

# Unstructured Data Analysis

## Setting up Anaconda Python, Jupyter, and spaCy

Fall 2020

Written by Yoonjung Kim in Fall 2017 with some edits by George Chen since then

# Preface

Anaconda Python can be installed on your local machine. You are expected to work on your local machine for homework and quizzes, although HW3 will also use Google Colab. 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.

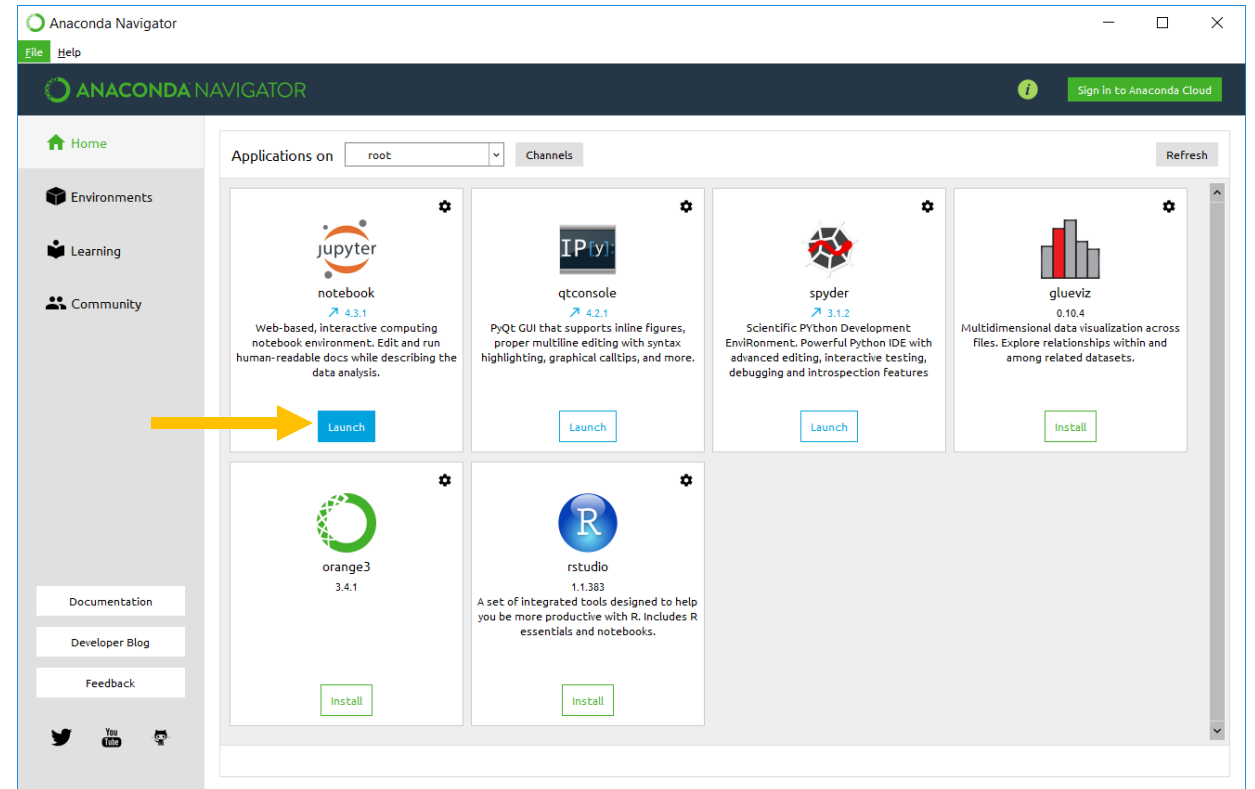
# Set up your local machine

Download & install Anaconda for Python 3 for your OS

<https://www.anaconda.com/distribution/>

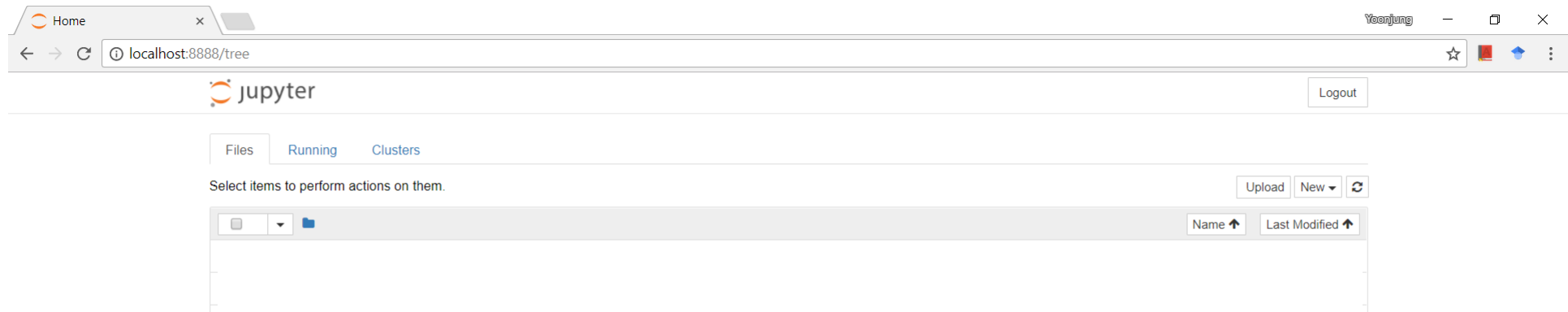
# Set up your local machine

- Open Anaconda Navigator
- Launch Jupyter notebook



# Run Jupyter notebook

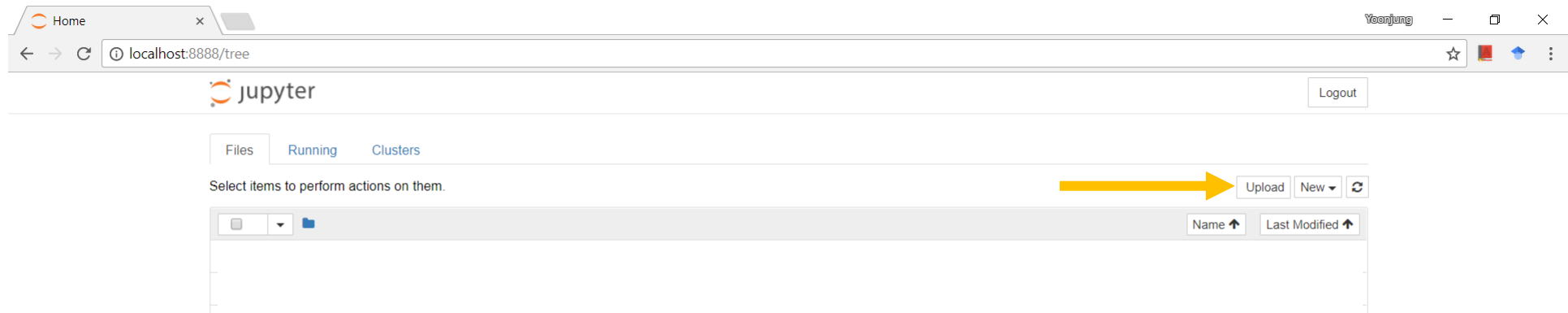
- If you see this... Jupyter setup completed!!



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

# Upload notebook

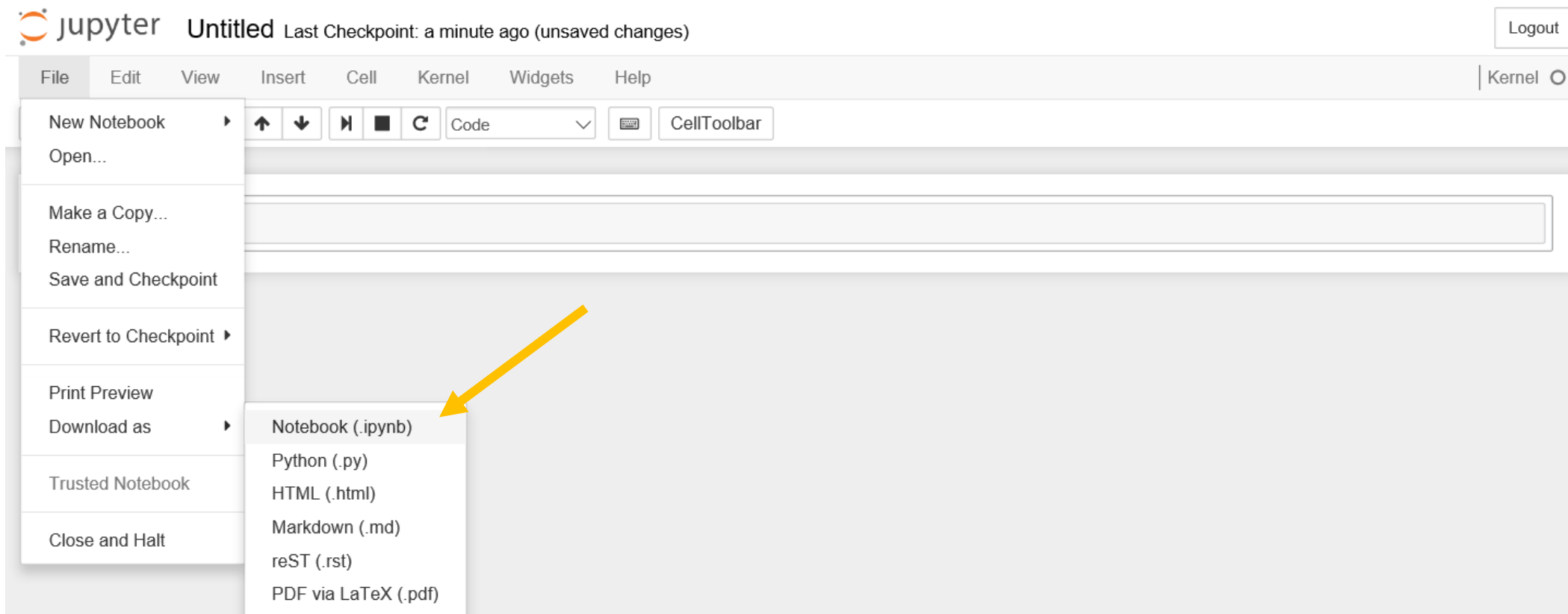
- Click Upload and select \*.ipynb file



- 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)



Python package installation

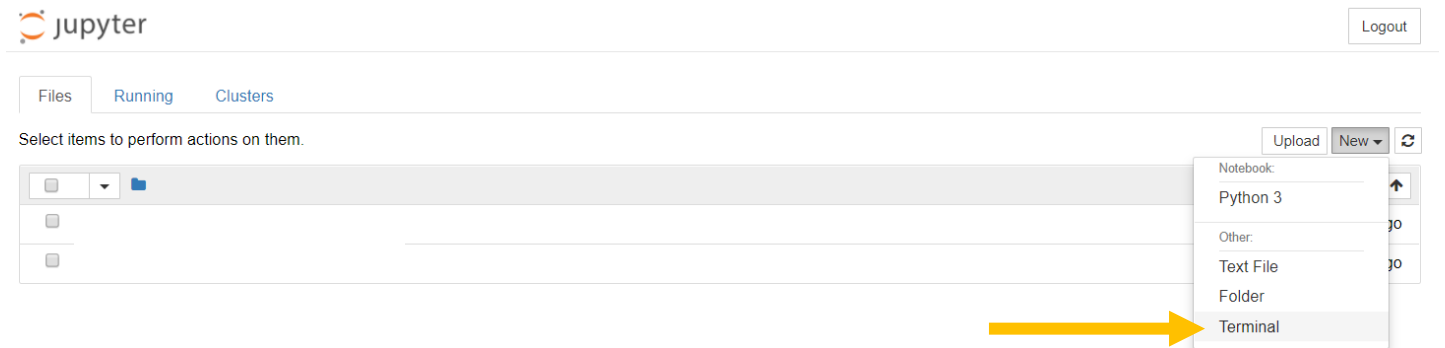


# Use `pip` or `conda`

- Using `pip` or `conda` makes installation (and uninstallation) straightforward (and if the installation fails, it will typically do so gracefully without screwing up any other packages)
- Note that many basic packages such as `numpy`, `scipy`, `scikit-learn` (`sklearn`), and `nltk` are already included with Anaconda, so you do *not* have to try to separately install them
- Before installing a package, refer to the official documentation

# Install spaCy

- Open Jupyter notebook
- Click New > Terminal



- Type the following lines of code

```
pip install -U spacy
pip install -U spacy-lookups-data
python -m spacy download en_core_web_sm
```