



# Jupyter

Open Tools for  
Human-in-the-Loop  
Computing

NCSU  
January 30, 2018

Peter Parente

# About Me



- My name is Peter Parente
- Work at Valassis Digital (previously MaxPoint) ~1.5 yrs
- Worked in IBM Emerging Tech ~10 yrs before that
- Received my doctorate from UNC-CH in comp. sci.
- Long time open source contributor
- Long time IPython user
- Relatively recent Jupyter contributor / maintainer

# Agenda



Share information about Jupyter:

- The Project
- The Technology
- The Community
- And You

# The Project



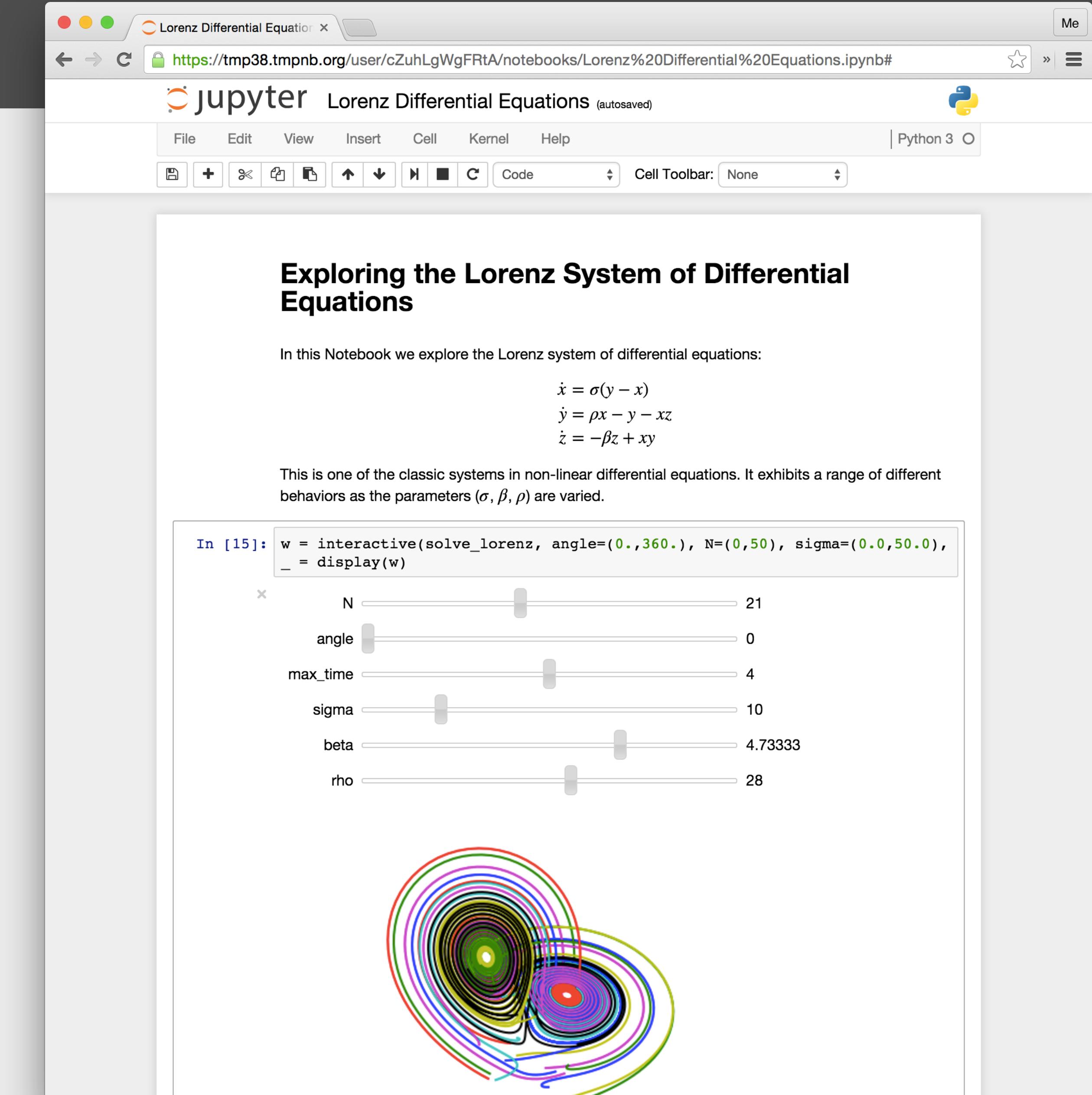
# Project Jupyter



- Born out of the IPython project circa 2011
- Open source software to support interactive computing
- Open formats, protocols, APIs to build an ecosystem
- Open community to promote usage and contribution

# Jupyter Notebooks

Text  
+  
Code  
+  
Visualizations  
+  
Widgets  
+  
Compute  
+  
Data



# Motivations

- Open science
- Reproducible research
- Computational thinking
- Computational narratives

The collage includes the following elements:

- A top right screenshot shows a Jupyter Notebook interface with a video overlay of a woman speaking.
- A middle left screenshot shows a blog post from "Probabilistic Programming and Bayesian Methods for Hackers" with a Jupyter notebook code cell.
- A bottom left screenshot shows a tweet from Mark Cuban (@mcuban) about the new NBA.
- A bottom right screenshot shows a Jupyter Notebook interface titled "Regex Golf with Arbitrary Lists".
- A bottom center screenshot shows a tweet from Project Jupyter (@ProjectJupyter) about LIGO data releases.

# The Technology



# Building Blocks



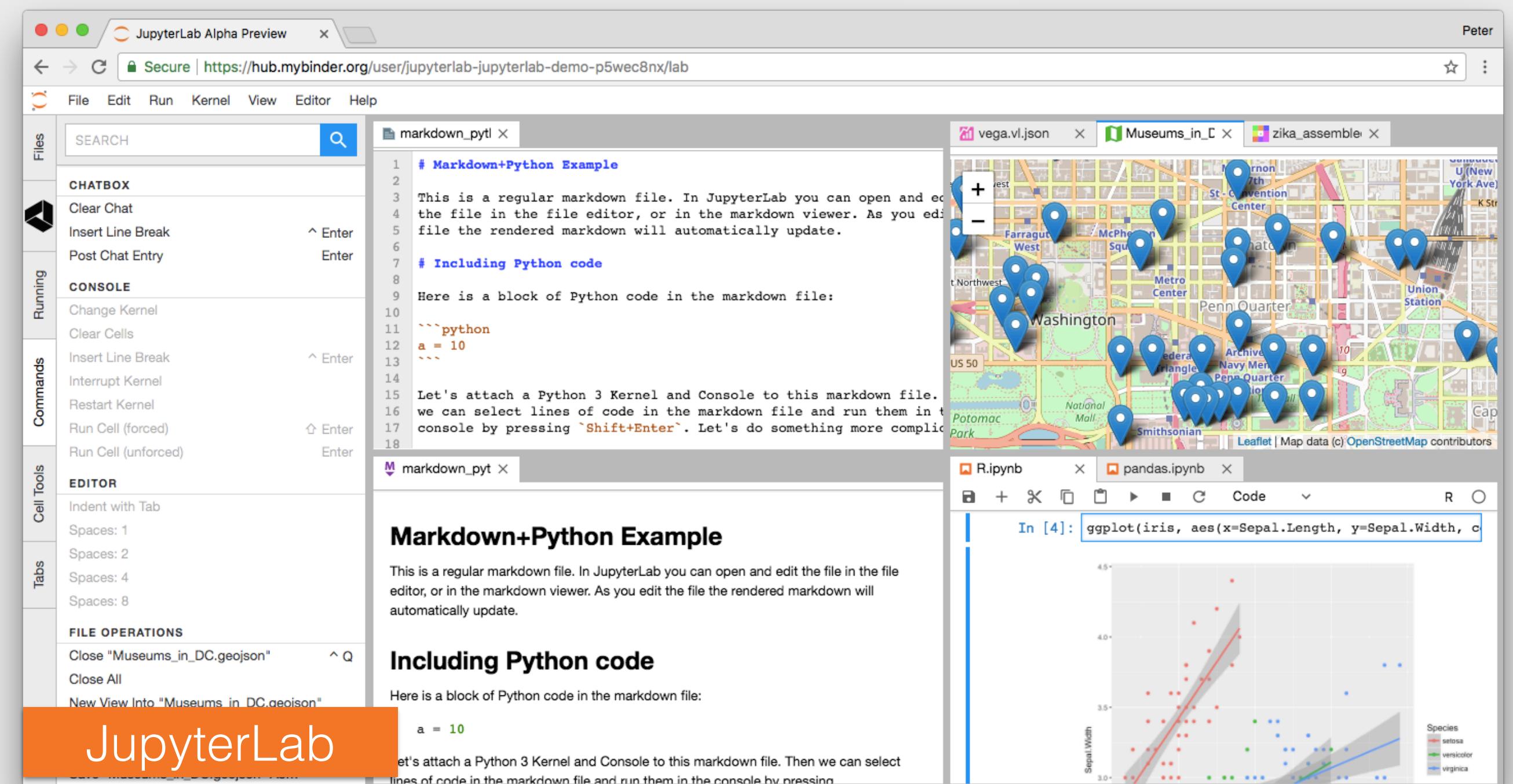
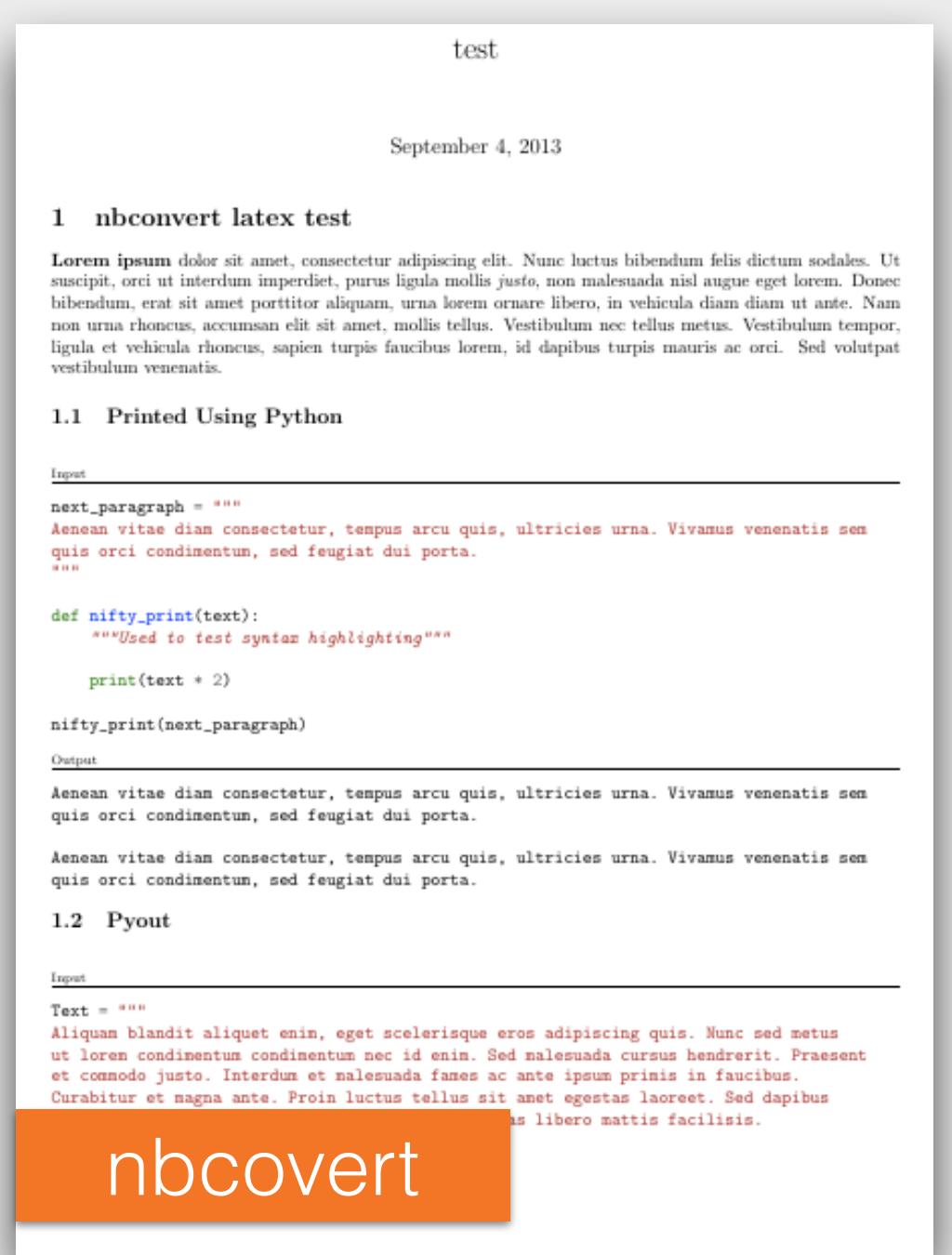
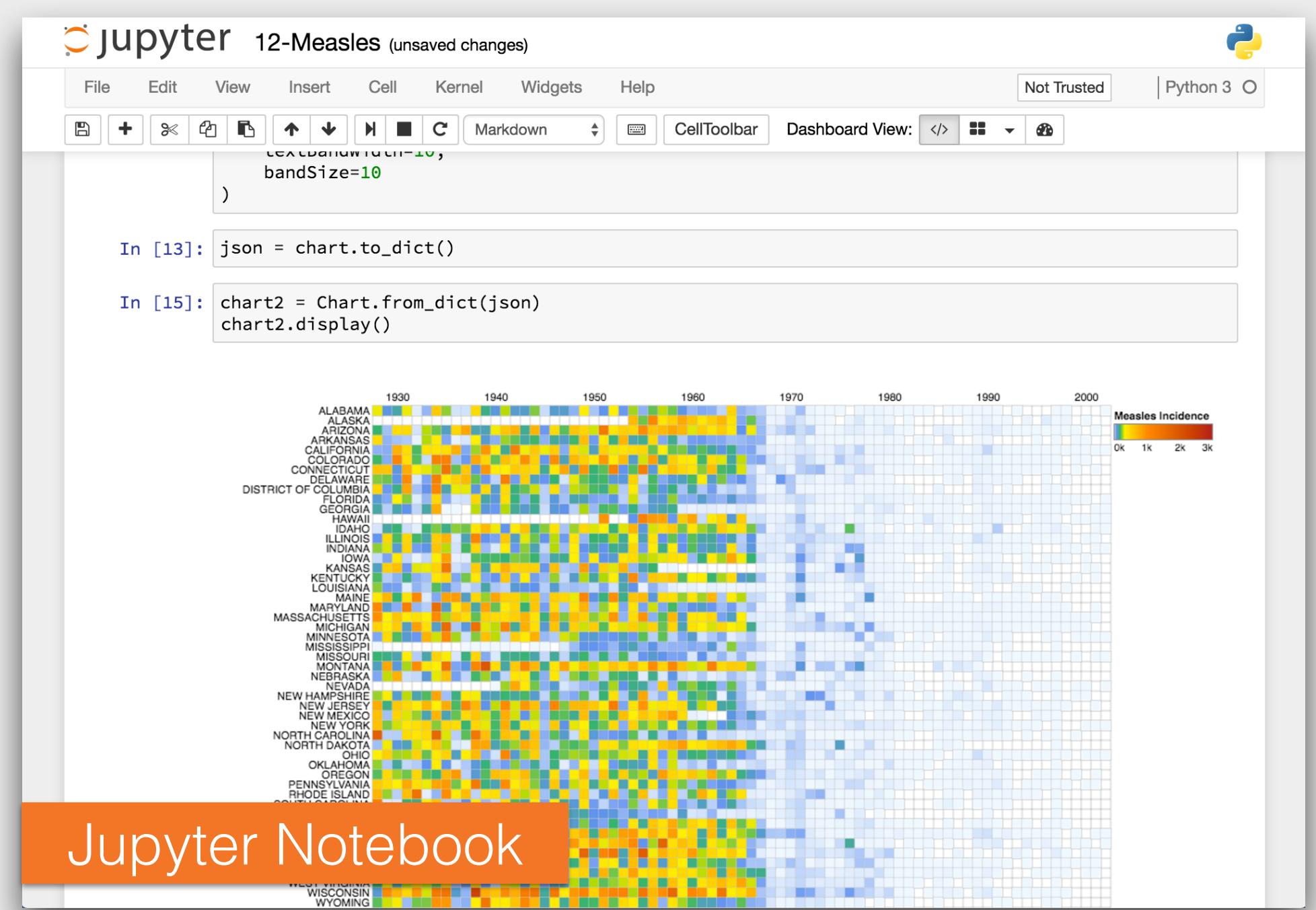
Services

Applications

Libraries

Standards

# Application Demos



# Service Demos



**nbviewer**

A simple way to share Jupyter Notebooks

Enter the location of a Jupyter Notebook to have it rendered here:

URL | GitHub username | GitHub username/repo | Gist ID Go!

**Programming Languages**

IPython

In [1]: In [2]: In [3]:

IP[y]: IPython Interactive Computing

File.open('lib/ruby/static/base/images/pythologo.png')

python

IRuby

In [1]: In [2]:

IRuby: Notebook

Books

Python for Signal Processing

O'Reilly Book

**binder** (beta)

Turn a GitHub repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Build and launch a repository

GitHub repo or URL

GitHub repository name or link

Git branch, tag, or commit

Path to a notebook file (optional)

master Path to a notebook file (optional) File ▾ launch

Copy the URL below and share your Binder with others:

Fill in the fields to see a URL for sharing your Binder.

Copy the text below, then paste into your README to show a binder badge: [launch](#) [binder](#)

# Technology Ecosystem



- Kernels for 60+ languages
- Extensions for slideshows, GeoJSON, WebGL, code folding, SQL magics, notifications, ...
- Support for Python, R, Julia, Scala, C++, ... libraries
- Alternative frontends
- Commercial offerings

**Google Colaboratory**

**Microsoft Azure Notebooks**

**IBM Data Science Experience**

**Hydrogen for Atom**

**Quantopian**

**nteract**

# The Community



# Users and Contributors



jupyterhub / jupyterhub

Unwatch 182 Unstar 2,677 Fork 600

Code Issues 148 Pull requests 7 Projects 0 Wiki Insights Settings

Pulse Contributors

# of ipynb files

THE LEONA M.  
HELM  
CHARITAB

Prack  
the #1 managed

yuvipanda 80 commits 1,688 ++ 780 --

barrachri 35 commits 527 ++ 336 --

# LIGO

News | Gravitational Waves De X Peter

Secure | https://www.ligo.caltech.edu/news/ligo20160211

LIGO Caltech Hanford Observatory | Livingston Observatory | LIGO MIT Search LIGO Lab

# LIGO

Laser Interferometer  
Gravitational-Wave Observatory  
Supported by the National Science Foundation  
Operated by Caltech and MIT

About Learn More News Gallery Educational Resources For Scientists Study & Work

Image credit: The SXS Project

## Gravitational Waves Years After Einstein's Death

News Release • February 11, 2016

Visit The Detectors

See also: LIGO Hanford Press Release

LIGO Opens New Window on the Universe by Listening to Colliding Black Holes

WASHINGTON, DC/Cascina, Italy

Getting Started

Data

- Events
- Bulk Data

Tutorials

- Software
- Detector Status
- Timelines
- My Sources
- GPS ↔ UTC
- About the detectors
- Projects
- Acknowledge LOSC

RELATED NEWS

LIGO Open Science Center Peter

Secure | https://losc.ligo.org/tutorials/

# LIGO Open Science Center

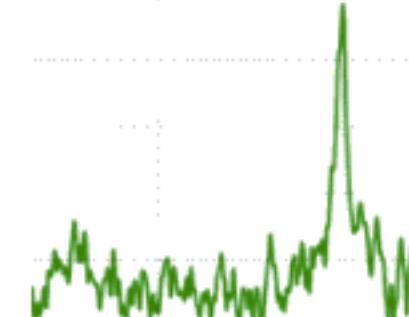
LIGO is operated by California Institute of Technology and Massachusetts Institute of Technology and supported by the U.S. National Science Foundation.

## Tutorials

Each tutorial will lead you step-by-step through some common data analysis tasks. While LIGO data can be analyzed using libraries in many software languages (C, C++, Matlab, etc.), most of these tutorials use Python. See also the [software examples page](#) for more examples.

See the [software setup page](#) for help installing software to run these tutorials.

### Binary Black Hole Events



Use matched filtering to find signals hidden in noise.

**Run:** Azure | mybinder (Beta)

**View:** [GW150914](#) | [LVT151012](#) | [GW151226](#) | [GW170104](#)

**Download:** [zip file with data](#) | [Jupyter notebook](#) | [python script](#)

### Quickview Notebook



Make summary plots for any short segment of LIGO data.

**Run:** Azure | mybinder (Beta)

# Buzzfeed



2016-01-tennis-betting-analysis x Peter

GitHub, Inc. [US] | https://github.com/BuzzFeedNews/2016-01-tennis-betting-analysis/blob/master/notebooks/tennis-analysis.ipynb

This repository Search Pull requests Issues Marketplace Explore

BuzzFeedNews / 2016-01-tennis-betting-analysis Watch 34 Star 213 Fork 51

Code Issues 2 Pull requests 0 Projects 0 Wiki Insights

Branch: master 2016-01-tennis-betting-analysis

jtemplon Initial commit

1 contributor

709 lines (708 sloc) | 24.1 KB

**Data and Analysis: Deep Dive**

The Python code below runs the anonymous analysis from "The Tennis Racket". The methodology contains many steps, including:

**Importing The Data**

```
In [1]: import pandas as pd  
import random
```

```
In [2]: betting_data = pd.read_csv("../data/betting.csv")
```

**Match Selection**

The code below excludes opening odds that are suspiciously low or high. It also excludes all bookmakers' opening odds for the matches that were excluded by OddsPortal. (This is a common method of suspicious betting.) The code also excludes matches where one player is a clear favorite — or "walkover" on OddsPortal.

The Tennis Racket

BuzzFeed NEWS / REPORTING TO YOU BuzzFeed Videos Quizzes Tasty More

Matt Chase for BuzzFeed News

A large image of a man in a white shirt and black shorts playing tennis on a red court. A yellow tennis ball is in the air above his racket. To the right of the image, the word "THE TENNIS RACKET" is written vertically in yellow letters, with each letter hanging from a thin white string.

A BUZZFEED NEWS / BBC INVESTIGATION

# Berkeley



# Kaggle



TensorFlow deep NN | Kaggle

validation\_accuracy => 0.9845

accuracy

step

When, we're happy with the outcome, we

Test data contains only images and labels

Predicted labels are stored into CSV file

In [25]:

```
# read test data from CSV file
test_images = pd.read_csv('../input/mnist-test.csv')
test_images = test_images.astype(np.float32)

# convert from [0:255] => [0.0:1.0]
test_images = np.multiply(test_im
```

Notebook Code Data (1) Output (2) Comments (172) Log Versions (36) Forks (1854) Fork Notebook

kaggle Search kaggle

Competition Datasets Kernels Discussion Jobs ... Sign In

Kirill Kliavin

**TensorFlow deep NN**

last run 2 years ago · Python notebook · 97155 views using data from Digit Recognizer · Public

491 voters

Notebook Code Data (1) Output (2) Comments (172) Log Versions (36) Forks (1854) Fork Notebook

Tags tutorial neural networks

## TensorFlow deep NN

A high-level tutorial into Deep Learning using MNIST data and TensorFlow library.

by @kakauandme and @thekoshkina

Accuracy: 0.99

**Prerequisites:** fundamental coding skills, a bit of linear algebra, especially matrix operations and perhaps understanding how images are stored in computer memory. To start with machine learning, we suggest [coursera course](#) by Andrew Ng.

Note:

Feel free to fork and adjust CONSTANTS to tweak network behaviour and explore how it changes algorithm performance and

# O'Reilly



Secure | https://www.safaribooksonline.com/oriole/# Peter

Everything

LEARN ALONGSIDE SMART PEOPLE SOLVING HARD PROBLEMS

# Oriole Online Tutorials

Powered by Jupyter

Oriole is a unique new medium for learning experience with executable code examples.

Led by some of the most brilliant minds in their fields, each lesson is an easily digestible, thought-by-thought tour of the solution to a problem in both narrative and code form. No set-up or installation is necessary. Tutorials require nothing more than a connection and a laptop. You can run the code within the environment. Make a mistake? Break it? Just delete and try again.

Start

- Setup
- Connecting to Cassandra
- Venues
- Events
- Explore events data
- Process mining

In this example), this specific query will first move the data from Cassandra to Spark, and then will perform filtering in Spark.

In general, it's good practice to move down and filter as much data as possible. This practice keeps throughput low and minimizes data transferred from one system to another.

Natalino Busa  
Head of Data Science

← Skip Here ▶ 0:00 / 29:18

Let us know what you think!  auto scroll/pause

Dataset: Events

Let's have a quick look at the events. They are stored in the Cassandra table `lbsn.events`.

```
1 val events = cc.sql("""select ts, uid, lat, lon, vid from lbsn.events where
2 lon>-74.2589 and lon<-73.7004 and
3 lat>40.65 and lat<40.75""")
```

↑

# Many, Many, More



GitHub, Inc. [US] | <https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks>

This repository Search Pull requests Issues Marketplace Explore Watch 294 Star 3,435 Fork 773

jupyter / jupyter Code Issues 83 Pull requests 0 Projects 0 Wiki Insights Settings Edit New Page

## A gallery of interesting Jupyter Notebooks

Ryan Francis edited this page 23 days ago · 51 revisions

This page is a curated collection of Jupyter/IPython notebooks that are notable. Feel free to add new content here, but please try to only include links to notebooks that include interesting visual or technical content; this should *not* simply be a dump of a Google search on every ipynb file out there.

**Important contribution instructions:** If you add new content, please ensure that for any notebook you link to, the link is to the rendered version using [nbviewer](#), rather than the raw file. Simply paste the notebook URL in the nbviewer box and copy the resulting URL of the rendered version. This will make it much easier for visitors to be able to immediately access the new content.

Note that [Matt Davis](#) has conveniently written a set of [bookmarklets and extensions](#) to make it a one-click affair to load a Notebook URL into your browser of choice, directly opening into nbviewer.

### Table of Contents

1. Entire books or other large collections of notebooks on a topic
  - Introductory Tutorials

Pages 6

 **Jupyter**

- Home
- A gallery of interesting Jupyter Notebooks
- Jupyter kernels
- Jupyter Notebook Server API

Clone this wiki locally

And You



# Get Started



- [jupyter.org](http://jupyter.org)
- [blog.jupyter.org](http://blog.jupyter.org)
- [groups.google.com/forum/#!forum/jupyter](http://groups.google.com/forum/#!forum/jupyter)
- [gitter.im/jupyter/jupyter](https://gitter.im/jupyter/jupyter)
- [try.jupyter.org](http://try.jupyter.org)
- pip (or conda) install notebook

# Get Involved



- Bug reports
- User feedback
- New ideas
- Documentation
- Graphic design
- Security audits
- Automation
- Test coverage
- Examples
- Plug-ins
- Code review
- Outreach
- User surveys
- Coordination
- Fundraising
- Bug fixes
- New features

# JupyterCon, August 2018, NYC

<https://conferences.oreilly.com/jupyter/jup-ny>

The screenshot shows the official website for JupyterCon 2018. At the top left is the O'Reilly logo. To its right, text reads "Brought to you by NumFOCUS Foundation and O'Reilly Media Inc.". The main title "jupytercon" is displayed in large orange and white letters. Below it, the text "THE OFFICIAL JUPYTER CONFERENCE" is followed by "AUG 21-22, 2018: TRAINING", "AUG 22-24, 2018: TUTORIALS & CONFERENCE", and "NEW YORK, NY". A testimonial from Tim Dobbins of The General Auto Insurance is shown: "Absolutely the best tech conference I've ever been to. A perfect mixture of networking, technical talks, and insight." - Tim Dobbins, The General Auto Insurance. A "APPLY TO SPEAK Proposals due March 6" button is visible. The footer includes links for "PROPOSALS", "SPONSORS", "ABOUT", and "ACCOUNT".

Discover how the world's most data-driven organizations are using Jupyter  
to analyze data, share insights, and create dynamic, reproducible data