





# Jupyter Notebooks 101

## A Reproducible Research Workshop

Simon Stone  
*Research Data Services*  
*Dartmouth College*





# About the Reproducible Research Group

- Joint venture of **Research Computing @ ITC** and **Research Data Services @ Library**
- Consult with **experts** on
  - research data management,
  - data visualization,
  - biomedical research support,
  - spatial data and GIS,
  - high performance and research computing,
  - statistical analysis,
  - economics and social sciences data
- **Meet** the people on campus that support your reproducible research lifecycle
- **Engage** in community discussions to learn from other researchers on campus
- Attend a workshop to **learn** practical tools and tips



# About Research Data Services

## Research Data Management

Data Management Plans (DMPs) for sponsored projects

Finding and using 3rd party data

Collection and cleaning of data

Organization and documentation

Publishing and Repositories

## Data Analysis/Visualization

Textual, numeric, spatial data

Reproducible research workflows

Scripting in R: tidyverse core package (i.e. ggplot, dplyr, tydr, tibble, etc.)

Scripting in Python: NumPy, SciPy, Pandas, Scikit-learn, Matplotlib, Seaborn, (OpenCV, PyTorch, TensorFlow, Tesseract, NLTK, etc.)

## Computational Scholarship

Computational project planning

Collections as Data

Storytelling with data and visualizations

Text and data mining

Digital Humanities support

Computational Pedagogy



# Work with us

[ResearchDataHelp@groups.dartmouth.edu](mailto:ResearchDataHelp@groups.dartmouth.edu)

**Jeremy Mikecz**

Research Data Science Specialist  
[jeremy.m.mikecz@dartmouth.edu](mailto:jeremy.m.mikecz@dartmouth.edu)  
[dartgo.org/jeremyappts](https://dartgo.org/jeremyappts)

**Simon Stone**

Research Data Science Specialist  
[simon.stone@dartmouth.edu](mailto:simon.stone@dartmouth.edu)  
[dartgo.org/meetwithsimon](https://dartgo.org/meetwithsimon)

**Lora Leligdon**

Head of Research Data Services  
[lora.c.leligdon@dartmouth.edu](mailto:lora.c.leligdon@dartmouth.edu)  
[dartgo.org/lora](https://dartgo.org/lora)

# Why talk about Jupyter Notebooks?

- 👉 Jupyter Notebooks are an important tool in computational research and education
- ▶ Many beginner courses use them, but only gloss over how they work
- 🌀 With only a superficial grasp of them, Jupyter Notebooks can get in the way of learning and understanding
- 😈 Some bad practices may be unwittingly adopted because of them



# What you will learn in this workshop

- **What** are Jupyter Notebooks
- **How** do you use Jupyter Notebooks
- **What to look out for** when working with Jupyter Notebooks
- **When to use** Jupyter Notebooks

# What we will work with in this workshop

- Jupyter Notebooks 🎉
- We will use a smattering of **Python** for illustrative code examples
- Materials: [www.dartgo.org/rr-notebooks101](http://www.dartgo.org/rr-notebooks101)







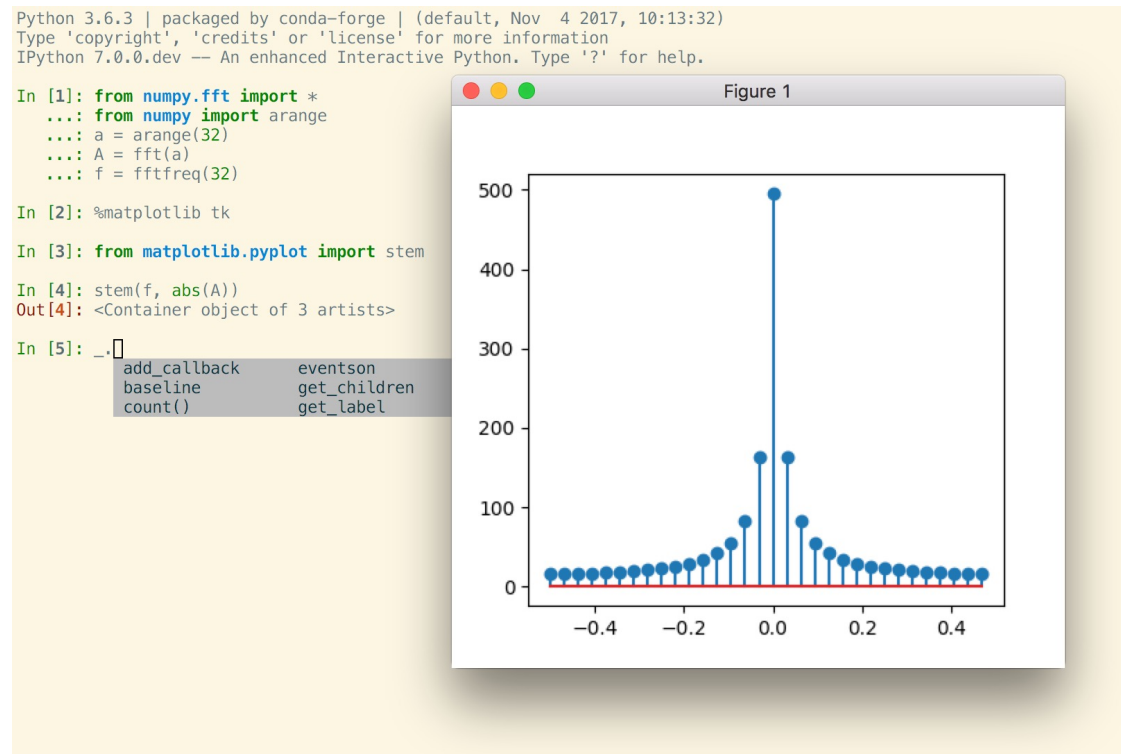
# Let's get started...

# A brief history of Jupyter

In 2001, Fernando Pérez wanted a better Python shell

He created IPython:

- Syntax highlighting
- Autocompletion
- Interactive visualizations

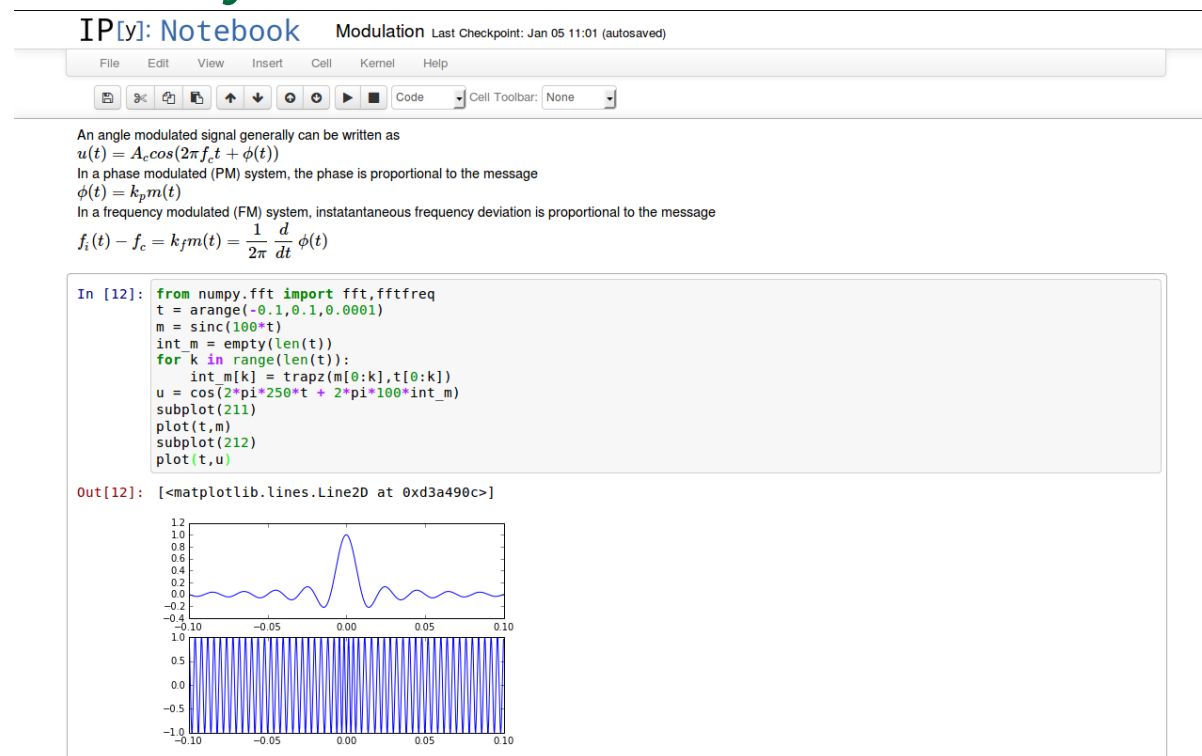


# A brief history of Jupyter

In 2001, Fernando Pérez wanted a better Python shell

He created IPython:

- Syntax highlighting
- Autocompletion
- Interactive visualizations
- A rich editor view mixing text, code, and visualizations






# A brief history of Jupyter

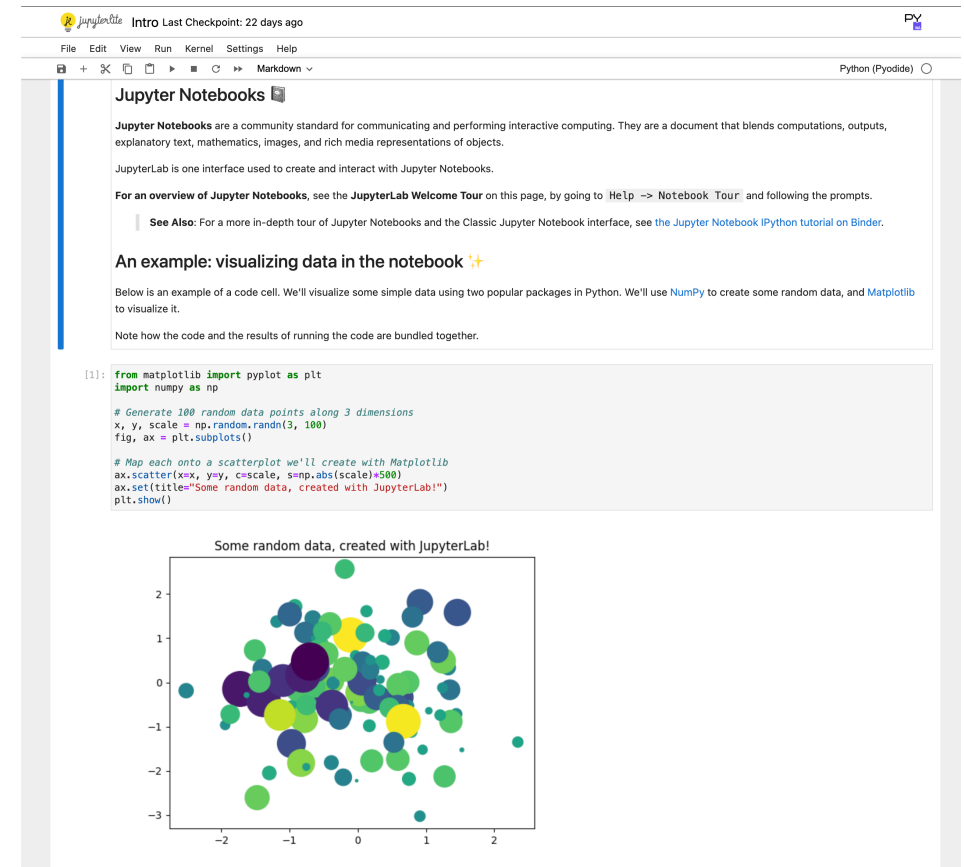
- In 2014, Pérez and his team realized that their editor functionality in IPython was actually independent of the programming language used
- Project Jupyter was spun off from IPython
- Jupyter: Julia, Python, R  
(the originally supported languages)
- Today, Jupyter Notebooks are used by all major cloud providers  
(Amazon Sagemaker, Google Colaboratory, Microsoft Azure Notebook)
- Jupyter Notebooks are virtually *everywhere* in research and education
- *The Atlantic*: [“The Scientific Paper is Obsolete”](#)



# What is a Jupyter Notebook?

Source: [https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what\\_is\\_jupyter.html](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html)

-  Notebooks are files produced by the Jupyter Notebook application, which contain both computer code (e.g., python) and rich text elements (paragraph, equations, figures, links, etc...).
-  The Jupyter Notebook App is a server-client application that allows editing and running notebook files via a web browser.
-  The Jupyter Notebook App can be executed on a local desktop requiring no internet access or can be installed on a remote server and accessed through the internet.



The screenshot shows a Jupyter Notebook interface. At the top, there's a header with the Jupyter logo and 'Intro Last Checkpoint: 22 days ago'. Below this is a menu bar with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Settings', and 'Help'. A toolbar with various icons is visible. The main content area has a text cell titled 'Jupyter Notebooks' which explains that Jupyter Notebooks are a community standard for communicating and performing interactive computing. It mentions that JupyterLab is one interface used to create and interact with Jupyter Notebooks. It also provides links for an overview and a more in-depth tour. Below the text cell is a code cell with the following Python code:

```
[1]: from matplotlib import pyplot as plt
import numpy as np

# Generate 100 random data points along 3 dimensions
x, y, scale = np.random.randn(3, 100)
fig, ax = plt.subplots()

# Map each onto a scatterplot we'll create with Matplotlib
ax.scatter(x=x, y=y, c=scale, s=np.abs(scale)*500)
ax.set(title="Some random data, created with JupyterLab!")
plt.show()
```

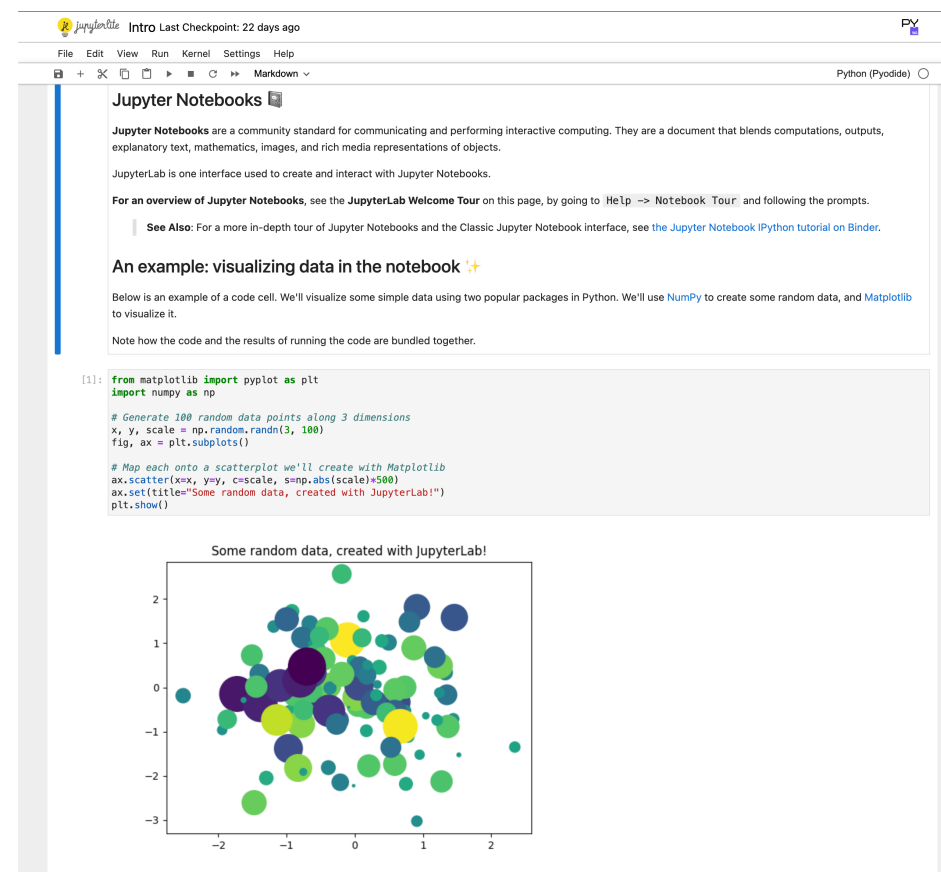
The code cell is followed by a scatter plot titled 'Some random data, created with JupyterLab!'. The plot shows 100 data points in a 2D space, with the x-axis ranging from -2 to 2 and the y-axis ranging from -3 to 2. The points are colored and sized based on a third variable, 'scale', which ranges from -1 to 1. The plot is a scatter plot with a white background and a black border.

Source: <https://jupyter.org/try>

# What is a Jupyter Notebook?

Source: [https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what\\_is\\_jupyter.html](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html)

- The code cells are executed by the *kernel*, a computational engine associated with the notebook
- There are many different kernels, each one offering a different programming language:
  - IPython, IRKernel, IJulia, Xeus (C++), many more
- Think of a kernel as a service that your notebook uses to run the code
- The kernel can run on your local machine or remotely (e.g., in the cloud or on an HPC cluster)



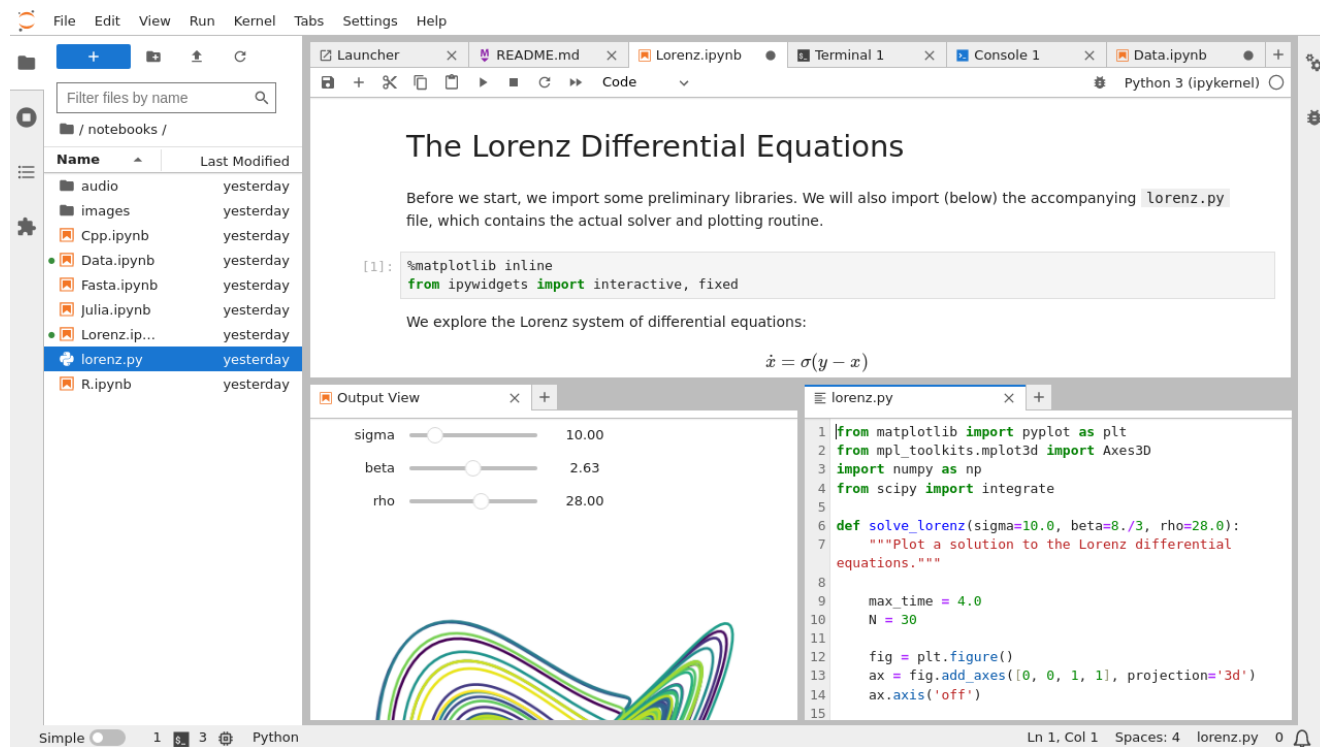
Source: <https://jupyter.org/try>

# How can you run Jupyter Notebooks?

Many ways lead to Jupyter, e.g.:

The official website ([www.jupyter.org/install](http://www.jupyter.org/install))

- The original Jupyter Notebook App
- The face-lifted, latest version JupyterLab



The screenshot shows the JupyterLab interface with the following components:

- File Browser (Left):** A sidebar showing a file tree with folders like 'audio', 'images', and files like 'Data.ipynb', 'Fasta.ipynb', 'Julia.ipynb', 'Lorenz.ip...', 'lorenz.py', and 'R.ipynb'.
- Code Editor (Center):** A window titled 'Lorenz.ipynb' showing the title 'The Lorenz Differential Equations' and introductory text. Below the text is a code cell with the following code:
 

```
[1]: %matplotlib inline
      from ipywidgets import interactive, fixed
```
- Output View (Bottom):** A window titled 'Output View' showing a 3D plot of the Lorenz attractor. The plot has three sliders for parameters: 'sigma' (10.00), 'beta' (2.63), and 'rho' (28.00).
- Code Editor (Bottom Right):** A window titled 'lorenz.py' showing the Python code for solving the Lorenz system:
 

```
1 |from matplotlib import pyplot as plt
2 |from mpl_toolkits.mplot3d import Axes3D
3 |import numpy as np
4 |from scipy import integrate
5
6 |def solve_lorenz(sigma=10.0, beta=8./3, rho=28.0):
7 |    """Plot a solution to the Lorenz differential
8 |    equations."""
9 |
10 |    max_time = 4.0
11 |    N = 30
12 |
13 |    fig = plt.figure()
14 |    ax = fig.add_axes([0, 0, 1, 1], projection='3d')
15 |    ax.axis('off')
```

Source: <https://jupyterlab.readthedocs.io/en/latest/>

# How can you run Jupyter Notebooks?

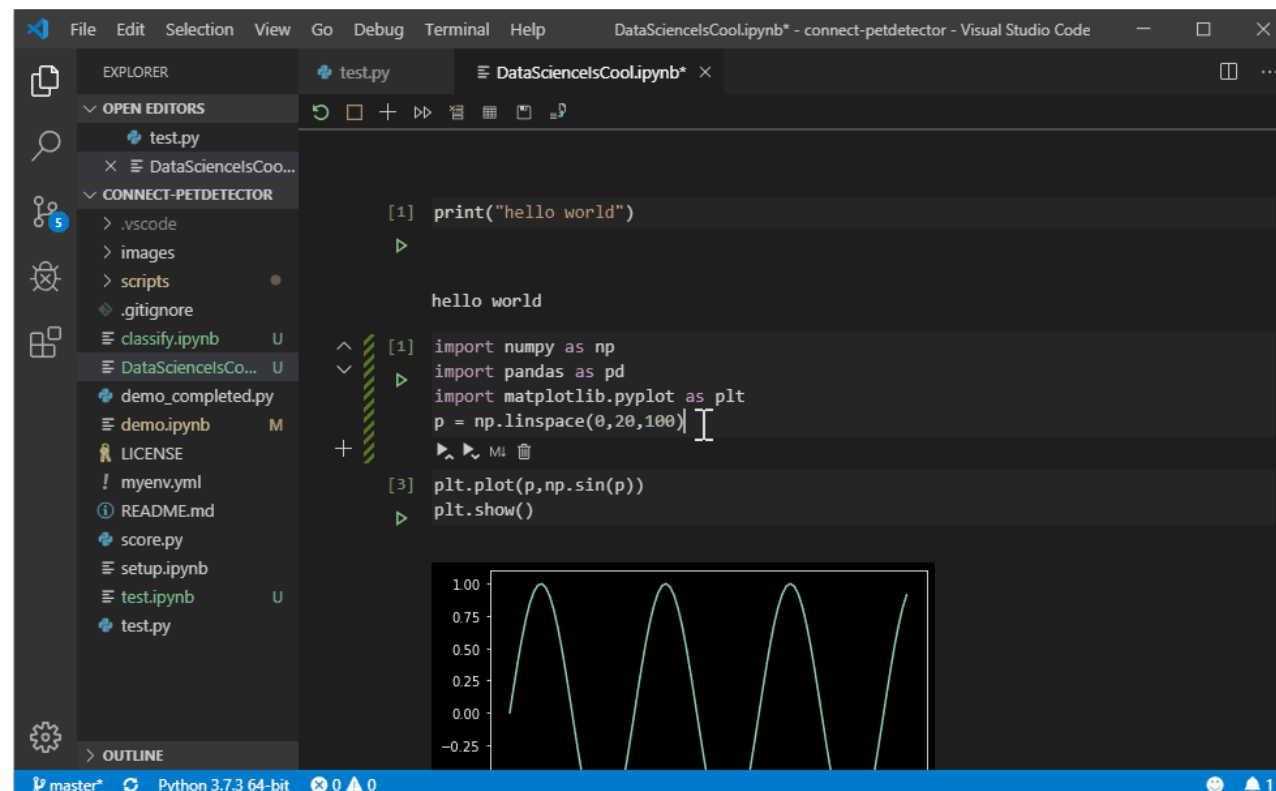
Many ways lead to Jupyter, e.g.:

**The official website** ([www.jupyter.org/install](http://www.jupyter.org/install))

- The original Jupyter Notebook App
- The face-lifted, latest version JupyterLab

**Using a Code Editor/IDE**

- E.g., [PyCharm](#) or [Visual Studio Code](#)



Source: <https://towardsdatascience.com/jupyter-notebook-in-visual-studio-code-3fc21a36fe43>



# How can you run Jupyter Notebooks?

Many ways lead to Jupyter, e.g.:

**The official website** ([www.jupyter.org/install](http://www.jupyter.org/install))

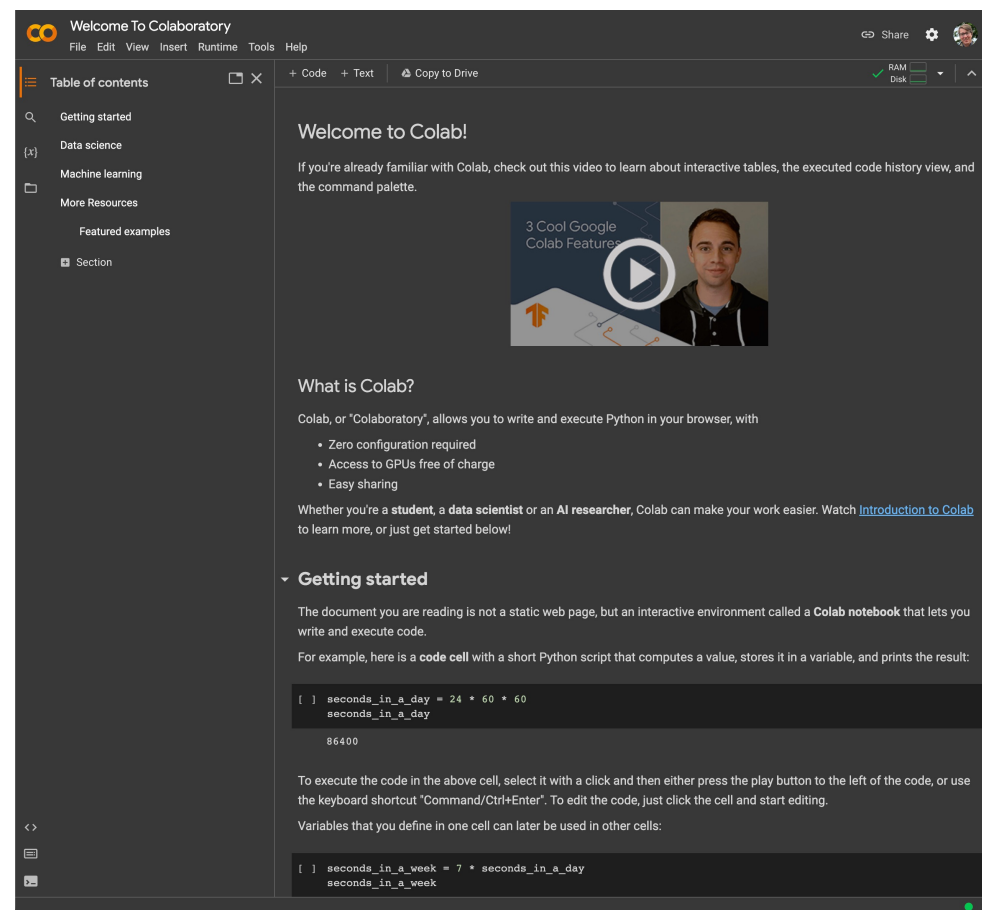
- The original Jupyter Notebook App
- The face-lifted, latest version JupyterLab

## Using a Code Editor/IDE

- E.g., [PyCharm](#) or [Visual Studio Code](#)

## Cloud service

- Google Colab, Amazon Sagemaker
- JupyterHub (e.g., [jhub.dartmouth.edu](http://jhub.dartmouth.edu))

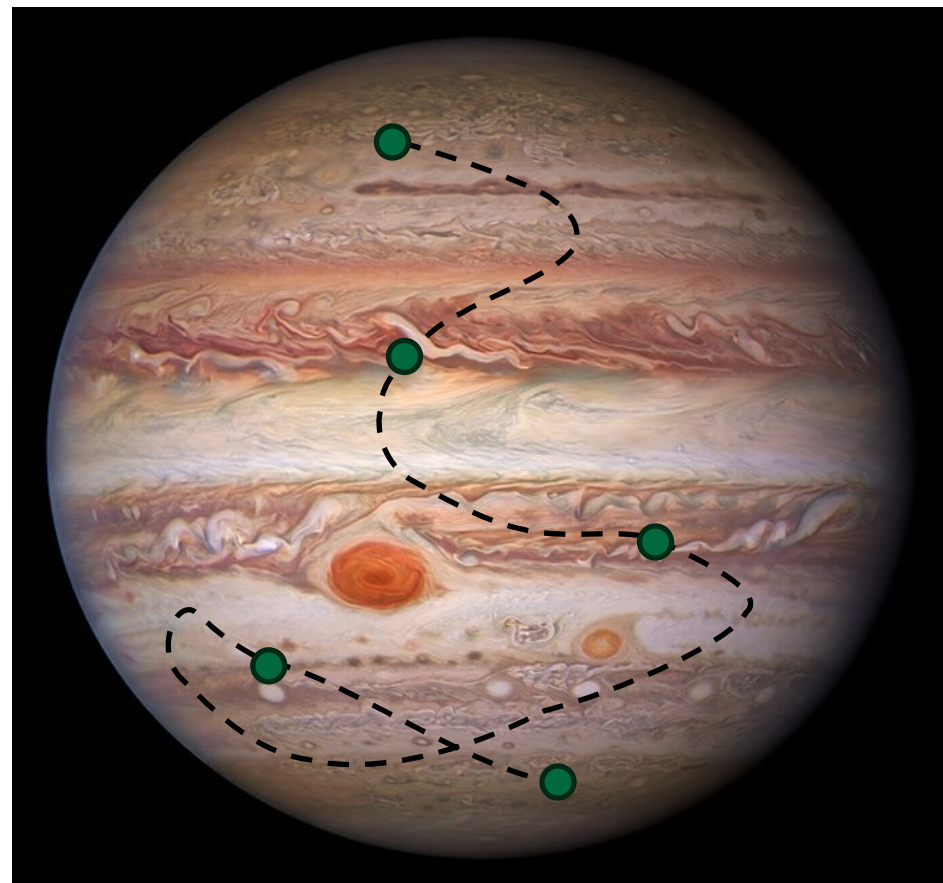


Source: <https://colab.research.google.com/>

# A Walk around Jupyter: Hands-on

## Itinerary

- General layout and interface elements
- Markdown cells
- Code cells
- Producing figures
- Magic commands

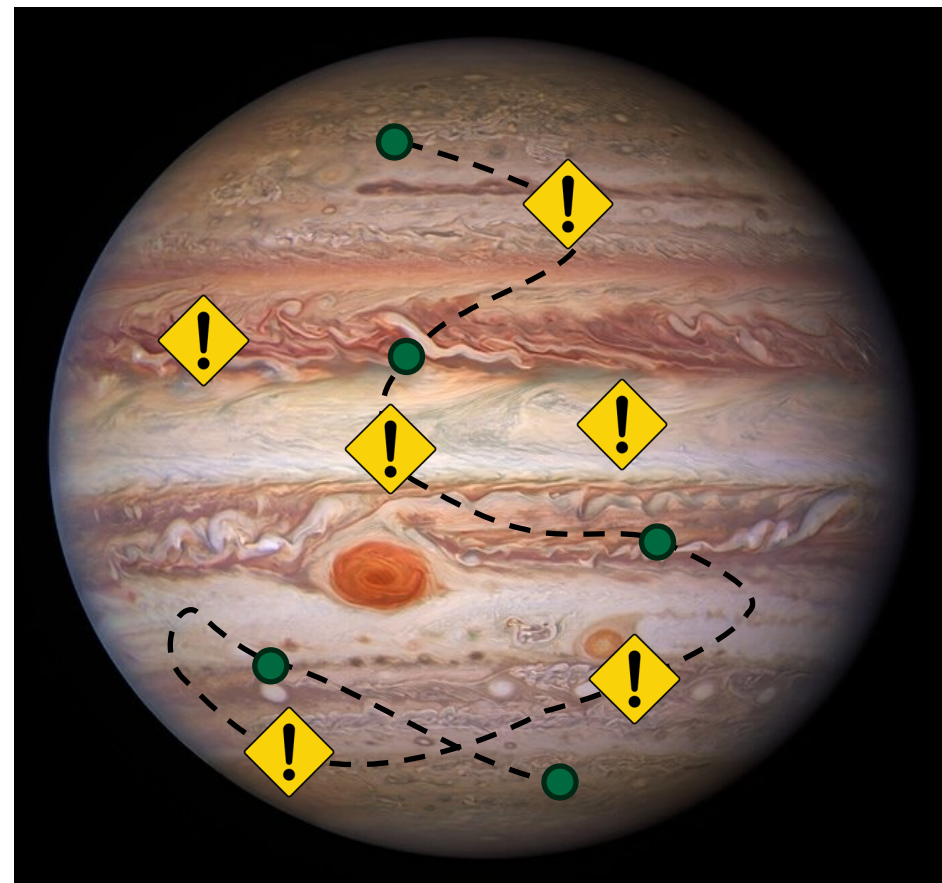


NASA/ESA/NOIRLab/NSF/AURA/M.H. Wong and I. de Pater (UC Berkeley)  
et al. Acknowledgments: M. Zamani, [CC BY 4.0](#), via Wikimedia Commons

# What to look out for







- Running code out of order
- “God notebooks”
- Discouraging modularity
- Difficulty to test and debug code
- Version control can be challenging
- Reproducibility may be an illusion:
  - Only 3 % of notebooks from scientific publications truly reproducible

Samuel, S., & Mietchen, D. (2023). Computational reproducibility of Jupyter notebooks from biomedical publications. *arXiv preprint arXiv:2308.07333*.



NASA/ESA/NOIRLab/NSF/AURA/M.H. Wong and I. de Pater (UC Berkeley)  
et al. Acknowledgments: M. Zamani, CC BY 4.0, via Wikimedia Commons

# When to use Jupyter Notebooks

-  Drafting, rapid prototyping
-  Creating visualizations
-  Reporting
-  Education and Teaching
-  Stand-alone tasks that are not part of a bigger pipeline
-  When reusability of code is not a concern

# Next steps

- Widgets:
  - Graphical user interface controls (sliders, checkboxes, text inputs, ...)
- Nbconvert
  - Convert your notebook into a static format (PDF, HTML, LaTeX, Markdown, ...)
- Voilà
  - Turn your notebook into a dashboard or web app



# Questions





# Thank you.

