



April 25th, 2018

Matthias Bussonnier

bussonniermatthias@gmail.com

GitHub: @carreau

Twitter: @mbussonn



About Me

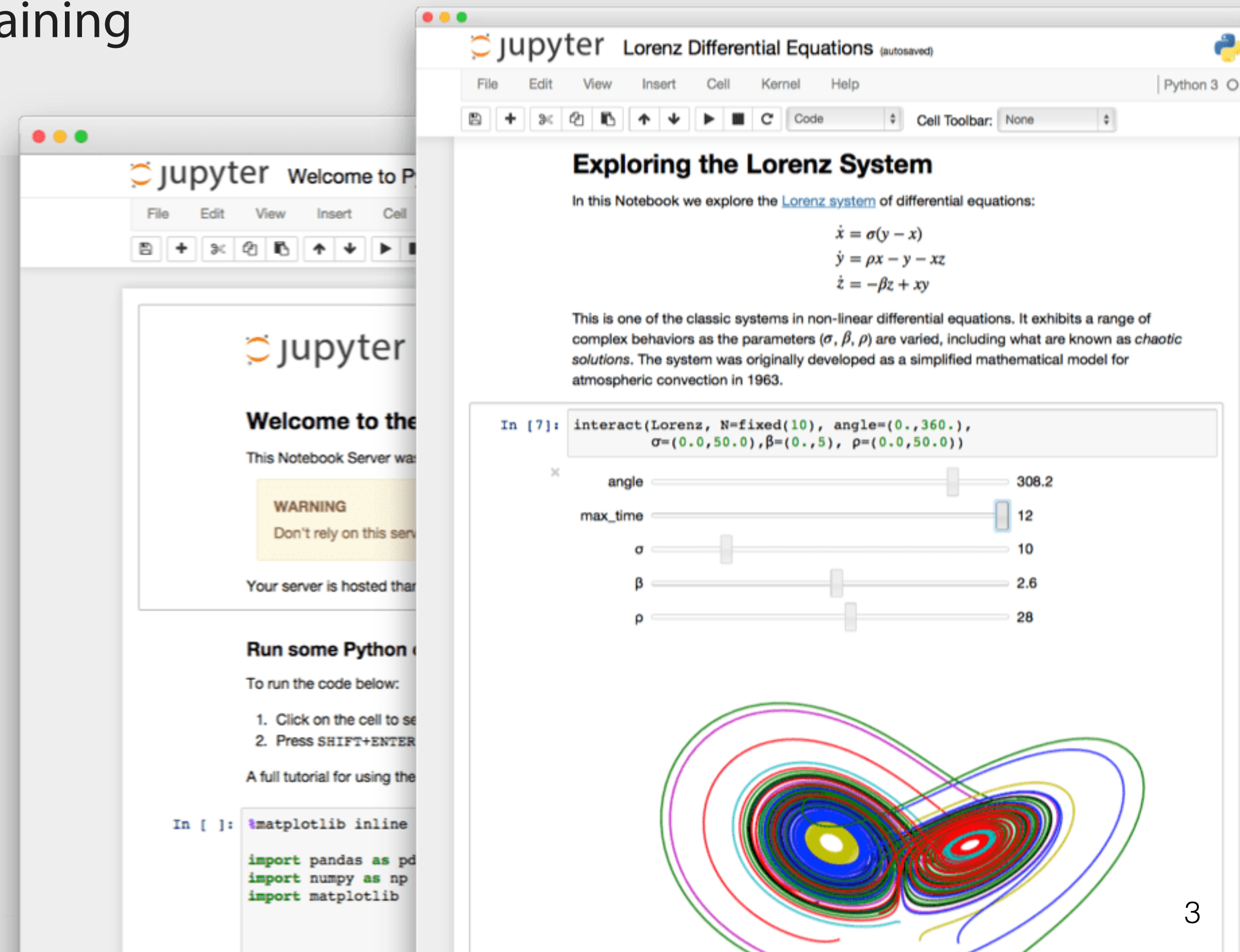
Matthias Bussonnier

- A Physicist/Bio-Physicist
- Core developer of IPython/Jupyter since 2012
 - Co-founder, and Steering Council member
- Post doctoral Scholar on Jupyter at BIDS



What is Jupyter

- Mainly Known for **The Notebook**
 - Web server, a web app, load .ipynb (json), containing code, narrative, math and results.
 - Attached to a **Kernel** doing computation.
- Results can be:
 - Static (Image)
 - Interactive (client-side scoll/pan/brush)
 - Dynamic (Call back into Kernel)



Open Organisation

- Organisation with Open Governance (<https://GitHub.com/jupyter/governance>)
- Funded by Grants and Donations, and Collaborations



Protocols and Formats

- Jupyter is also a set of **Protocols and Formats** that reduce the **N-frontends × M-backends** problem to a **M-Frontends + N-backends**,
 - Open, Free and Simple.
 - JSON (almost) everywhere
 - Notebook document format,
 - Wire protocol
 - Thought for Science and **Interactive** use case.
 - Results embedded in documents no "Copy past" mistake.
 - Scale from Education to HPC jobs.

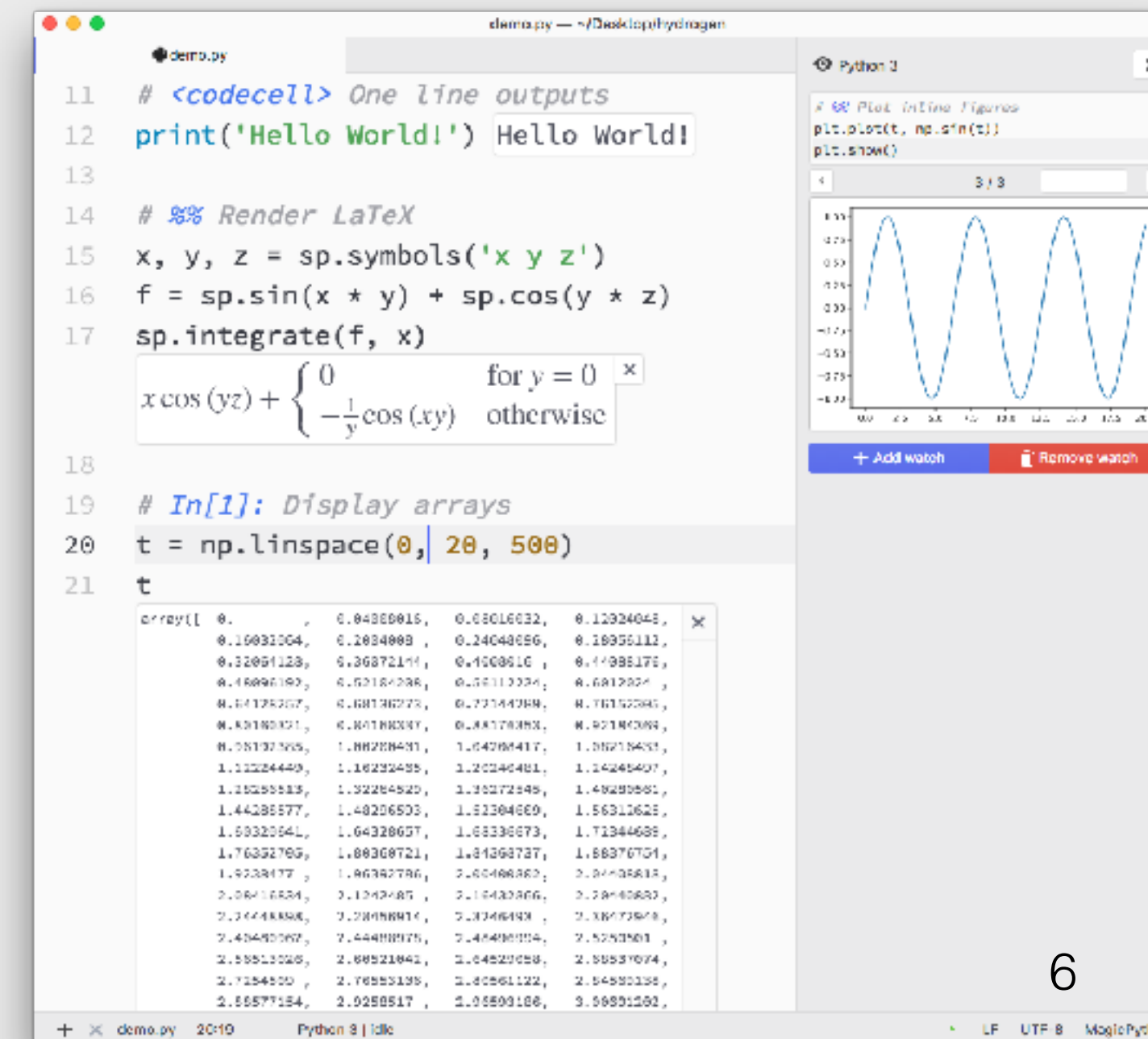
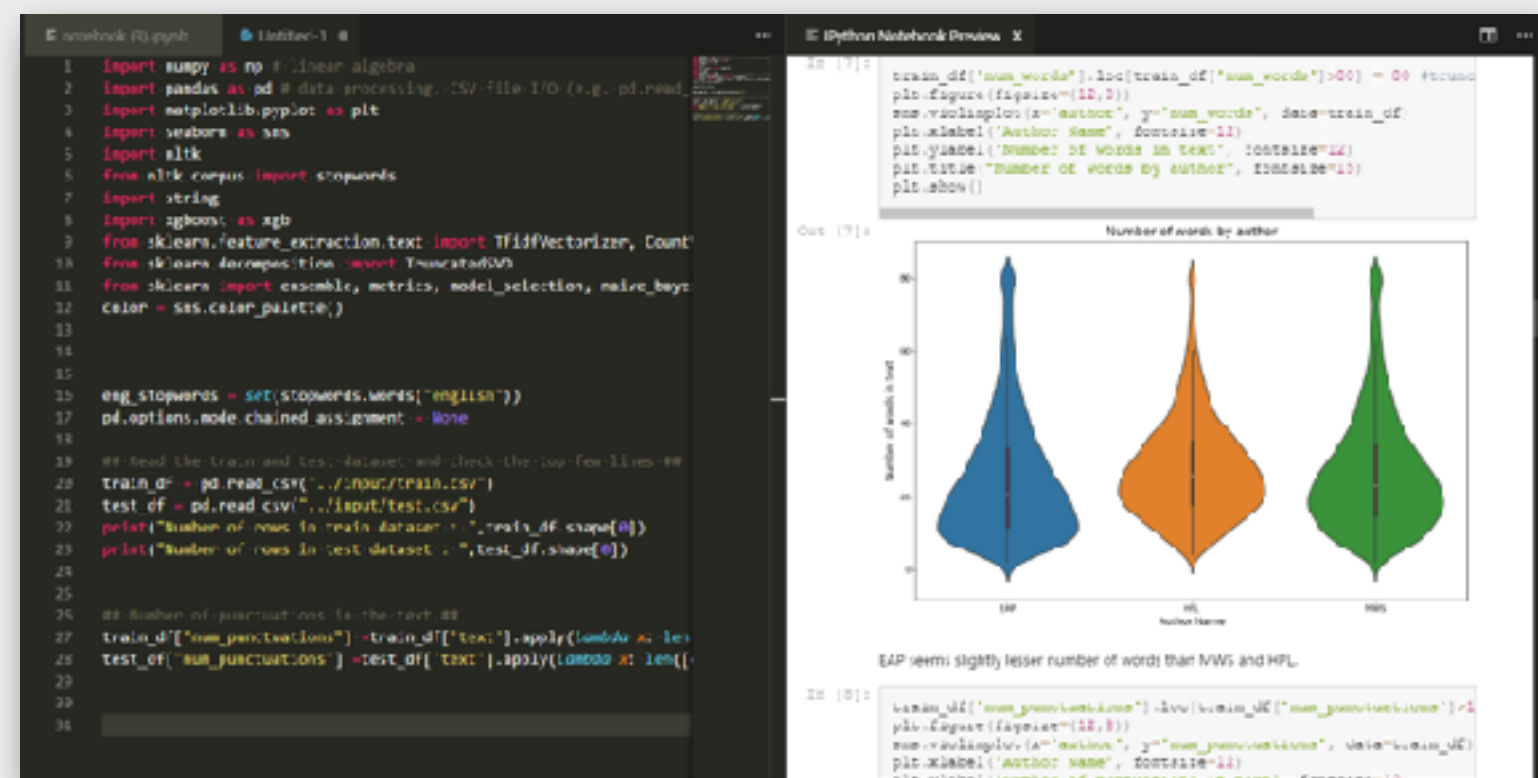


Ecosystem

Frontends: Notebook, JupyterLab, CLI, *Vim, Emacs, Visual Studio Code, Atom, Nteract, Juno...*

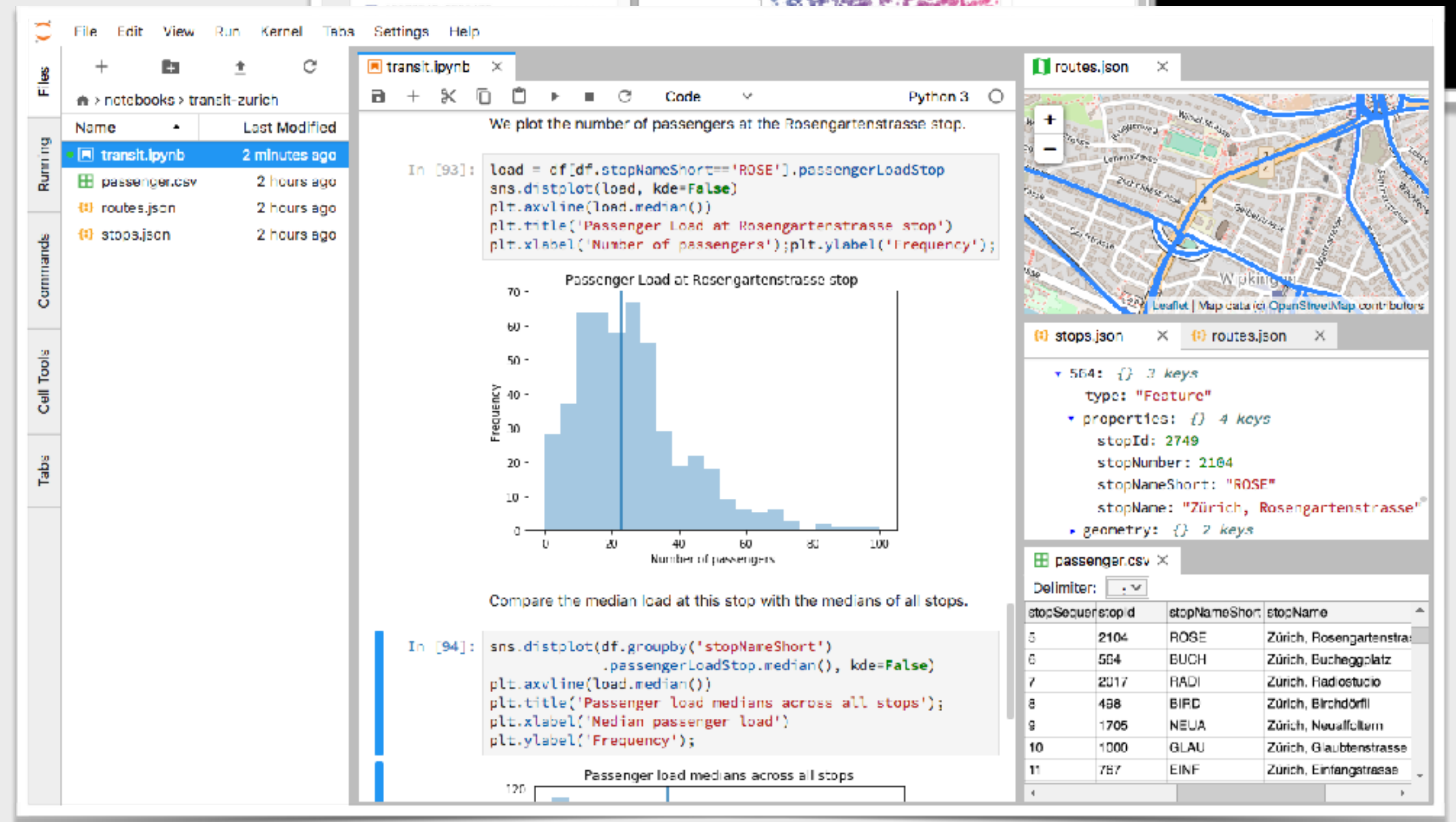
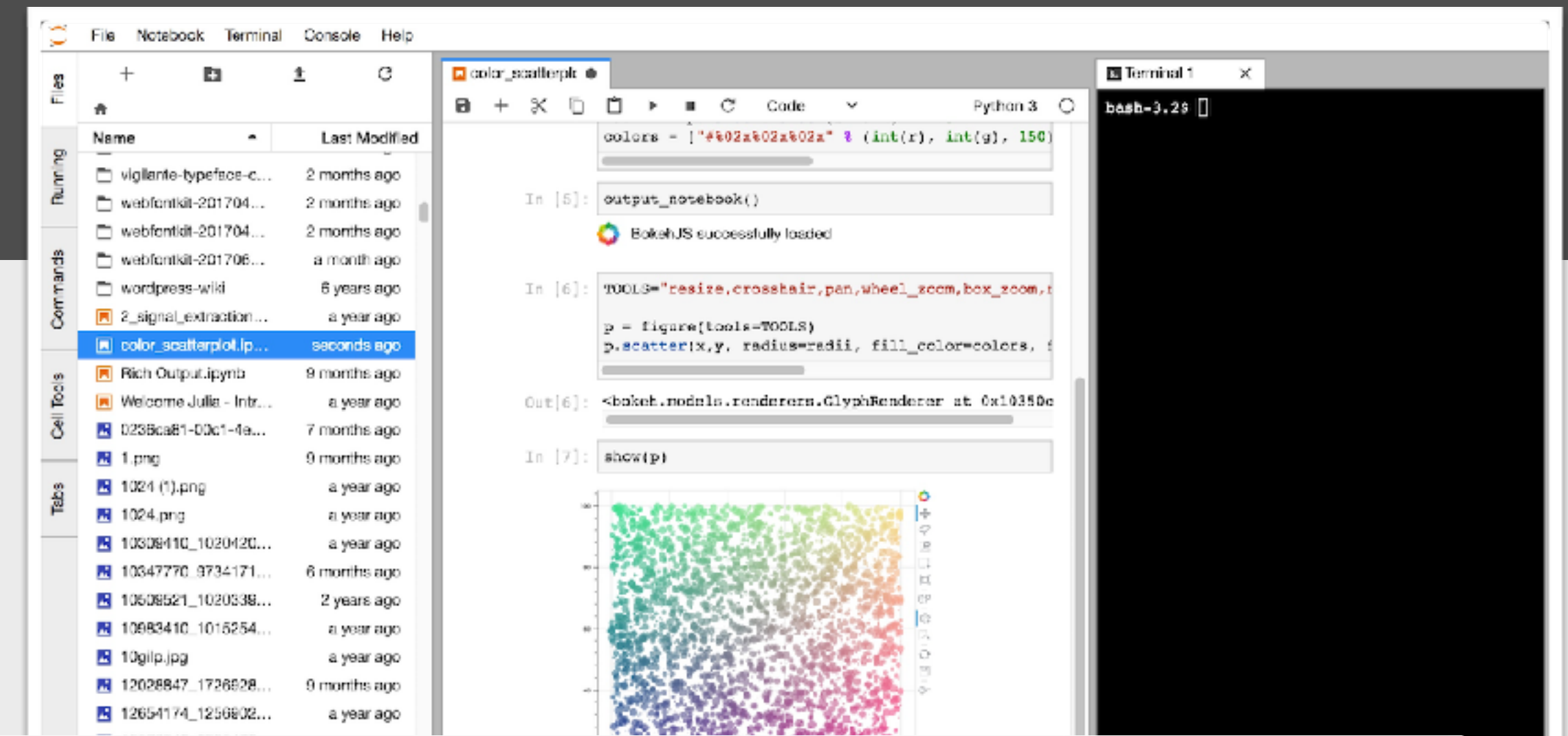
Kernels: Python, Julia, R, Haskell, Perl, Fortran, Ruby, Javascript, C/C++, Go, Scala, Elixir... 60+

Building Blocks: Nbformat, JupyterHub, Kernel Gateway...



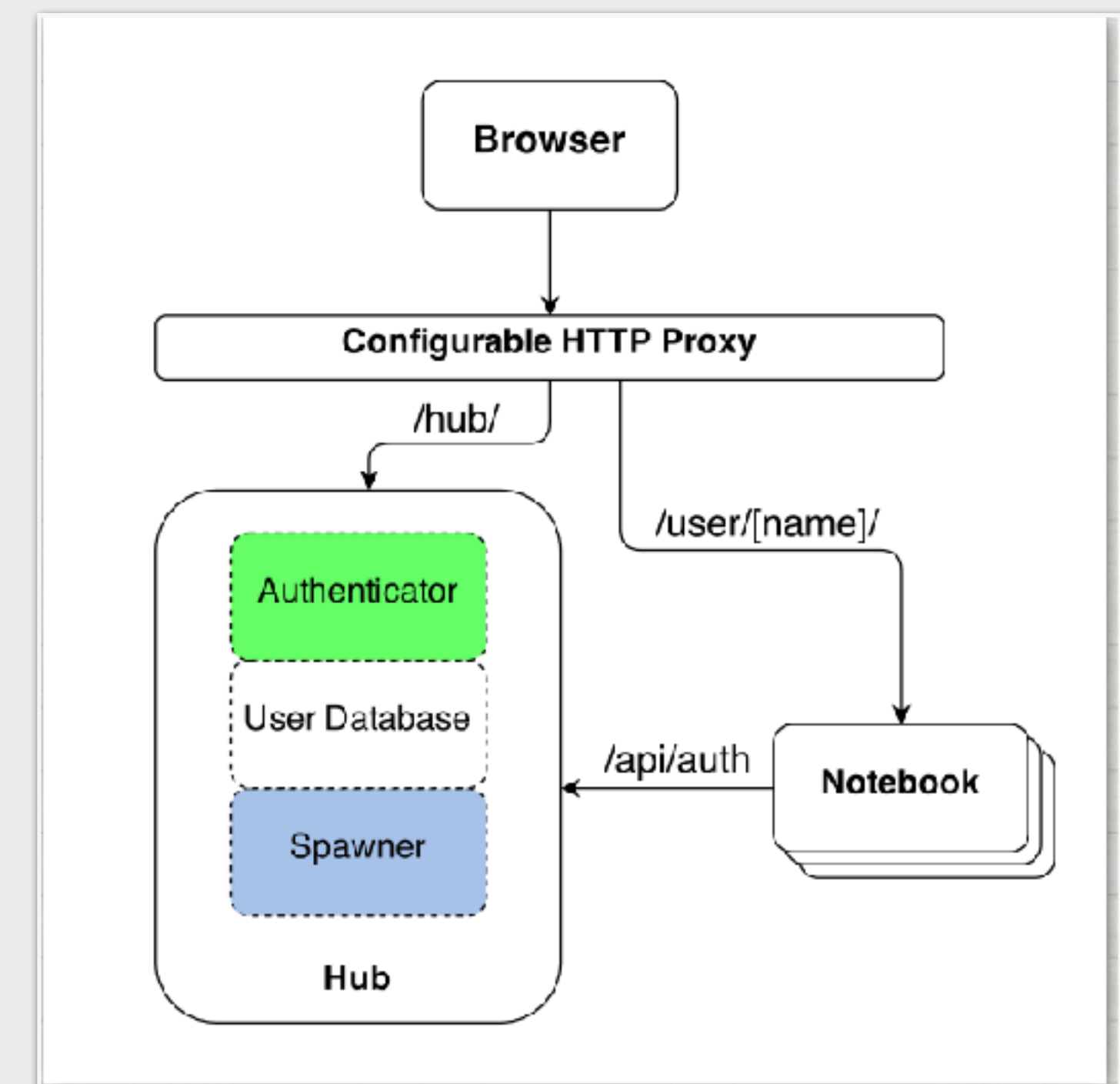
JupyterLab

- Extends the notebook interface with text editor, shell, ...etc
- is it and IDE ?
- If by I you mean Interactive, then yes



JupyterHub

- Multi-users Jupyter deployment
 - Not (Yet) Realtime collaboration
- Each user can get their own process/version(s)/configuration(s)
 - Hooks into any Auth
 - Only requires a browser
- Not limited to running Jupyter (e.g. work with RStudio, OpenRefine...)



Follow Along !



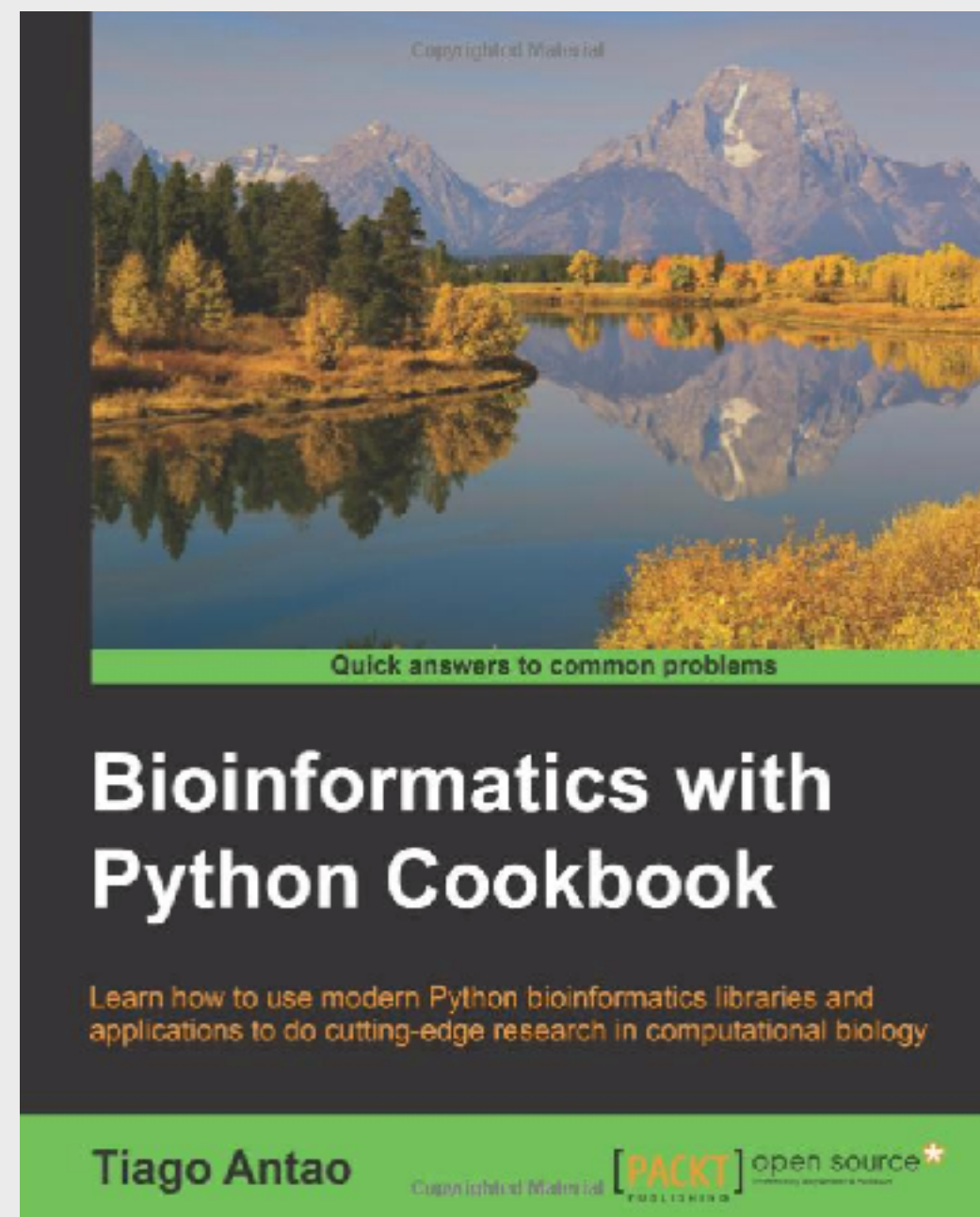
MyBinder.org

- Hosted, Ephemeral, On demand notebook in Docker Container.
- Can take time to start

<https://github.com/Carreau/JGI-demo>



Jupyter In Bioinformatic



- See [A gallery of interesting notebooks](#)
 - [An Introduction to Applied Bioinformatics](#)
 - [Bioinformatics with Python cookbook's notebooks](#)
 - [An open RNA-Seq data analysis pipeline tutorial with an example of reprocessing data from a recent Zika virus study](#)
 - [Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss](#)
- ...



The Shape of Things to come



The Shape of Things to come

Classic Notebook -> JupyterLab transition

- Stabilisation
- Transfer of extensions
- Collaboration:
 - Google retired Real-time API
 - Who “executes” problem
- Long Running Jobs

JupyterHub

- Horizontal (and Vertical) Scaling
- Audits APIs (Hippa Compliance ?)
- “Federation” (binder) / Intercommunication





O'REILLY*

jupytercon

Brought to you by NumFOCUS Foundation and O'Reilly Media Inc.

August 21–24, 2018
New York, NY

jupytercon.com

Thanks,
Questions ?





HPC Misconceptions

We need to run JupyterHub on the cluster: No

- Hub, Server, Kernels, (and Workers):
 - Do not have to be on the same machine
 - Do not have to use the same environment

A Kernel is a (single) language: No

- A Kernel is a preconfigured computation environment. It can be:
 - A queue, a hardware resource (GPU, SSD...), A location (like a beam-line)
- Example of Python, Cython, Julia, R, Fortran, Rust, C calling each other in same notebook

Every User have the same environment: No/No

- Kernels and notebook server can be configured independently
- Subset of users could use different server versions w/ different extensions.

JupyterHub is Limited to Jupyter: No

JupyterHub Can run RStudio, Open Refine.



Danger !

Despite Notebook being great, some limitations:

Most if not all **document state** is **in your browser** !

- Watch out for flaky network connections !
- Do Not close your Laptop Lid*/Tab* !

Workaround:

- Wrap computations (especially long), in Futures
- Use Caching.

Interrupting in compiled code is hard.

Large outputs/notebooks can crash the browser



*recent versions have buffering if `ifzlier`