USTH MASTER SPACE

Earth Observation - Astrophysics - Satellite Technologies

USTH

VIETNAM FRANCE UNIVERSITY

Data Analysis and Visualization

Hoàng Đức Thường

Department of Space and Applications (DSA), USTH

Hanoi, 2021

Useful Reading Material

- Book: Practical Statistics for Astronomers - Wall & Jenkins
- Book: Statistical Methods in Experimental Physics - Frederick James
- http://astronomy.swin.edu.au/~cblake/stats.html (Statistics course-Swinburne University)
- http://cs229.stanford.edu/section/ (Machine Learning-Stanford University).
- http://ircamera.as.arizona.edu/Astr_518/ (Instrumentation and Statistics-University of Arizona)

My Class Rules

- Be kind to people
- Try your best
- Be a good friend
- Do not laugh at others
- Listen to the teacher

Syllabus/Evaluation

Table1: Schedule

Items	Lecture	Tutorial/ Exercise	Practice/ Assignment	Lab-work	Total
No. of hours	14	0	0	0	14

Table 2: Assessment/Evaluation

Component	Attendance	Exercises	Practical	Reports	Midterm	Final
Percentage %	10		20	0	0	70

After this course

- Understand the statistic quantities: Mean, median, variation, correlation.
- Know and distinguish some popular probability distributions: Gauss (Normal), Poisson, Uniform, Chi-square.
- Understand χ^2 statistic for hypothesis testing and parameters estimation of a model.
- Visualization/manipulation plotting data using Numpy, Scipy, Astropy, Matplotlib/Python.

Content

- **Lecture 0**: Introduction to Jupyter notebook and Python.
- **Lecture 1**: Fundamental statistics quantities: Mean, median, standard deviation (variance), correlation.
- **Lecture 2**: Probability distributions: Binomial, Uniform, Normal (or Gaussian), Poisson, Gamma, T-Student's, Chi-square, the central limit theorem.
- **Lecture 3**: Hypothesis testing, model fitting and parameter estimation.
- **Lecture 4**: Principal components analysis (PCA) and Bayesian methods.
- **Lecture 5**: Basic Machine Learning using Scikit-learn tool.

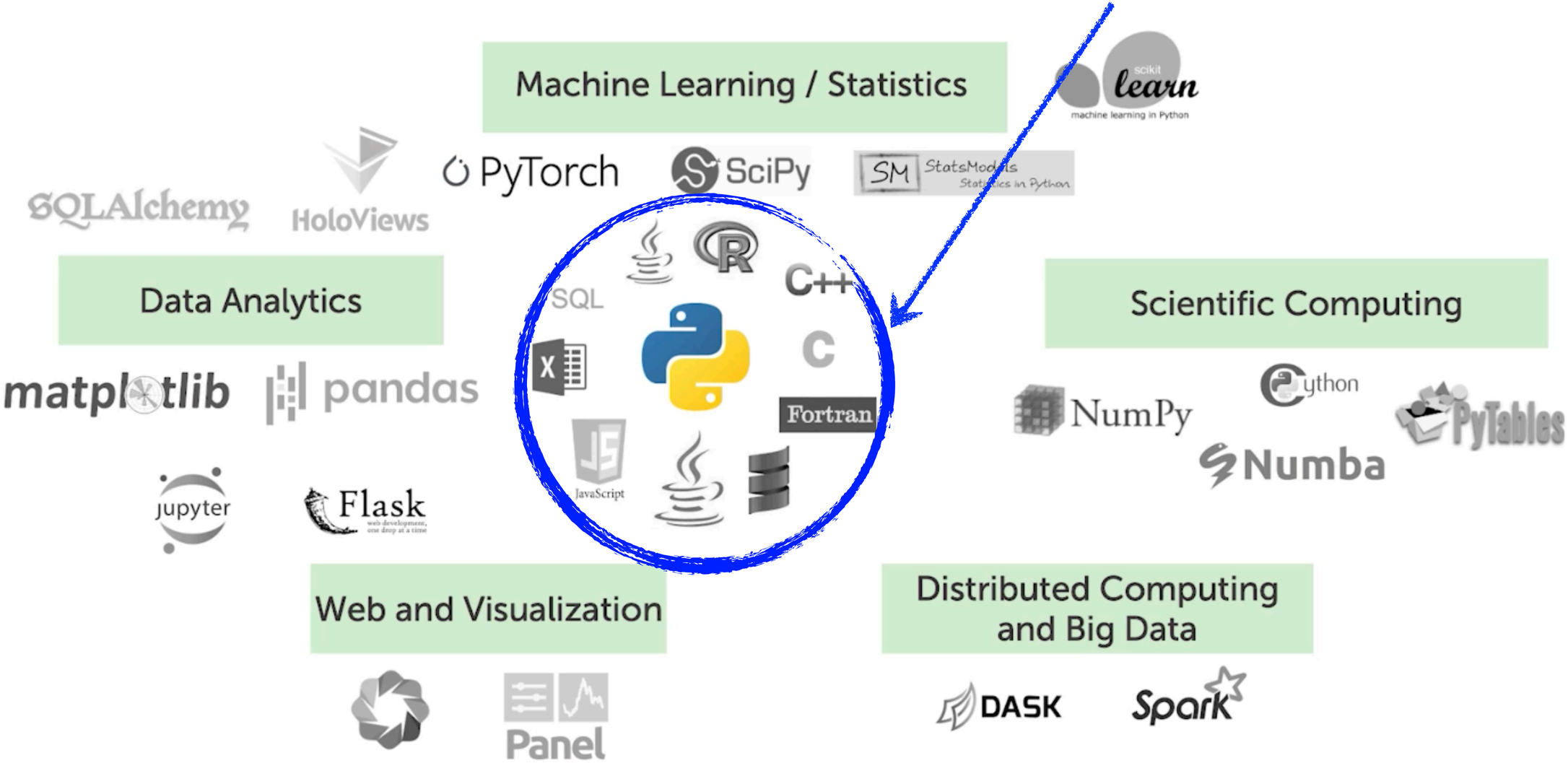
Note: Lecturers are associated with Jupyter Notebooks (python)

Lecture 0: Introduction to Jupyter notebook and Python.

- The required tools: Jupyter notebook/jupyterlab and python.
- Anaconda is a platform for data science, it manages, installs packages: https://www.anaconda.com/products/individual
- We can obtain Jupyter notebook via Anaconda.

In this class I will review tools for data analysis and visualization: Python and Notebook.

Data Science & Anaconda

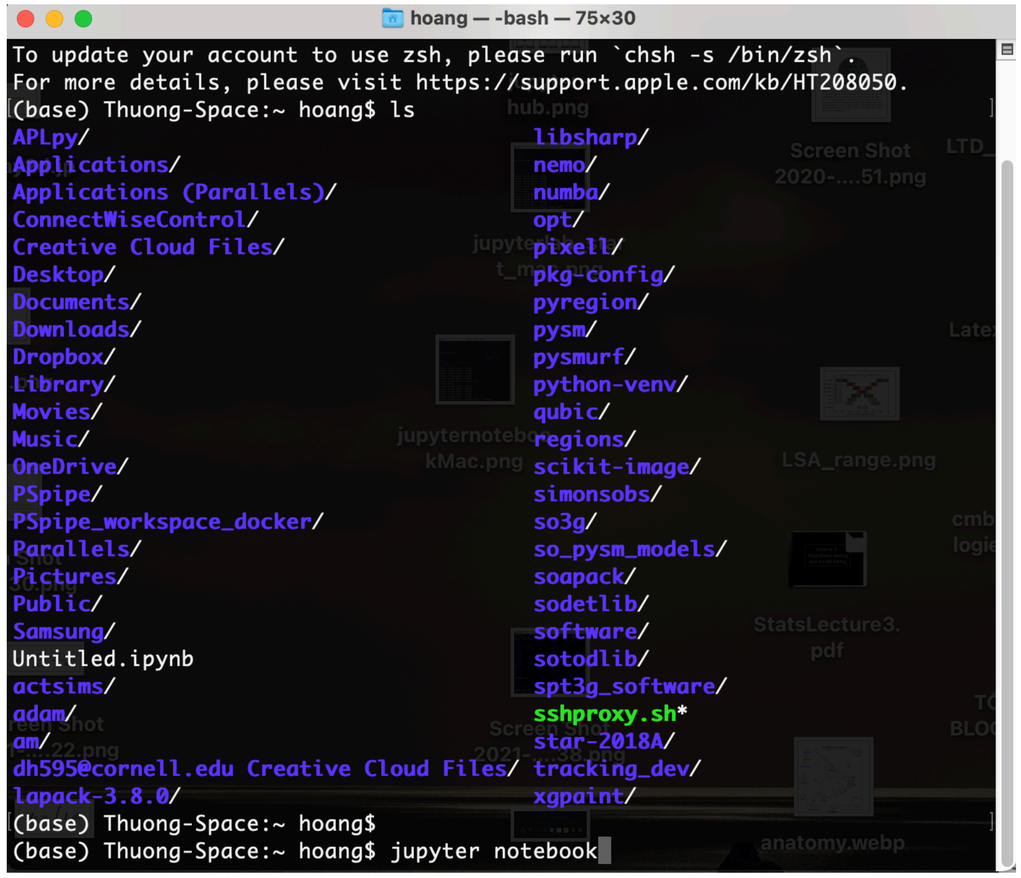


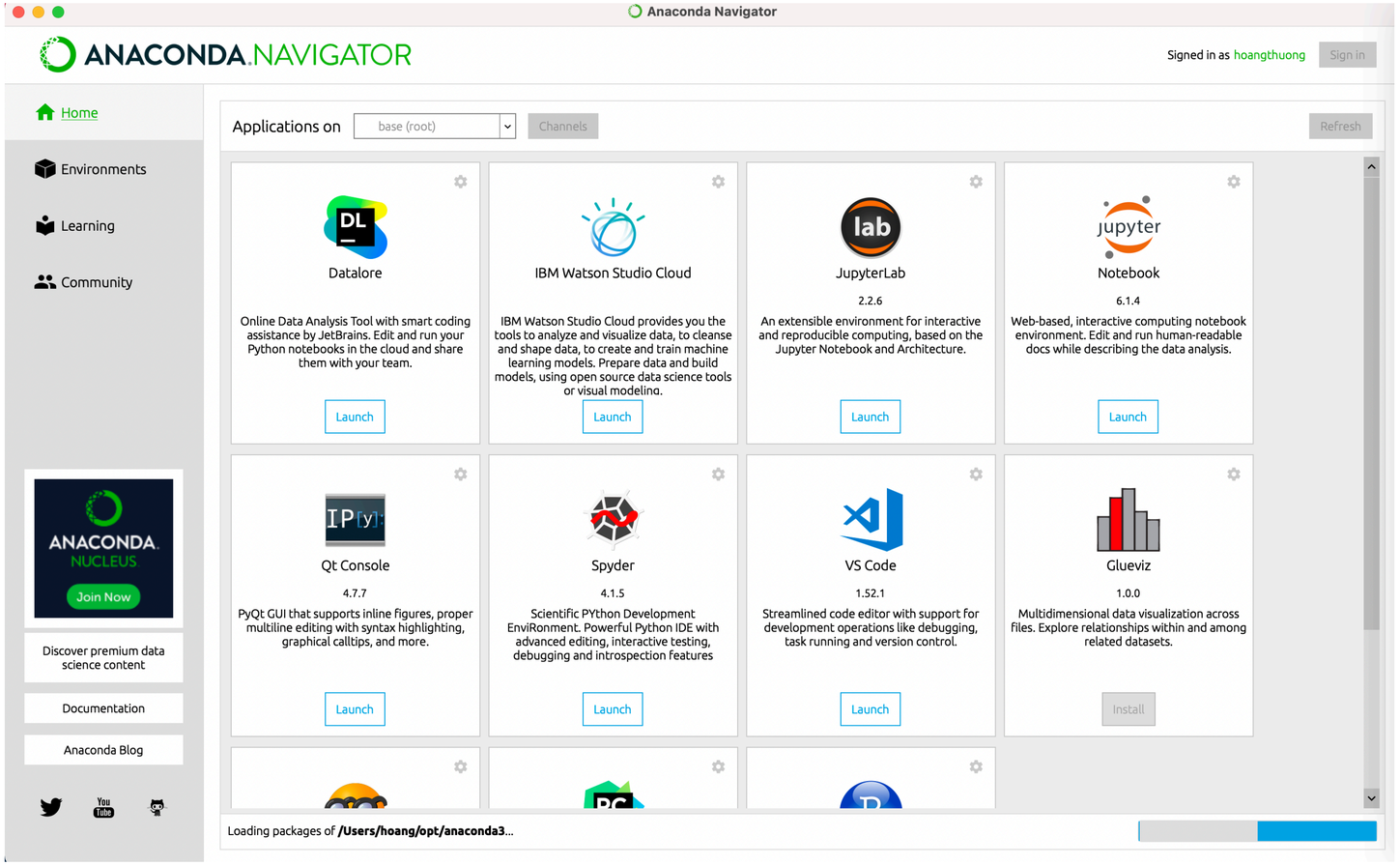
- Data science is so complicated with many packages, programming languages.
- Anaconda platform helps to manage, install, upgrade, run these packages on different OS: Mac, Linux, windows.

Anaconda features

- Anaconda (Open source Toolkit for Data Science): https://www.anaconda.com/open-source
  - DOWNLOAD: https://www.anaconda.com/products/individual
  - Getting start tutorial (login require!): https://anaconda.cloud/tutorials/8d29a356-46f8-4c5f-9fe8-3b3458b5a252
  - **Fundamentals: Jupyter (interactive programming web-base), Numpy (array format), Scipy (scientific libraries), pandas (data structures).**
  - **Data Visualization: Matplotlib, bokek (interactive web-base), Grafana, plotply, Holoviz**
  - Machine Learning: Keras (neural network), Tensorflow (tensor calculus), pytorch (deep learning), scikit-learn (classification, regression, clustering, …).
  - Image processing: Pillow (general), scikit-image (algorithms), OpenCV (computer vision library).
  - Data pipeline, Natural Language Processing, …
  - Interplay with other languages: Fortran, C, Spyder (Matlab-liked), …

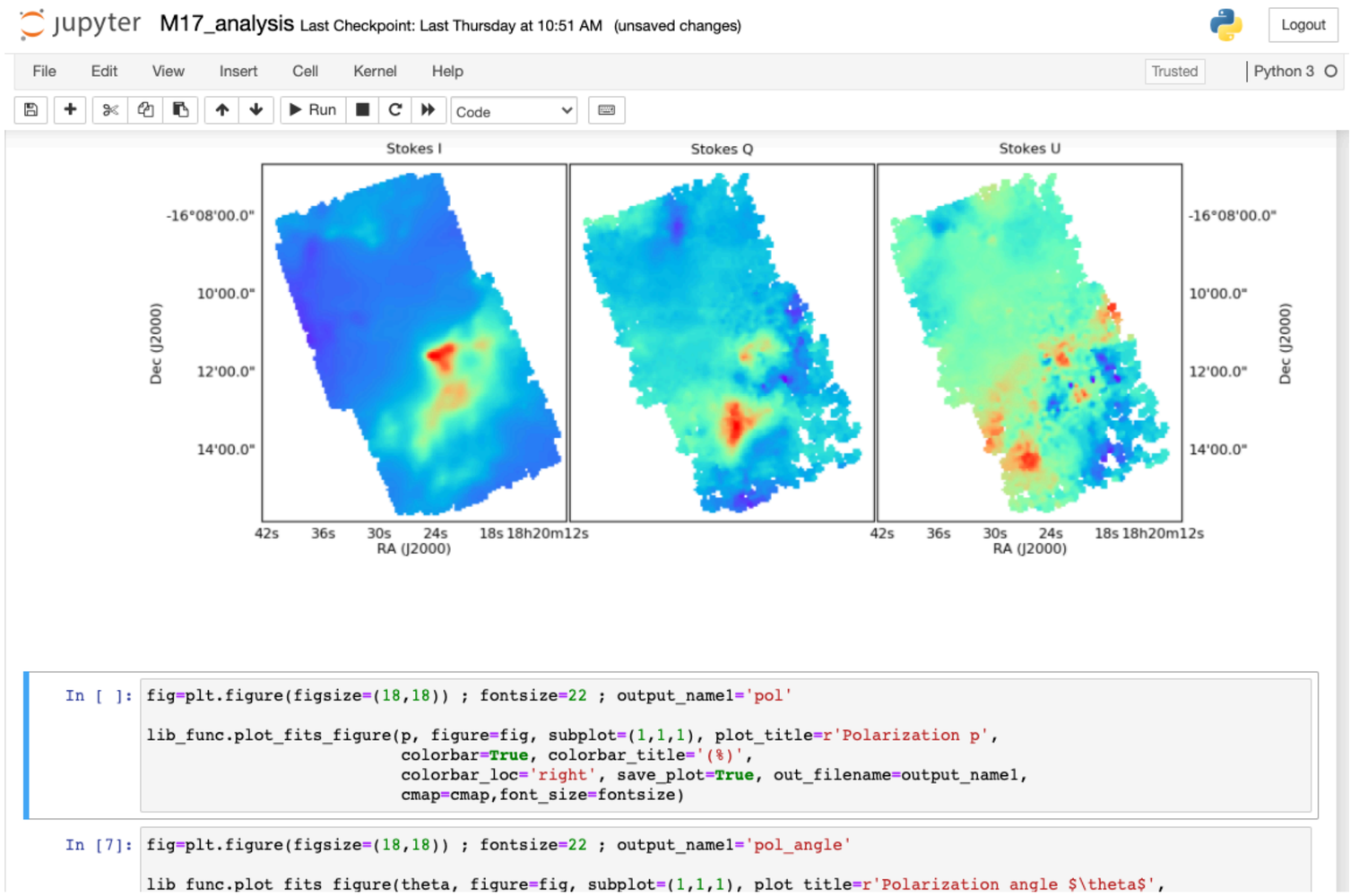
1. Jupyter notebook



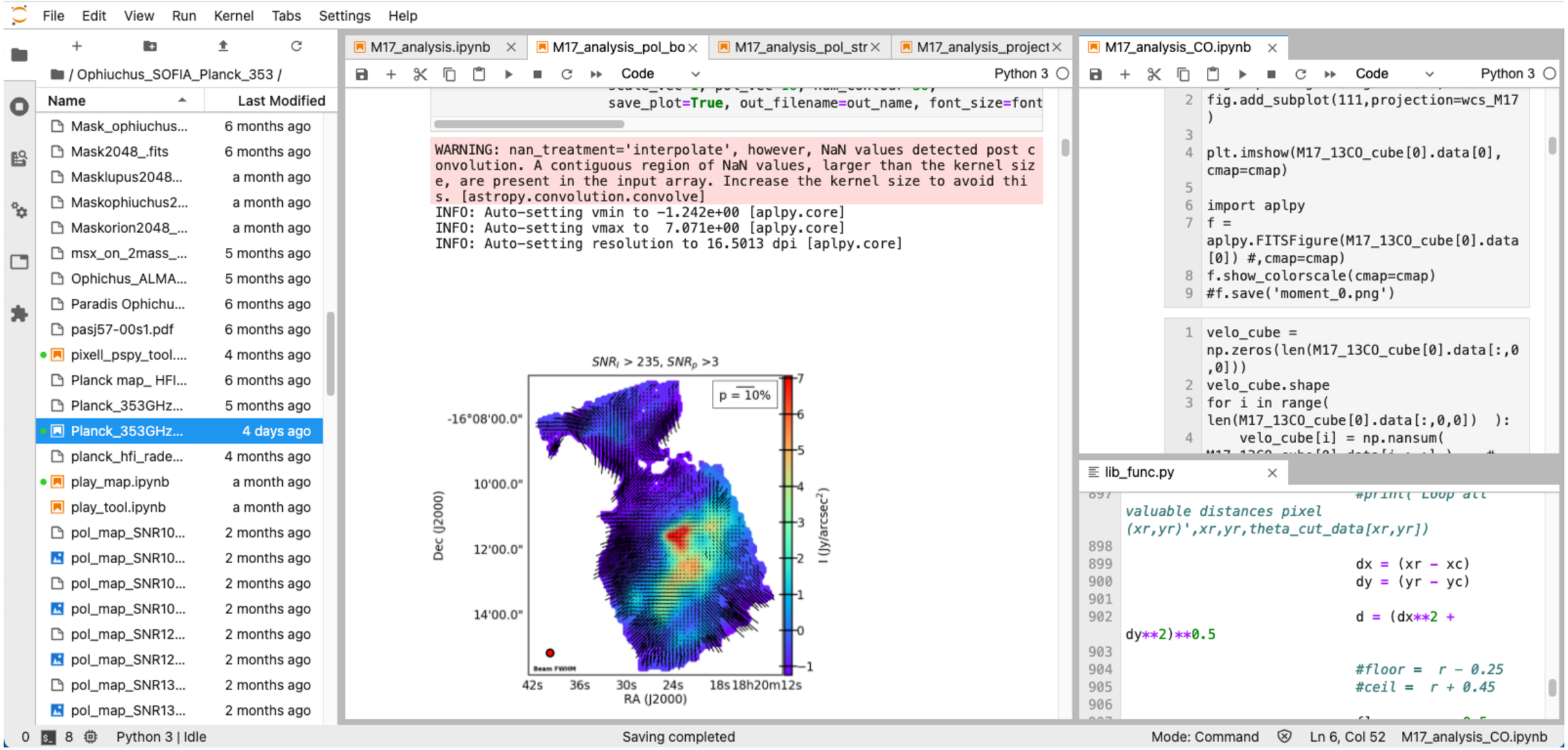


Jupyter notebook: A coding editor

- An interactive web-based programming.
- Allow to create and share live code, equations, visualizations via Github, nbviewer: https://nbviewer.jupyter.org/
- Support over 40 programming languages: Python, R, …
- Export to HTML, Latex, pdf, …
- It is useful for lecture notes, released software tutorials.



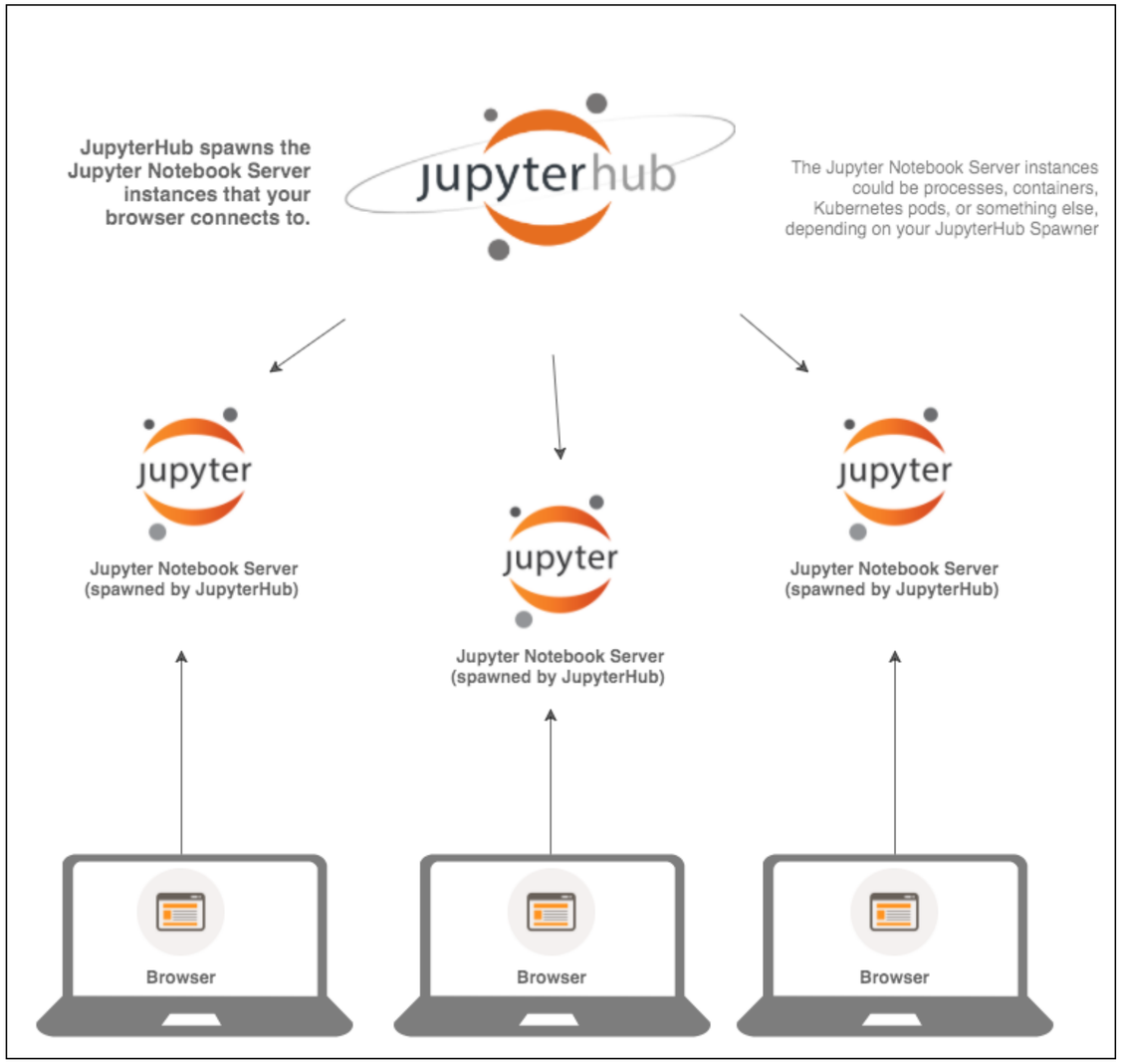
2. Jupyterlab



- JupyterLab is the development of Jupiter notebook, it is flexible interactive web-based environment.

3. JupyterHub for Multi-Users

- Source code: https://github.com/jupyterhub/jupyterhub
- JupyterHub tutorial: https://github.com/jupyterhub/jupyterhub-tutorial
- This is a good option for a remote serve



Summary

- The easiest way to have JupyterLab is to install via Anaconda, especially a windows system.
- Using Anaconda/pip to install packages: Jupyter notebook, JupyterLab, numpy, scipy, astropy, Spyder.
- If you are using a Mac OS, Linux, you can install JupyterLab independently.