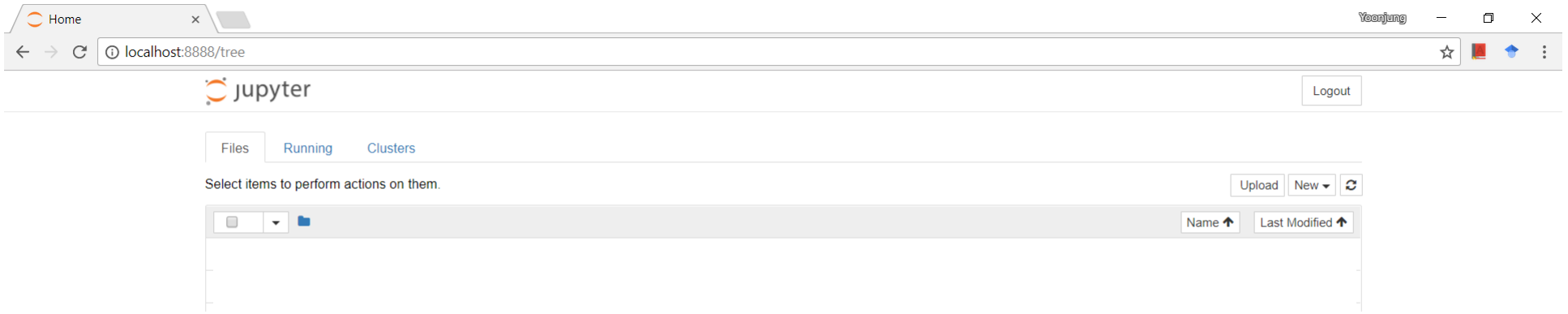
Set up your local machine

- Open Anaconda Navigator
- Launch Jupyter notebook



Run Jupyter notebook

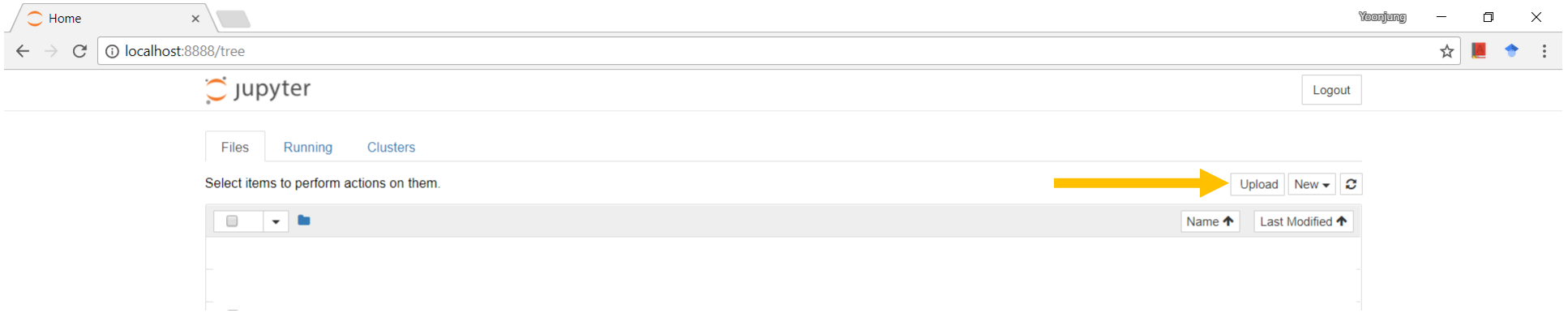
- If you see this… Jupyter setup completed!!



- Jupyter notebook lets you share your code and its result
- You can also use markdown language

Upload notebook

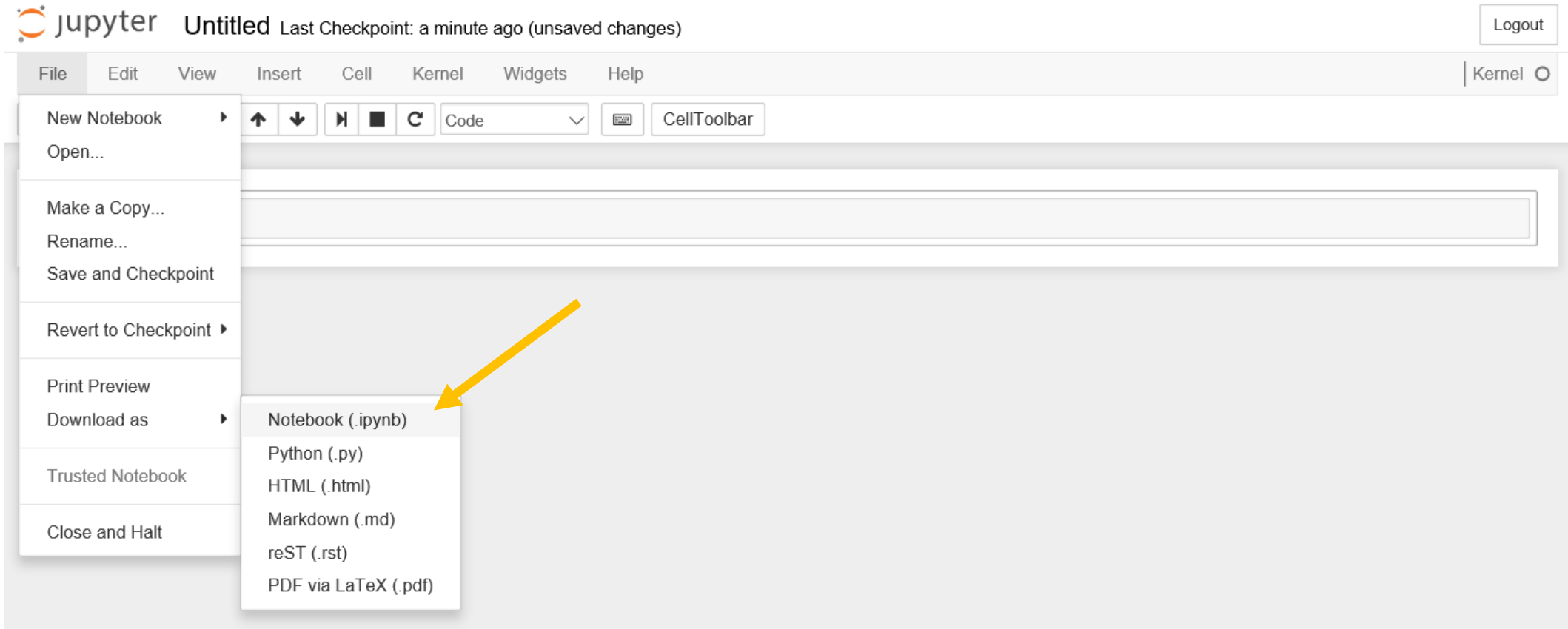
- Click Upload and select *.ipynb file



- Uploaded file will appear on the list below
- Click the file to open
- More tutorial on Jupyter notebook: https://youtu.be/k7WXVWej-NY

Download IPython notebook

- File > Download as > Notebook (.ipynb)



Notes

- You can upload and download any type of file using the interface
- Jupyter notebook will terminate in the server if you close the first the bash shell
  - To prevent this, use SCREEN command in Linux
  - Ref: https://www.rackaid.com/blog/linux-screen-tutorial-and-how-to/
- localhost:8888 connection will disappear if you close the second bash shell

Python package installation

Use Conda

- Using Conda makes your life much easier
- Basic packages such as numpy, scipy, scikit-learn (sklearn), and nltk are already installed with Anaconda
- Before installing a package, refer to the official documentation

- I'll show you an example of spaCy and nltk which you need for HW1

Install spaCy

- Open Jupyter notebook
- Click New > Terminal

- Type the following lines of code

  conda config --add channels conda-forge

  conda install spacy

  python -m spacy download en

This is an additional line downloading the English model in spaCy