Introduction to Jupyter

2020-01-30

Jupyter Notebook

“The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain *live code*, *equations, visualizations and narrative text*.”

Sample Notebooks

- A gallery of interesting Jupyter Notebooks · jupyter/jupyter Wiki
- Iris Species
- Heart Disease UCI

JupyterLab

Documentation: Overview — JupyterLab 1.2.6 documentation

“JupyterLab is a next-generation web-based user interface for Project Jupyter.”

Installation (JupyterLab)

Instruction for installation: https://jupyterlab.readthedocs.io/en/stable/getting_started/installation.html

What are pip/conda/pipenv?

- In general, they are package managers. Once you have them, you can install packages you want.
  - pip (recommended): https://pip.pypa.io/en/stable/installing/
  - pipenv: https://pipenv-fork.readthedocs.io/en/latest/
  - conda: https://docs.conda.io/en/latest/

Installation (Jupyter Notebook)

Check Project Jupyter | Installing the Jupyter Software

Follow the instructions under “Getting started with the classic Jupyter Notebook”

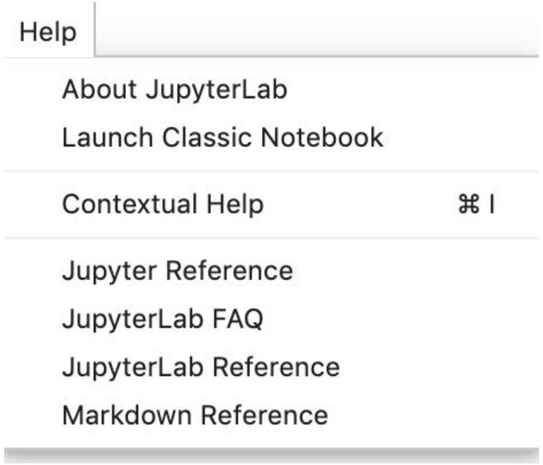
Start a Notebook

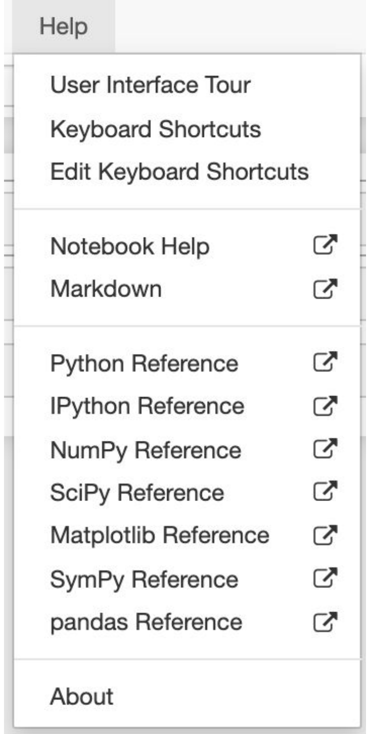
Start Jupyter notebook by running the command: `jupyter notebook`

Start JupyterLab by running: `jupyter lab`

Edit/Command Mode and Shortcuts

- Edit mode
- Command mode
- Shortcuts
- Documentation





Basic Operations

- Create and rename a notebook
  - A notebook document is in the format of *.ipynb
- Run a cell
- Add a cell
- Delete a cell
- Run all cells

- Run a line in console

Run a command line

Use `!` in front of the command line, and then run the cell

For example:

```
!pip list
```

* what if you want to run command lines in a for loop or with some conditions? Check the post here: Running for loop terminal commands in Jupyter

Preparation: downloading the dataset

- Download the dataset for the novel coronavirus: Novel Coronavirus 2019 Dataset

You can check the excellent dashboard from JHU here:
https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

Practice 1: Install libraries

Try to pip install `numpy`, `pandas`, and `seaborn` using Jupyter

For example, you can run `!pip install numpy`

`numpy` and `pandas` are libraries for data preprocessing

`seaborn` contains some standard visualization methods

We will introduce more about them in the later labs.

Practice 2: exploring the dataset

- Read the .csv file using `pandas.read_csv(filename)`
- Check the file information by using `.info()`
- Check a specific column
- Check the first a few rows in the dataframe by using `.head()`
- Check the statistics of each numeric column by using `.describe()`

Practice 3: simple visualization

- processing the data for plotting
  - Get unique dates: `list(df['Last Update'].unique())`
  - Get the data for `confirmed, recovered, and death` for each date
  - Create a new dataframe for plotting,
- Plot a line chart

Markdown cell and Code cell

In default, a created cell is a code cell.

Change it into a markdown cell:

Hit key `Esc` into Command Mode, then hit key `m`

（hit key `y` to change to the code mode）

Or

Jupyter Notebook: for selected code cell, click `Cell`, then `Cell Type`, then `Markdown` on the toolbar

Jupyter Lab: for selected code cell, click `Code` then `Markdown` on the toolbar

Markdown Cells

“Markdown is an easy-to-read, easy-to-write syntax for formatting plain text.”

Find the grammar here: Markdown Cells — Jupyter Notebook 6.0.3 documentation

Your can also check JupyterLab -> Help -> Markdown Reference,

Or Jupyter Notebook -> Help -> Markdown

*built-in magic

Documentation: [Built-in magic commands]()

Check all the magic by running: `%lsmagic`

For example, try to run:

```
%pwd
```

```
%%HTML

<iframe width="560" height="315"
src="https://www.youtube.com/embed/HW29067qVWk"
frameborder="0" allow="accelerometer; autoplay;
encrypted-media; gyroscope; picture-in-picture"
allowfullscreen></iframe>
```