ETH zürich

IT Services

Interactive Scientific Computing with Computational Notebooks

Henry Lütcke
Scientific IT Services, ETH Zurich

ETH Library Coffee Lecture, Nov. 17 2021

CC BY NC SA

Scientific Computing – A Requirement in Most Research Fields

Steven Ge
@StevenXGe

OMG! Deep Learning is hot!
Six out of 10 articles published by Nature Methods this week have "deep" in their titles! Another uses the convolutional network. It is not a special issue.
It could a revolution or hype/herding effect at its peak.
nature.com/nmeth/volumes/...

6:44 PM · Nov 10, 2021 · Twitter Web App

https://twitter.com/StevenXGe/status/1458490843245973504

Traditional Scientific Computing – Editors, Terminals, IDEs

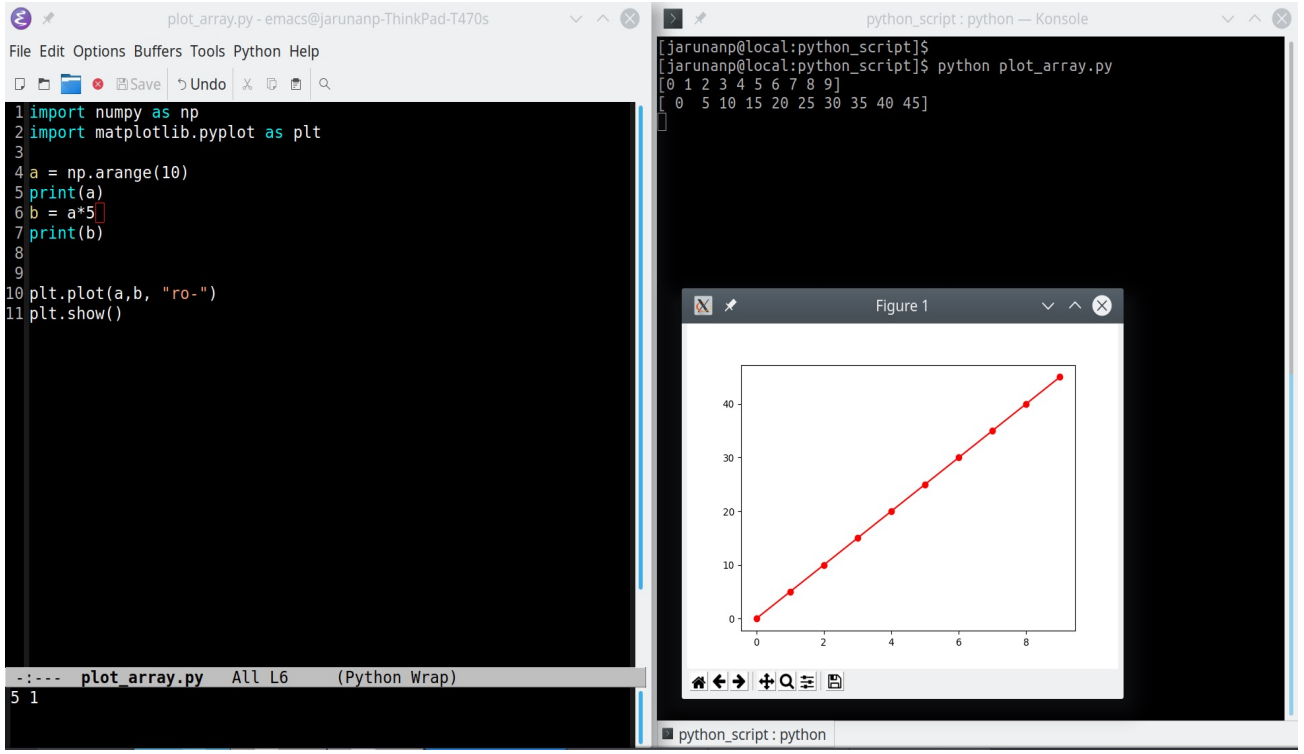
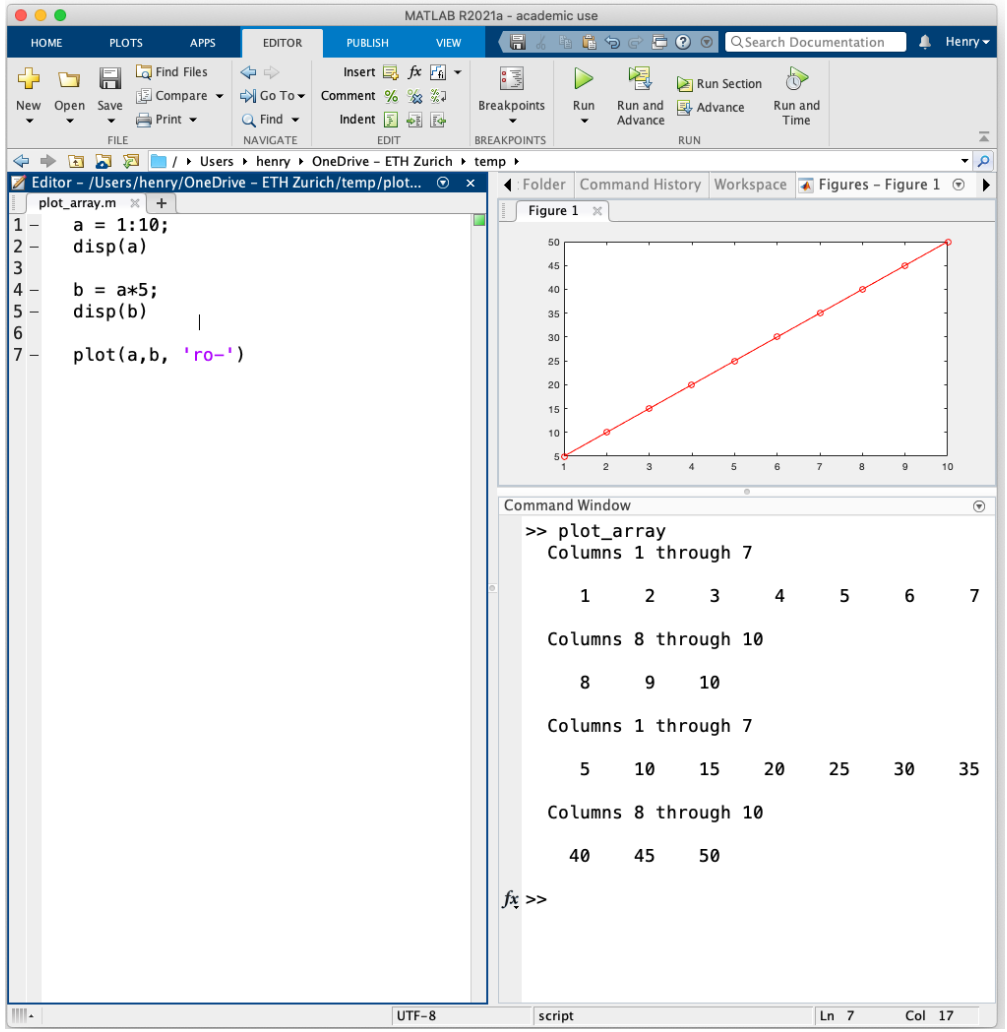


Image credit: Jarunan Panyasantisuk



Computational Notebooks – Keep Together What Belongs Together

jupyter Jupyter_Image_Analysis_Notebook Last Checkpoint: câteva secunde în urmă (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

Image Analysis with Python and Jupyter

This Jupyter notebook demonstrates the analysis of a simple image with Python.

First, make sure you have the image `blobs.tif` in the `data` folder of your notebook dashboard.

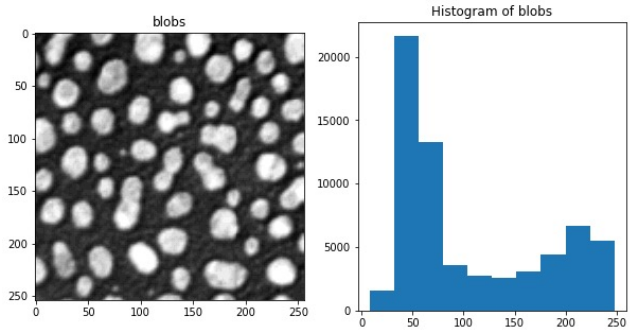
Import required modules

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
from skimage.io import imread
from skimage.morphology import reconstruction, remove_small_objects
from skimage.measure import label, regionprops

%matplotlib inline
```

Read and display the data

In [2]:
```
img = imread('data/blobs.tif')
```

In [3]:
```
fig = plt.figure(figsize=(10,5))
fig.add_subplot(121)
plt.imshow(img, cmap='gray')
plt.title('blobs')
fig.add_subplot(122)
plt.hist(img.ravel(), bins=10)
plt.title('Histogram of blobs')
plt.show()
```



Interactive Computational Notebooks

- Applications that combine documentation, code, input and output generated by the code, e.g. graphs, plots (*Nature 515, 151–152*)
- Useful for exploratory data analysis, sharing and reproducibility

RStudio

- Open source + commercial edition
- Mainly for development in R but other languages supported

jupyter

- Open source
- > 40 languages supported (Python, R, Julia, Matlab, IDL, etc.)

WolframAlpha NOTEBOOK EDITION™

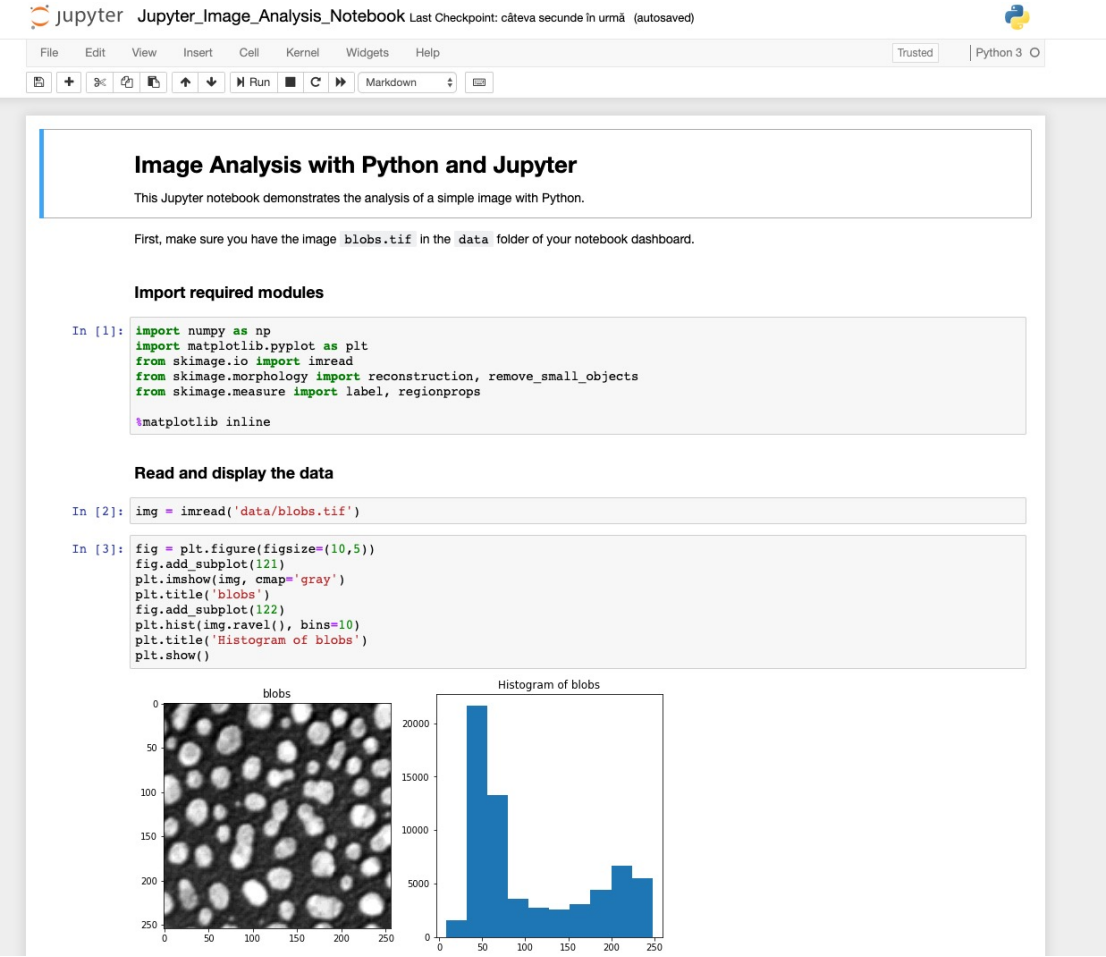
- Commercial
- Used in mathematical fields

MATLAB

- Commercial
- Used in scientific, engineering, mathematical fields

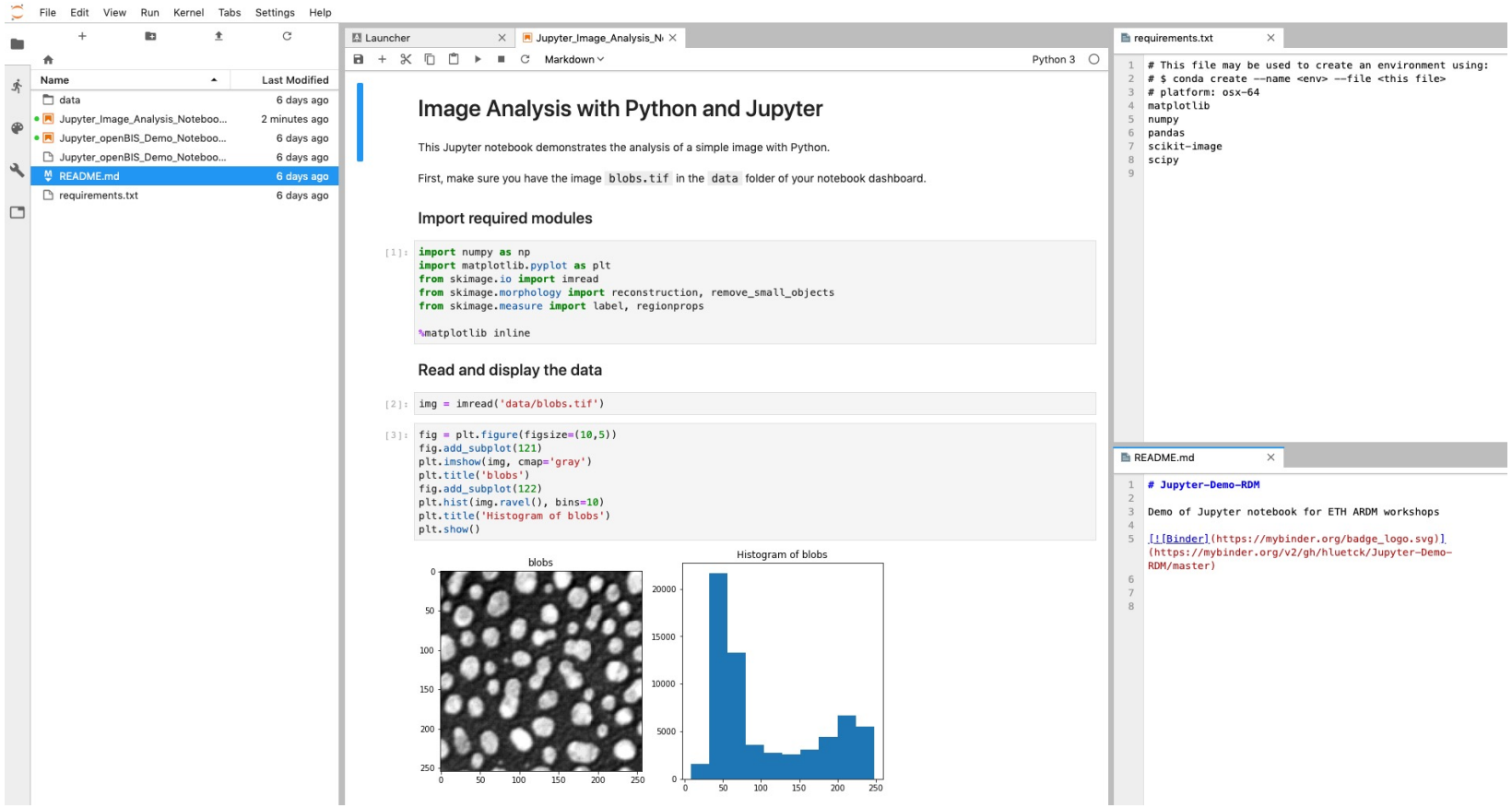
Interactive Computational Notebooks: Jupyter

- **Jupyter notebook:** web-based interactive computational environment



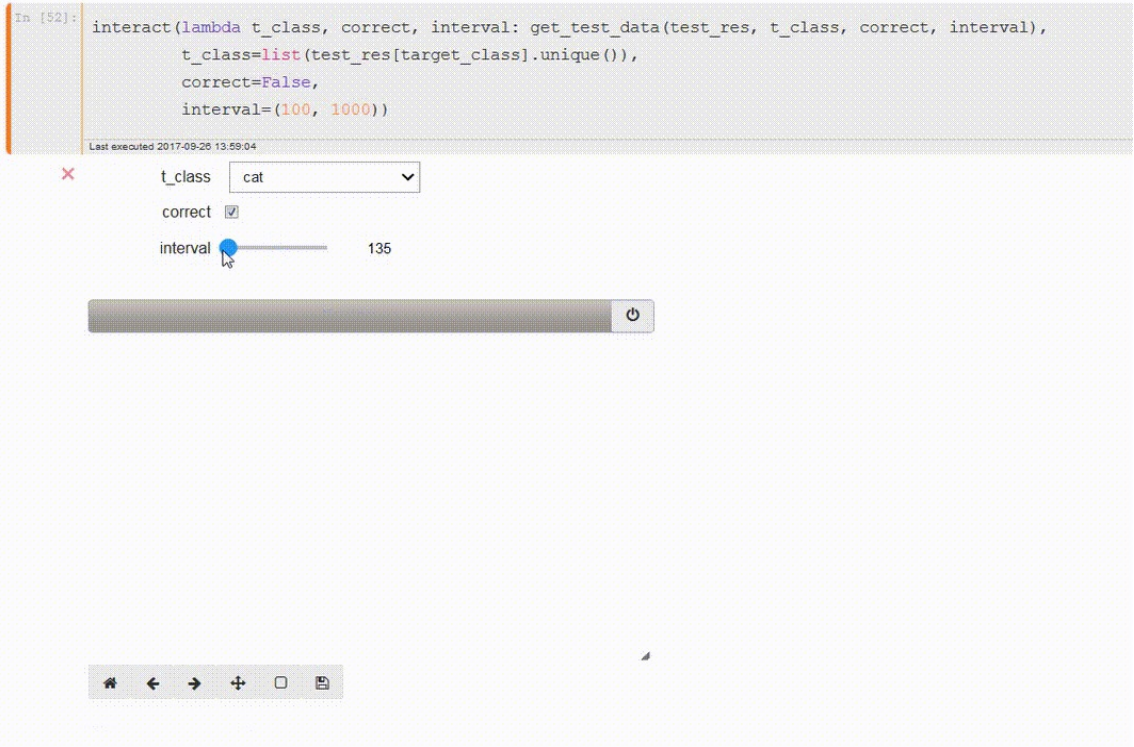
Interactive Computational Notebooks: Jupyter

- **Jupyter notebook:** web-based interactive computational environment
- **JupyterLab:** next-generation for Jupyter notebooks (and more)



Interactive Computational Notebooks: Jupyter

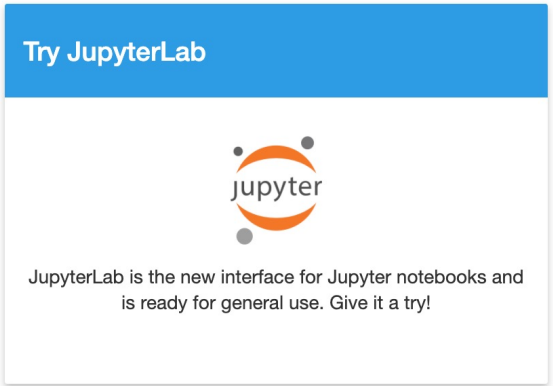
- **Jupyter notebook:** web-based interactive computational environment
- **JupyterLab:** next-generation for Jupyter notebooks (and more)
- Dozens of programming languages supported (core: Julia, Python, R)
- Extensions to build user interfaces (sliders, buttons etc.)



https://towardsdatascience.com/interactive-visualizations-in-jupyter-notebook-3be02ab2b8cd

Interactive Computational Notebooks: Jupyter

- **Jupyter notebook:** web-based interactive computational environment
- **JupyterLab:** next-generation for Jupyter notebooks (and more)
- Dozens of programming languages supported (core: Julia, Python, R)
- Extensions to build user interfaces (sliders, buttons etc.)
- Notebook export in various formats (HTML, PDF, Python …)
- Sharing of interactive notebooks on reproducibility platforms (e.g https://mybinder.org/)
- Integration with ETH scientific computing infrastructure (see https://gitlab.ethz.ch/sfux/Jupyter-on-Euler-or-Leonhard-Open)
- To get started: https://jupyter.org/try



Interactive Computational Notebooks: Jupyter

Gravitational wave physics

gwastro / o2-bbh-pe

Watch 8 | Star 4 | Fork 1

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Security | Insights

Branch: master | o2-bbh-pe / data_release_o2_bbh_pe.ipynb | Find file | Copy path

soumide1102 Update contour plots adding boundary bias code, add skymap notebook f00120b on 26 Apr

1 contributor

11.6 MB | Download | History

Posterior samples of the parameters of binary black holes from Advanced LIGO, Virgo's second observing run

Soumi De[1], Christopher M. Biwer[2], Collin D. Capano[3,4], Alexander H. Nitz[3,4], Duncan A. Brown[1]

[1]Department of Physics, Syracuse University, Syracuse, NY 13244, USA

[2]Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[3]Albert-Einstein-Institut, Max-Planck-Institut for Gravitationsphysik, D-30167 Hannover, Germany

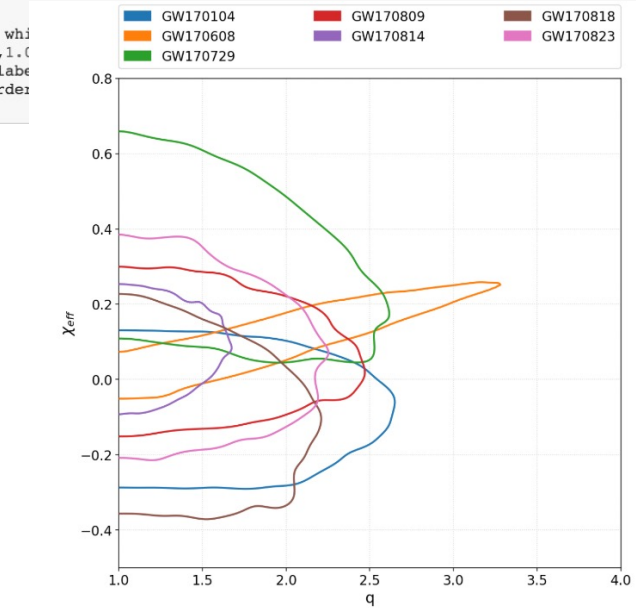
[4]Leibniz Universitat Hannover, D-30167, Hannover, Germany

License

This work is licensed under a https://creativecommons.org/licenses/by/4.0/deed.ast.

To plot Fig. 2 of the paper : mass ratio---effective spin ($q - \chi_{\rm eff}$) posteriors

In [36]:
```
fig, ax = pyplot.subplots(figsize=(9.5, 9.5))


handles = []
colors = itertools.cycle(["C{}".format(i) for i in range(10)])

ndim = 2
# read samples
params = [None] * ndim
params[0] ="(primary_mass(mass1, mass2))/(secondary_mass(mass1, mass2))"
params[1] = "chi_eff_from_spherical(mass1, mass2, spin1_a, spin1_polar, spin2_a, spin2_polar)"

for filename, label in zip(files, labels):
    with InferenceFile(filename, "r") as fp:
        # Read samples from the inference output file
        samples = fp.read_samples(params)
    color = colors.next()

    # Bounds on the domain for evaulating KDE
    xlow_bc, xhigh_bc = 1.0, None
    ylow_bc, yhigh_bc = -1.0, 1.0

    # Make density plot
    create_contour_plot(params[0], params[1], samples, xlow_bc, xhigh_bc,
                        ylow_bc, yhigh_bc, fig=fig, ax=ax, plot_contours=True,
                        xmax=4.0, ymin=-0.5, ymax=0.8, contour_color=color)

    handles.append(patches.Patch(color=color, label=label))


pyplot.xlabel(r"q", fontsize=16)
pyplot.ylabel(r"$\chi_{eff}$", fontsize=16)
pyplot.xlim(right=4.0)
pyplot.ylim(-0.5, 0.8)
pyplot.tick_params(axis='both', whi
pyplot.legend(bbox_to_anchor=(0,1.0
              handles=handles, labe
              mode="expand", border
fig.show()
```



Computational Notebooks – What Can Go Wrong?

- **Versioning**
  - Version control of even moderately complex NBs is challenging
  - Tracking NB history is harder than for traditional source code
  - Some tools may help (e.g. nbdime)

```
$ diff a.ipynb b.ipynb
76,77d75
<     "plt.rc('axes', grid=False)\n",
<     "plt.rc('axes', facecolor='white')\n",
90c88
<      "image/png": "iVBORw0KGgoAAAANSUhEUgAABLkAAAMQCAYAAADLj7dlAAAABHNCSVQICAgIfAhki
AAAAAlwSFlz\nAAAWJQAAFiUBSVIk8AAAIABJREFUeJzsvXeYZFd57b12h0maPNJII2lGOaCAkEBCFgozIxkBAp
lY\n1waDyDZg8MX+zMU2F4Mx1x8PwWAwxmBjg4yNi2BfQMa20iiAQFkIjXKWRtJIE3tSz3TXuX+8vV2n\nqyucv
N+9z/o9zzynprvq1D6nnqqtqr1prbRNFEQghhBBCCCGEEEII8Zkh1wMghBBCCCGEEEIIISQv\nFLkIIYQQQgghB
BCiPdQ5CKEEEIIIYQQQggh3kORixBCCCGEEEIIIYR4D0UuQgghBBCCCGEEOI9\nFLkIIYQQQgghBBCiPdQ5CK
EEEIIIYQQQggh3kORixBCCCGEEEIIIYR4D0UuQgghBBCCCGEEOI9\nFLkIIYQQQgghBBCiPdQ5CKEEEIIIYQQ
Qggh3kORixBCCCGEEEIIIYR4D0UuQgghBBCCCGEEOI9\nFLkIIYQQQjzEGHOJMaZljPmo67EkZWq8D7EkeByGEE
ELChCIXIYQQQirDGPOmKaFj3BhzkMNx/H/G/H/G\nmG3GmP/pagwFEbkeQJUYY75gjNlijHmD67EQQgghRB8UuQghhB
BSJe+DCDMjAH7L4TjeAmA+gLc5\nHEMRGNcDqJi3AVgI4DddPQgghhJBKMMacCuBMAFsg4sy7jTH
DjobzZwBuBxR/dP\nsvERADcC+LTrgRBCCCFEHxS5CCGEEFIVH4C4uP4SIlQcBOD1LgzYSRVEziqIXR1H0fRf3
T7IRRdFf\nRlH0K1EUXe96LIQQQgjRB0UuQgghhJSOMWYpgP8BoAxkhhhhBBC\
```

Computational Notebooks – What Can Go Wrong?

- **Versioning**
  - Version control of even moderately complex NBs is challenging
  - Tracking NB history is harder than for traditional source code
  - Some tools may help (e.g. *nbdime*)
- **Scalability**
  - Scaling to large datasets is challenging (due to browser limitations)
- **Reproducibility**
  - Interactive working mode can result in hard-to-reproduce notebooks
  - Discipline is needed! Regular pruning & refactoring; *"Restart kernel & Run all"* is your friend
- **Collaboration**
  - Collaborative editing generally not supported
- **Security**
  - Data confidentiality & access controls may be problematic

Chattopadhyay et al. *(2020). What's Wrong with Computational Notebooks?* doi:10.1145/3313831.3376729

Contacts

Henry Lütcke
henry.luetcke@id.ethz.ch

sis.helpdesk@ethz.ch
https://sis.id.ethz.ch/

https://twitter.com/ETH_SIS