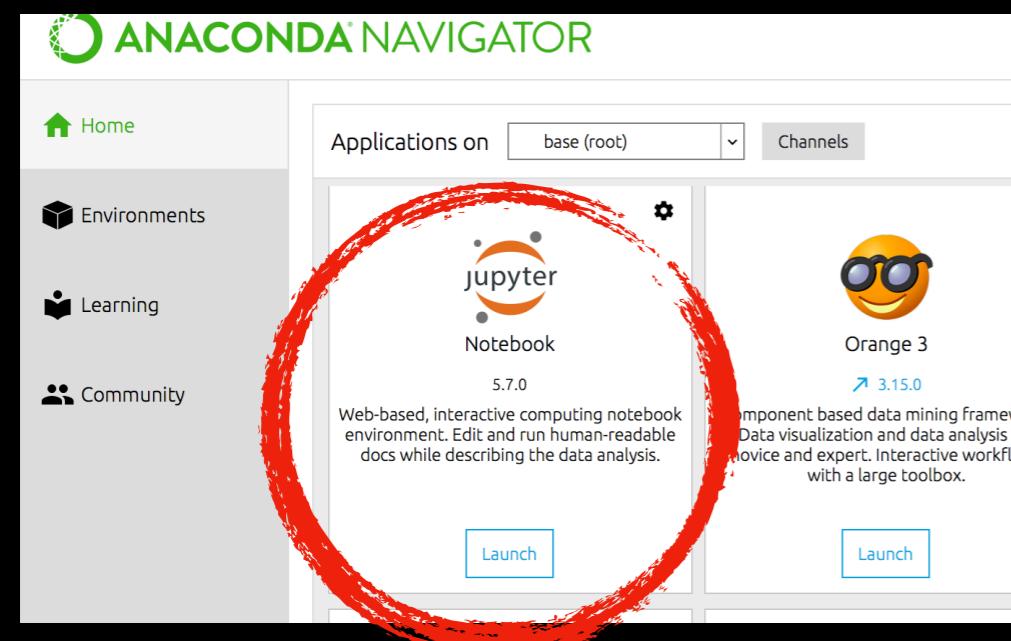


An Introduction To



Workshop for Digital Humanities
Michael Grossberg - 18/11/27

While we talk



Install
Anaconda
Open Notebook

Use your account at

ccnyhub.org

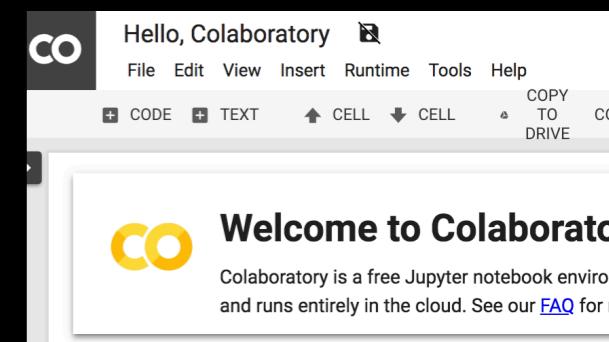
OR

OR



colab.research.google.com

OR



What is Jupyter?

Document Format

Corner detectors

If you think of edges as being lines, then corners are an obvious choice for features as they represent a change in both directions. One of the earlier corner detectors was introduced by Harris, and it is still a very effective corner detector. It's reliable and it's fast. There's a tutorial explaining how Harris works on the OpenCV site.

In [36]:

```
harris_test=input_image.copy()
#grayscale it
gray = cv2.cvtColor(harris_test, cv2.COLOR_BGR2GRAY)

gray = np.float32(gray)
blocksize=4 #
kernel_size=3 # sobel kernel: must be odd and fairly small

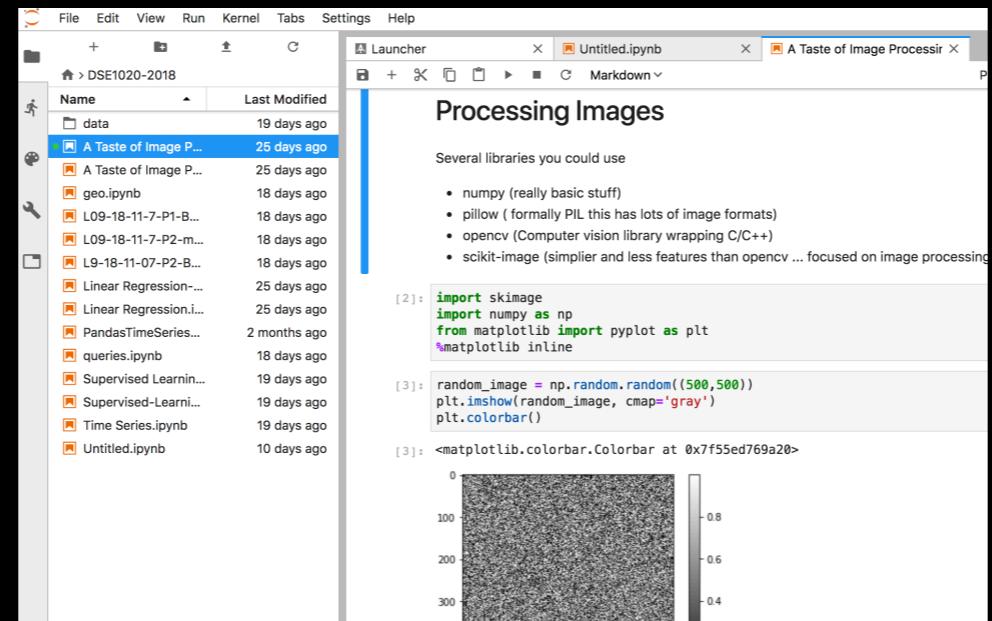
# run the harris corner detector
dst = cv2.cornerHarris(gray,blocksize,kernel_size,0.05) # parameters are b
r and Harris threshold

#result is dilated for marking the corners, this is visualisation related
er
dst = cv2.dilate(dst,None)
#we then plot these on the input image for visualisation purposes, using b
harris_test[dst>0.01*dst.max()]=0,0,255
plt.imshow(cv2.cvtColor(harris_test, cv2.COLOR_BGR2RGB))
```

Out[36]:

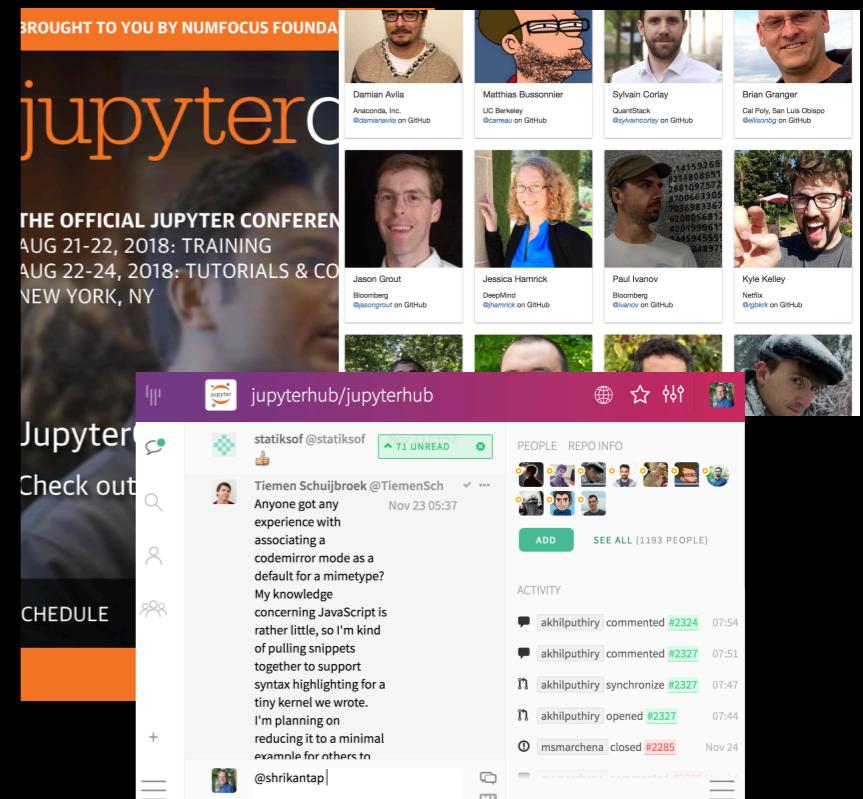


Tools/Services Using Format



The screenshot shows a Jupyter Notebook interface. On the left, there's a code cell (In [36]) containing Python code for corner detection on a dog's face image. On the right, there's a larger section titled "Processing Images" with sub-sections like "Several libraries you could use" and "import statements". A plot of a grayscale image with a color bar is also visible.

Community



The screenshot shows the GitHub repository for "jupyterhub/jupyterhub". It features a banner for the "OFFICIAL JUPYTER CONFERENCE" in New York. The repository page includes a list of contributors with their GitHub profiles, a "Check out" section, a "SCHEDULE", and a "PEOPLE" section showing recent activity and comments.

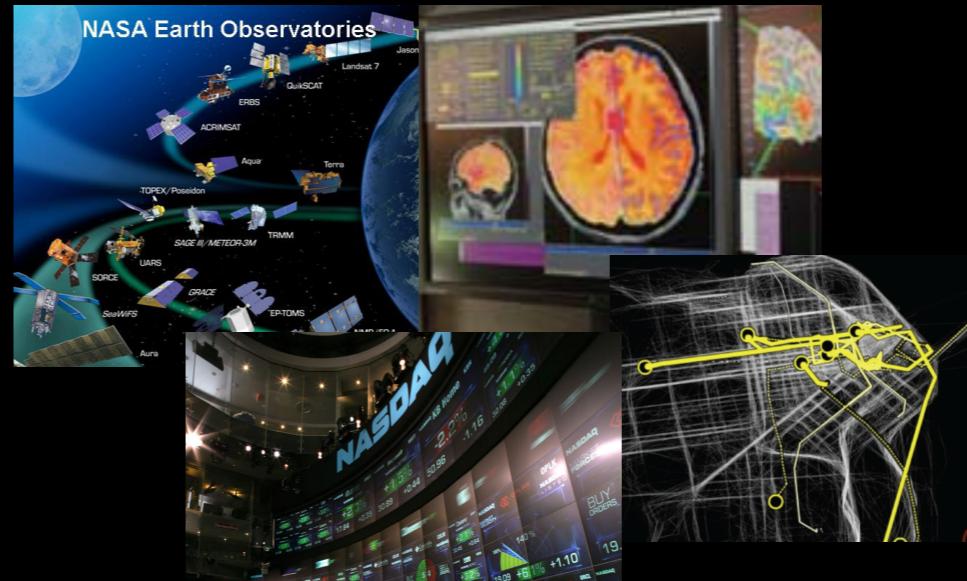
Open Source, Free (as in Beer and Speech), Widely Used

Motivation, Why would I care?

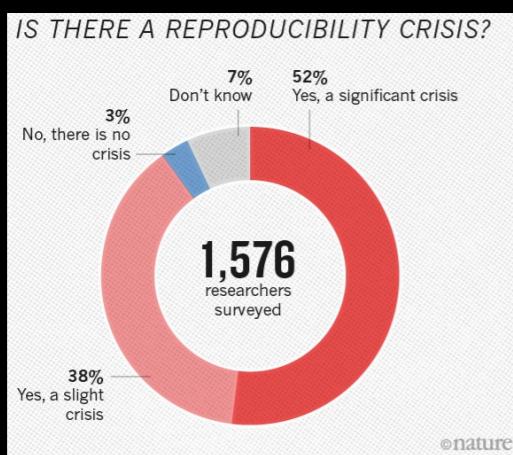
Ubiquitous Computing



Deluge of Data



Lack Reproducibility



Passive Learning FAIL



Ubiquitous Computing

Computing Everywhere



Easily Available Cloud Computing



New Literacy? How to Learn?

The Scratch website homepage features a blue header with the Scratch logo, navigation links for Create, Explore, Tips, and About, a search bar, and buttons for Join Scratch and Sign in. Below the header, there's a call-to-action: "Create stories, games, and animations Share with others around the world". It includes three icons: a cat (TRY IT OUT), a blue character (SEE EXAMPLES), and a yellow flower (JOIN SCRATCH it's free!). A central box displays a Scratch script for a "Welcome to Scratch" animation. Below this, a message states "A creative learning community with 41,046,888 projects shared" and links for ABOUT SCRATCH, FOR EDUCATORS, and FOR PARENTS. A "Featured Projects" section shows five thumbnail images: Wave (Version-1.0 Or jenniferandjim), Wall Jump run_script, membrane Briyani, Guess the CUP! WETRON_TV, and BackForth Spaacer.

Kids Games (Scratch)

The Social Drink Machine

I wish there was more information on this project, because I want to build it. Robofun who does some kick ass technology based marketing created [The Social Drink Machine](#).

The Social Drink Machine is a custom-built bar counter featuring a Raspberry Pi setup. It has a display screen showing various drink options and a QR code for users to scan. Behind the counter, several bottles of alcohol are lined up, and a row of glasses is ready for dispensing. The machine is illuminated with blue lights, creating a modern and interactive atmosphere.

Hardware Projects (Raspberry PI)

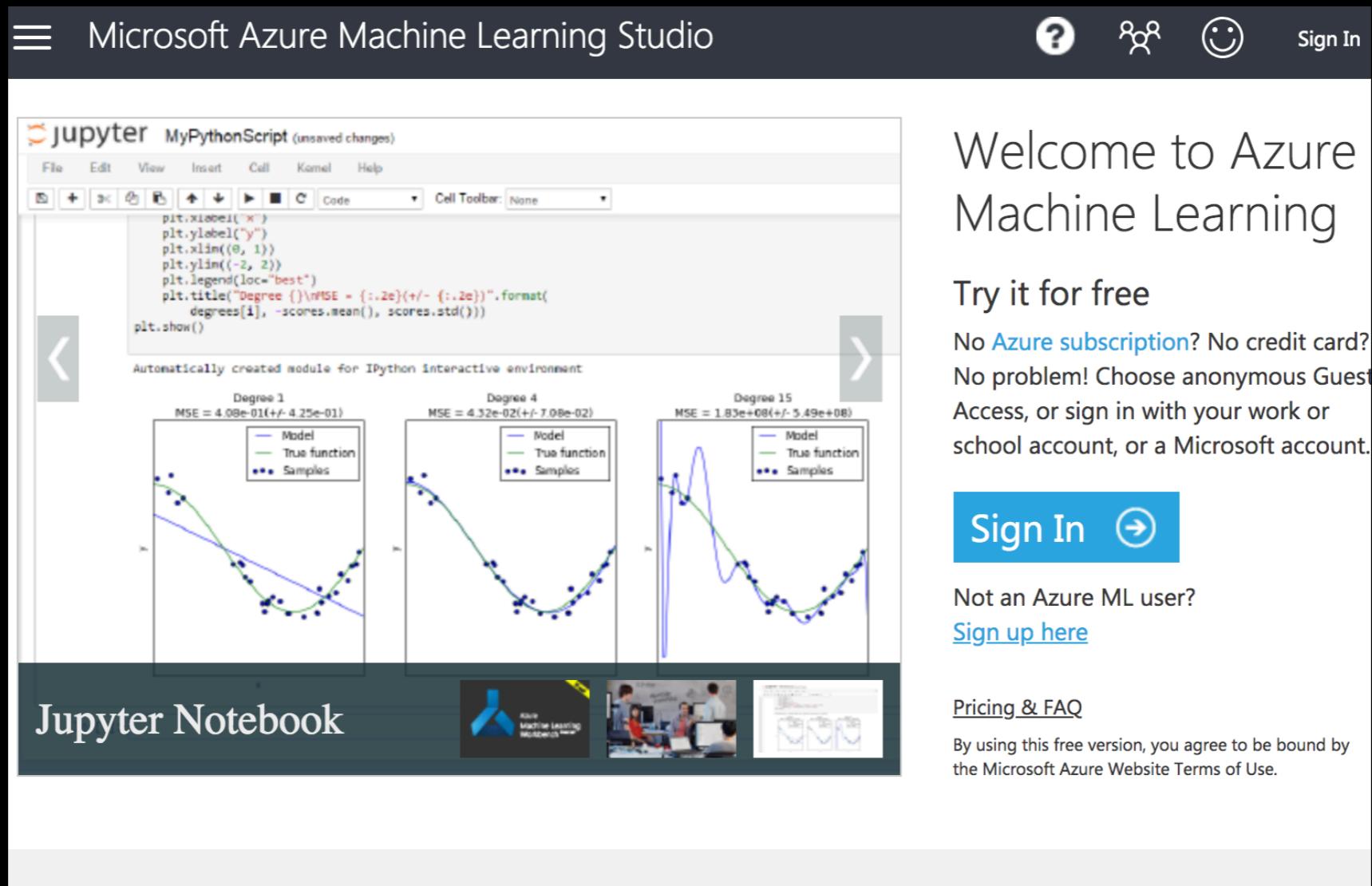
Mostly appeals to technophiles

What tasks can be made easier?



Rocket Powered Excel?

Jupyter Hubs in Cloud



The screenshot shows the Microsoft Azure Machine Learning Studio interface. At the top, it says "Microsoft Azure Machine Learning Studio". On the right, there are icons for help, user profile, and sign in, along with a "Sign In" button. The main area displays a Jupyter Notebook titled "MyPythonScript (unsaved changes)". The notebook contains Python code for plotting polynomial regression models. Below the code, three plots are shown for Degree 1, Degree 4, and Degree 15, comparing the Model (blue line), True function (green line), and Samples (blue dots). The plots illustrate how increasing the degree of the model leads to overfitting. The bottom of the notebook interface has a "Jupyter Notebook" label and several small icons. To the right of the notebook, a large "Welcome to Azure Machine Learning" message is displayed, followed by "Try it for free" and instructions for guest access or sign-in. A prominent blue "Sign In" button with a circular arrow icon is located below the welcome message. Further down, there is a link for non-users to "Sign up here" and a "Pricing & FAQ" section.

Welcome to Azure Machine Learning

Try it for free

No [Azure subscription](#)? No credit card?
No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

[Sign In](#) 

Not an Azure ML user?
[Sign up here](#)

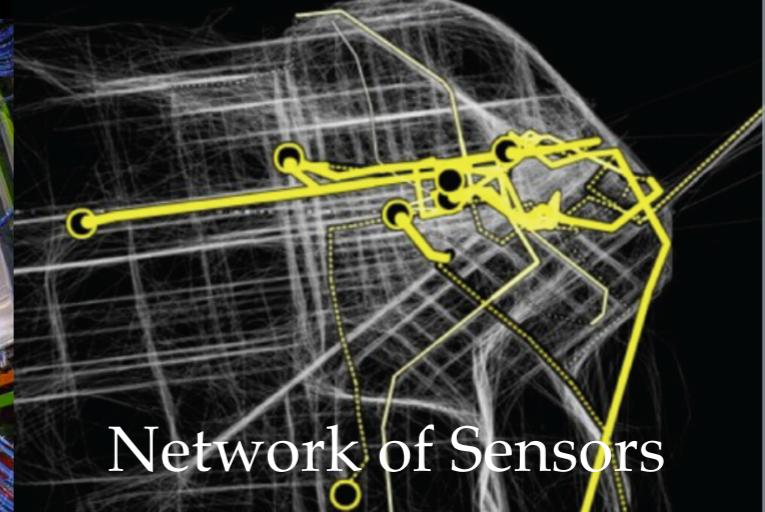
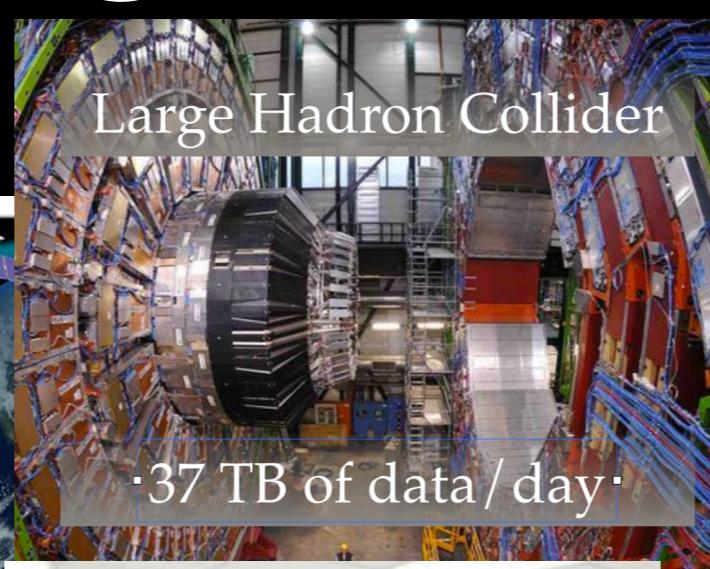
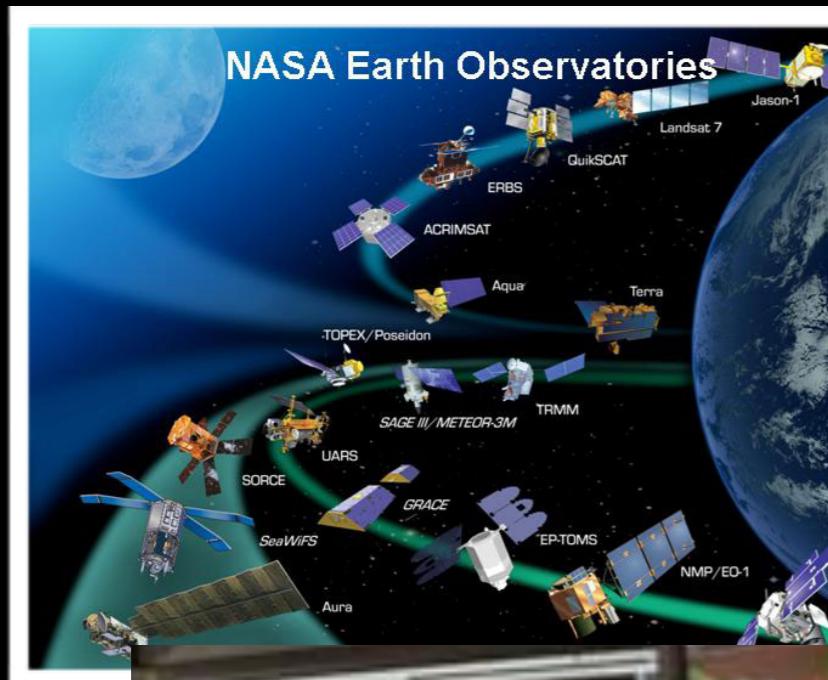
[Pricing & FAQ](#)

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

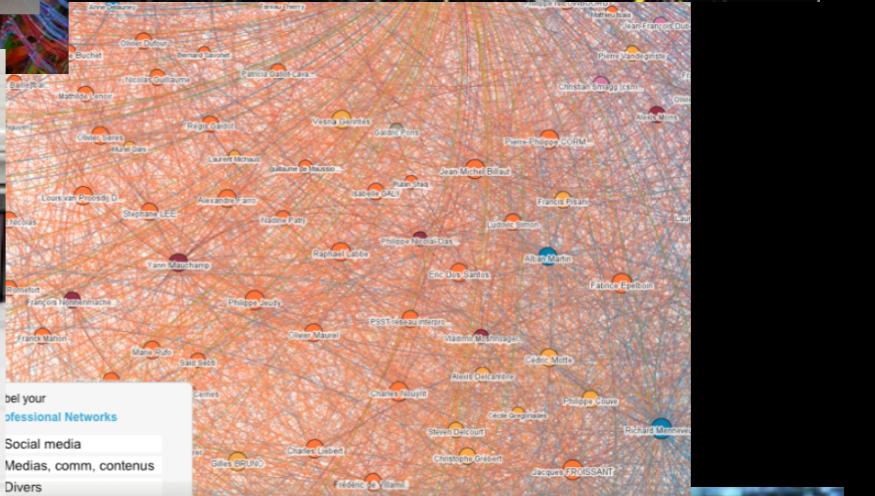
Access to Arbitrary Computation Power
Think Google Docs For Computation

Deluge of Data

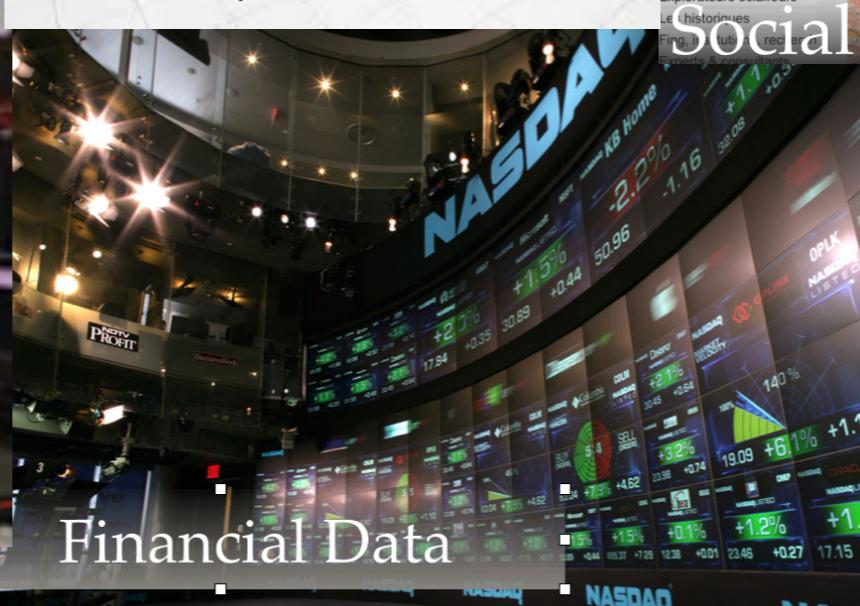
Exploding New Sources



Model / Simulation



Security Data



Financial Data

Every two days now
we create as much
information as we
did from the dawn of
civilization up until
2003

Eric Schmidt, former CEO
Google, 2010



Digital Humanities Corpora

Literature

Project Gutenberg

Free ebooks - Project Gutenberg

Book search · Book categories · Browse catalog · Mobile site · Report errors · Terms of use

Some of the Latest Books



Welcome

Project Gutenberg offers over 57,000 free eBooks. Choose among free epub books, free kindle books, download them or read them online. You will find the world's great literature here, with focus on older works for which copyright has expired. Thousands of volunteers digitized and diligently proofread the eBooks, for enjoyment and education.

No fee or registration is required. If you find Project Gutenberg useful, please consider a small donation to help Project Gutenberg digitize more books, maintain

Browse Catalog
Bookshelves
Main Page
Categories
Contact Info

Project Gutenberg appreciates your donation!

Donate

Why donate?

in other languages
Português
Deutsch
Français

hosted by ibiblio

Project Gutenberg
+57K Books

Art History

Search the Collection

Search for Artist, Culture, Title, Accession #, Gallery, etc. All Fields

Filter By: Object Type / Material, Geographic Location, Date / Era, Department

Show Only: Highlights, Artworks With Image, Artworks On Display, Open Access

Sort: Relevance



Girdle ornament
China
Bowl
13th century
Lamp
4th century
Design for Victoria, no. 213
Brewster & Co. 1884

The Metropolitan Museum of Art
open access +400K images

History

CHRONICLING AMERICA
Historic American Newspapers

Search America's historic newspaper pages from 1789-1963 or use the U.S. Newspaper Directory. Search Pages Advanced Search All Digitized Newspapers 1789-1963

All states from 1789 to 1963 enter one or more search words

Pages Available: 14,329,958



100 Years Ago Today: 11/25/1918 (75 issues)

The daily times. [volume] (600p.) Wilson, N.C.

The evening world. (18pp.) New York, N.Y.

Hickory daily record. (4pp.) Hickory, N.C.

About Chronicling America
About the Site and API
Recommended Topics
Help

More Resources
National Digital Newspaper Program
NDNP Award Recipients
Newspaper and Current Periodicals Reading Room
Ask LC Newspaper & Current Periodicals Librarian
Historic Newspapers on Flickr (part of the LC Flickr Commons photostream)
Beyond Words (Help find pictures in historic newspapers!)

US. Library of Congress

Many many other digital databases

Keep Data with Computation in The Cloud

Registry of Open Data on AWS



Landsat 8

[earth observation](#) [satellite imagery](#) [gis](#) [natural resource](#) [sustainability](#) [disaster response](#)

An ongoing collection of satellite imagery of all land on Earth produced by the Landsat 8 satellite.

[Details →](#)

Usage examples

- EOS Land Viewer by [Earth Observing System](#)
- Spectator - tracking Landsat 8, accessing the data and quick preview by [Spectator](#)
- Integrate imagery from the full Landsat archive into your own apps, maps, and analysis with [Landsat image services](#) by [Esri](#)
- COG-Explorer - View Cloud Optimized GeoTIFF images in the browser directly from object storage by [EOX](#)
- Using Vector tiles and AWS Lambda, we can build a really simple API to get Landsat and Sentinel images by [Remote Pixel](#)

[See 13 usage examples →](#)

Common Crawl

[encyclopedic](#) [machine learning](#) [natural language processing](#) [internet](#)

A corpus of web crawl data composed of over 25 billion web pages.

[Details →](#)

Usage examples

- N-gram counts and language models from the [Common Crawl](#) by [Christian Buck](#), [Kenneth Heafield](#), [Bas van Ooyen](#)
- Index to WARC Files and URLs in Columnar Format by [Sebastian Nagel](#)
- Learning word vectors for 157 languages by [Facebook AI Research](#)

Lack Reproducibility

Incentives for Bad Research Practices

Research Article

Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling

Leslie K. John¹, George Loewenstein², and Drazen Prelec³
¹Marketing Unit, Harvard Business School; ²Department of Social & Decision Sciences, Carnegie Mellon University; and ³Sloan School of Management and Departments of Economics and Brain & Cognitive Sciences, Massachusetts Institute of Technology

Abstract
Cases of clear scientific misconduct have received significant media attention recently, but less flagrantly questionable research practices may be more prevalent and, ultimately, more damaging to the academic enterprise. Using an anonymous elicitation format supplemented by incentives for honest reporting, we surveyed over 2,000 psychologists about their involvement in questionable research practices. The impact of truth-telling incentives on self-admissions of questionable research practices was positive, and this impact was greater for practices that respondents judged to be less defensible. Combining three different estimation methods, we found that the percentage of respondents who have engaged in questionable practices was surprisingly high. This finding suggests that some questionable practices may constitute the prevailing research norm.

Keywords
professional standards, judgment, disclosure, methodology

Received 5/20/11; Revision accepted 10/20/11

Although cases of overt scientific misconduct have received significant media attention recently (Altman, 2006; Deer, Louis, 1993). In the study reported here, we measured the percentage of psychologists who have engaged in QRP.

Psychological Science
23(5) 524-532
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611430953
<http://pss.sagepub.com>

 Association for Psychological Science

 SAGE

Research misconduct - The grey area of Questionable Research Practices

30 September 2013

We all know that fabrication, falsification and plagiarism fall under the definition of research misconduct. However, the VIB definition of research misconduct is broader than that. It also includes practices "that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting or reporting research."

VIB guidelines and recommendations

There is a continuum from truly correct to truly deceptive scientific research. The grey area in between is often referred to as 'Questionable Research Practices' (QRP). Whether a form of QRP will qualify as research misconduct is very much determined by the intent of the researcher.

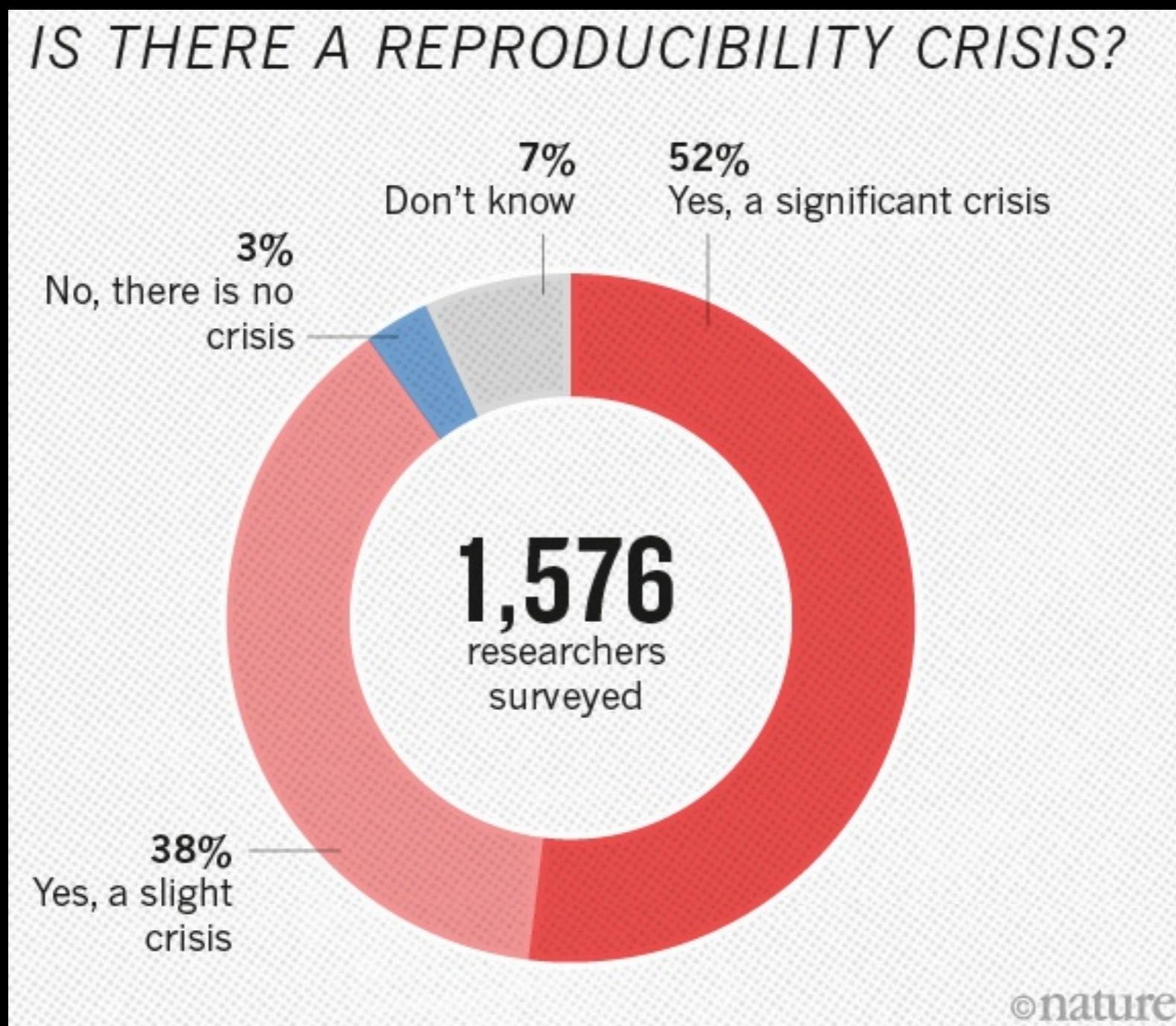
Examples of QRP:

- Neglecting negative outcomes
- Using inappropriate statistics to support one's hypothesis
- Inappropriate research design
- Leaving out relevant controls
- Inappropriate re-use of controls
- Removal of 'outliers'
- Conscious bias
- Unethical experimentation
- Peer review abuse

Examples of actions that lead to incorrect data being taken up in the scientific record but are not considered scientific misconduct

- Honest error

Most See Crisis



Writing on the wall

The screenshot shows a portion of the NIH Grants and Funding: Extramural Programs (EP) website. At the top, there's a blue header with the NIH logo, "U.S. National Library of Medicine", and "NLM Customer Support" with social media links. Below the header is a search bar and a menu icon. The main content area has a teal background and displays the following text:

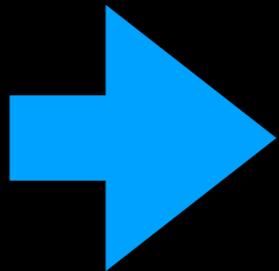
Grants and Funding: Extramural Programs (EP)

[EP Home](#) | [Grant Programs](#) | [Awards](#) | [Deadlines & Forms](#) |
[Help](#) | [EP Site Map](#)

[Home](#) > [EP Home](#) > [Help](#)

Data Sharing Plan Requirements

Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health. It is NIH policy that any investigator submitting a grant application seeking



**Reproducibility
Requirements**

NOW

SOON

Jupyter Notebooks Present Results **with** Reproducibility

The Atlantic

Popular

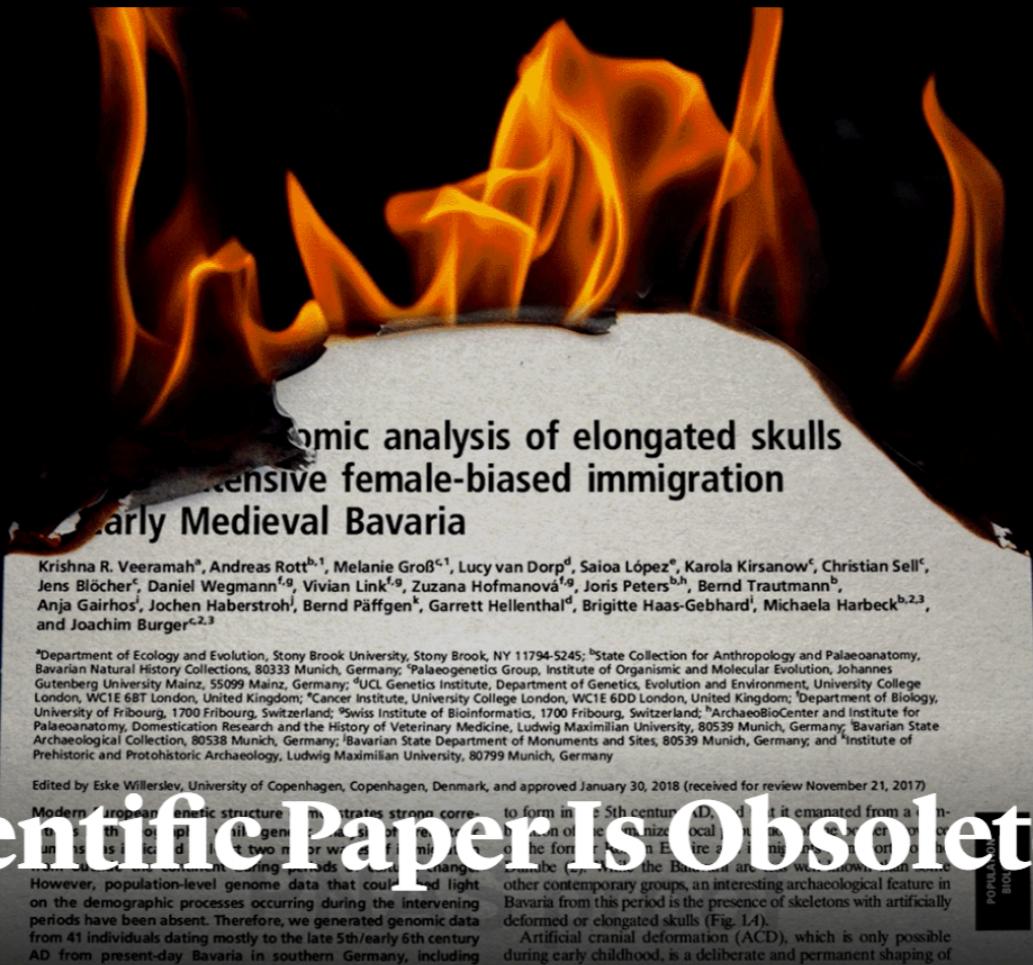
Latest

Sections ▾

Magazine ▾

More ▾

Subscribe



The Scientific Paper Is Obsolete

Here's what's next.

Passive Learning
FAIL

Where great ideas
go to **die**



Lectures are Broken

The Washington Post
Democracy Dies in Darkness

[Answer Sheet](#) • Analysis

It puts kids to sleep – but teachers keep lecturing anyway. Here's what to do about it.

By Valerie Strauss
July 11, 2017



TIME [Subscribe](#)

EDUCATION

Why Long Lectures Are Ineffective

If students can only focus for 15-minute intervals, shouldn't we devote precious class time to something more engaging?

By Salman Khan | Oct. 02, 2012

[Share](#) [Like 4.4K](#) [Tweet](#) [G+](#) [Share](#) [Save](#) [Read Later](#)

Each school day, millions of students move in unison from classroom to classroom where they listen to 50- to 90-minute lectures. Despite there being anywhere from 20 to 300 humans in the room, there is little actual interaction. This model of education is so



Science [Home](#) [News](#) [Journals](#) [Topics](#) [Careers](#)

[Log in](#) | [My account](#) | [Contact us](#)

[Become a member](#) [Renew my subscription](#) | [Sign up for newsletters](#)

[f](#) 1K [@ Save](#) [in](#)



Blah? Traditional lecture classes have higher undergraduate failure rates than those using active learning techniques, new research finds. WIKIMEDIA

Lectures aren't just boring, they're Ineffective, too, study finds

By Aleszu Bajak | May. 12, 2014, 3:00 PM

Active Learning



Smaller Class Discussions



Clickers for Large Classes

Code Along/Problem Solve

University of California, Berkeley

Data 8: Foundations of Data Science

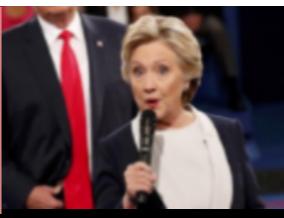
Jupyter



Foundations of Data Science: A Data Science Course for Everyone



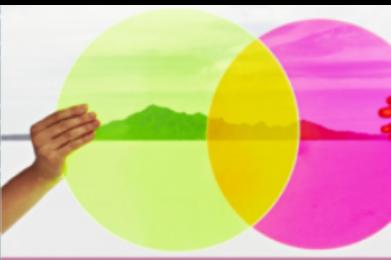
ART W23AC
Data Arts



XRHETOR R1A
The Craft of Writing



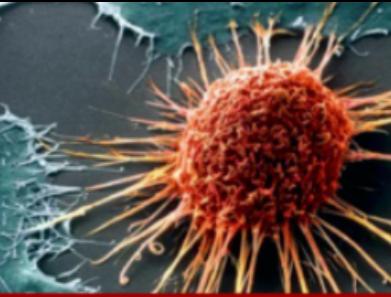
LEGALST 88
Taking Measure of the Justice System



COGSCI 88
Data Science & the Mind



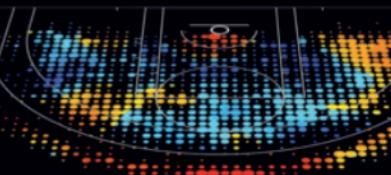
PSYCH 88
DS for Cognitive Neuroscience



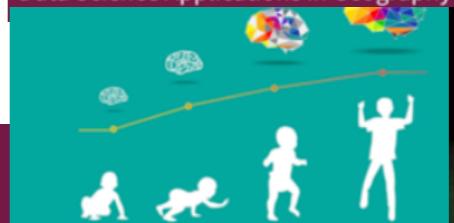
MCB 88
Immunotherapy of Cancer



GEOG 88
Data Science Applications in Geography



L&S 88-2
Sports Analytics



L&S 88-1
Children in the Developing World



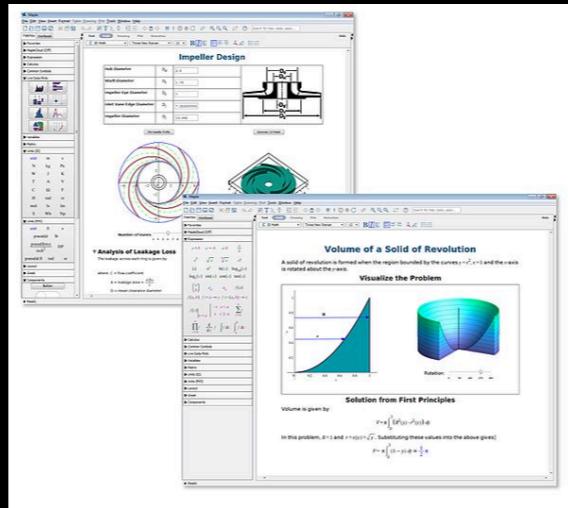
STAT 88
Probability & Mathematical Statistics

Facets of Jupyter

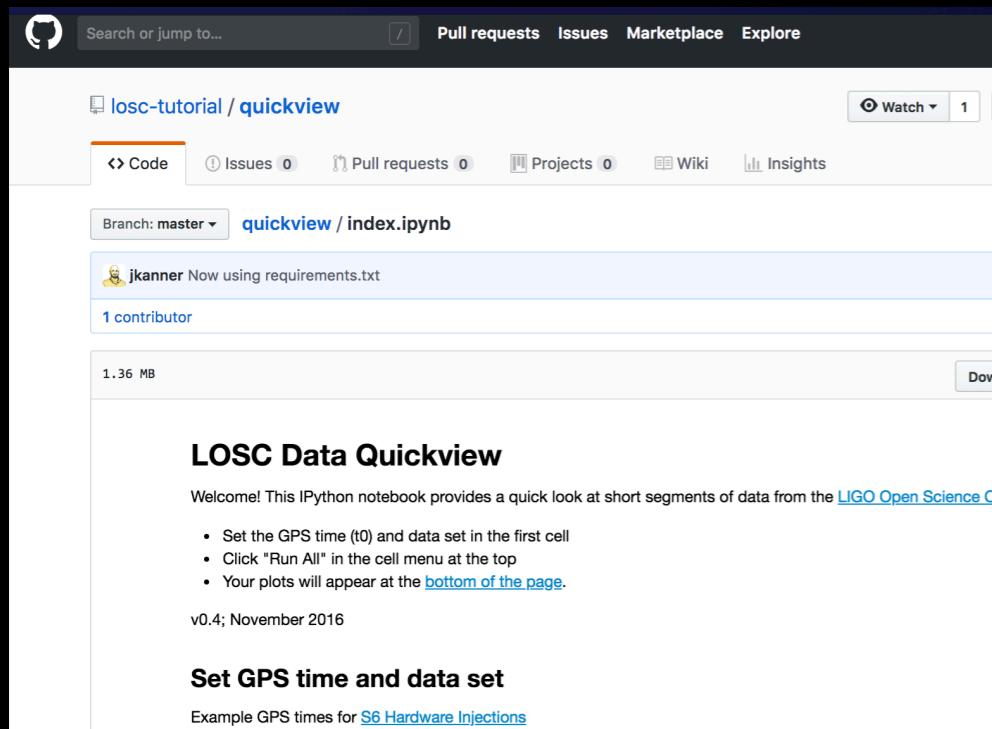
Computational Notebook Interface

A screenshot of a Mathematica notebook titled "bifurcation.nb". The code defines a function `bifurcate` that takes parameters `x0`, `a0`, `k0`, `white`, and `r`. It uses `NestWhileList` to iterate until `k0 + k - 1` is reached. The code also includes a logistic map iteration and a plot of `plotData` against `r` from 3.5 to 4.0. The resulting plot shows a complex, fractal-like bifurcation diagram.

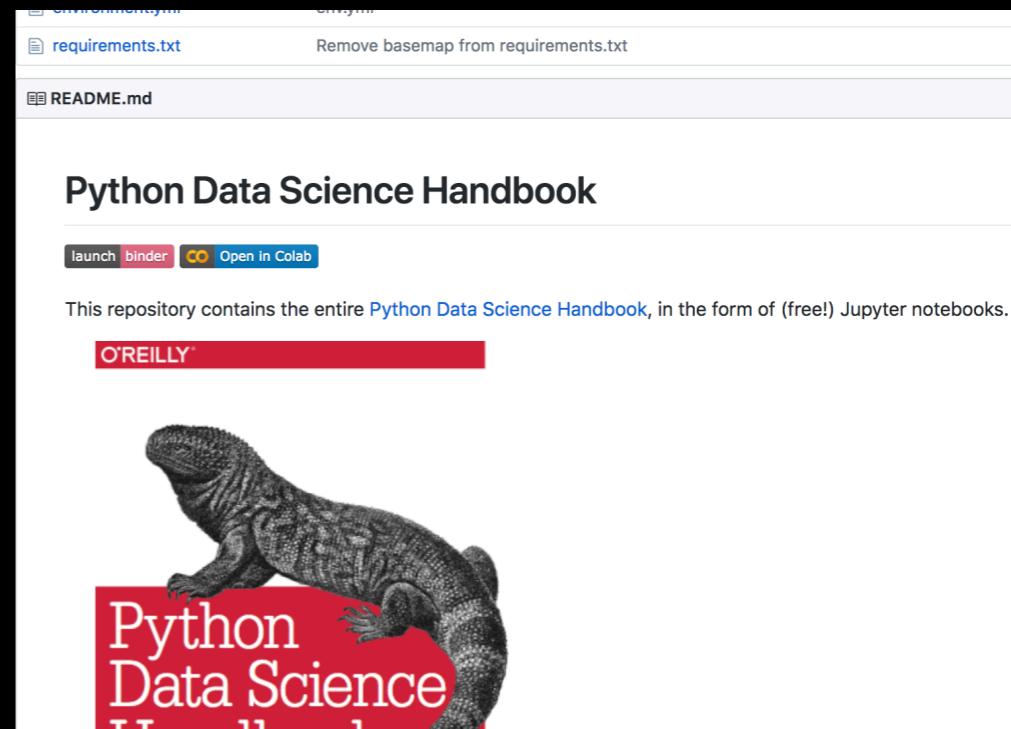
Mathematica Notebook
1988



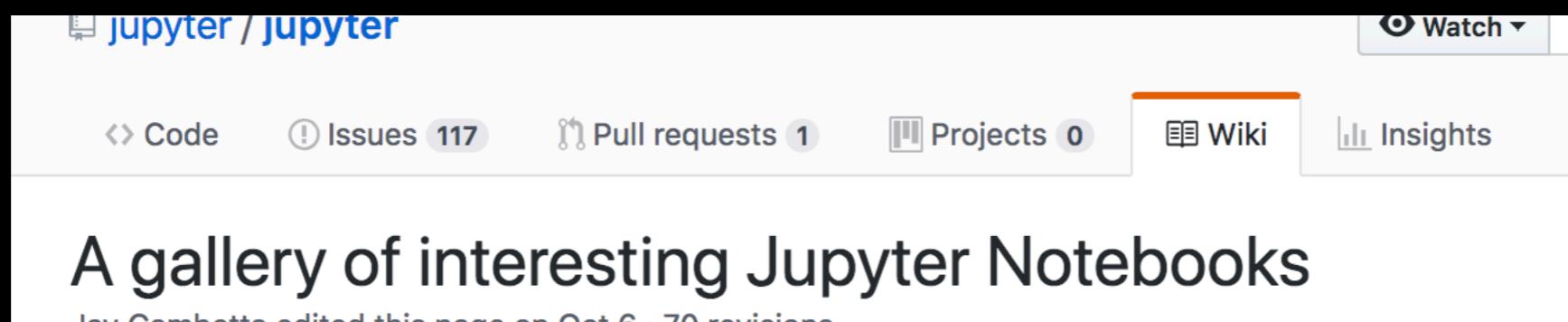
Format: Experiment Archive, HW, Tutorials, Books



**Reproducible LIGO Events
Colliding Black Hole on Github
also Azure**

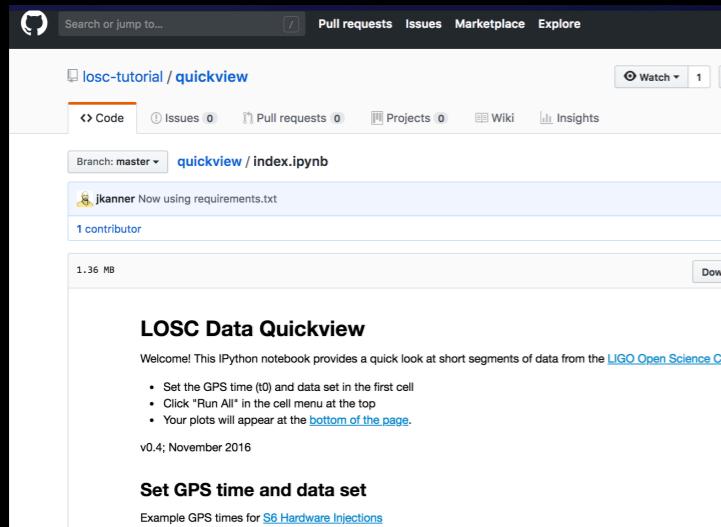


**Data Science Book Published
as Jupyter Notebooks**

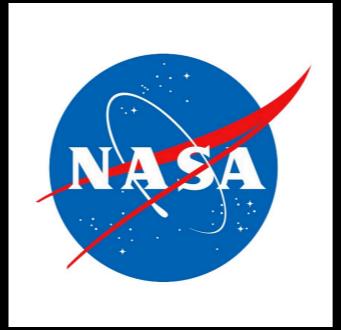


<https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks#statistics-machine-learning-and-data-science>

Who uses it?



Academics



Mayor's Office of
Data Analytics
MODA-NYC

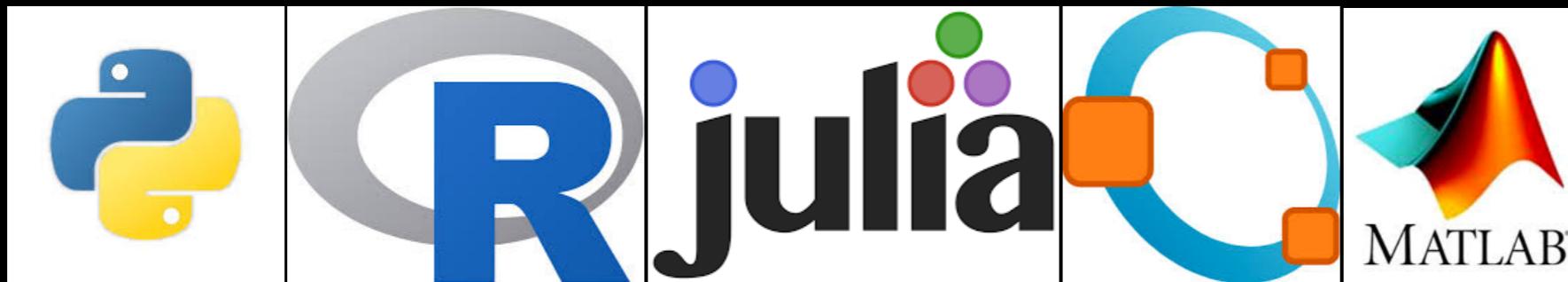


Government



Industry

Programming Languages



Python

R

Julia

Octave

Matlab



Fortran

Scheme

C++

Ruby

Haskell



JavaScript

C#

Scala

Erlang

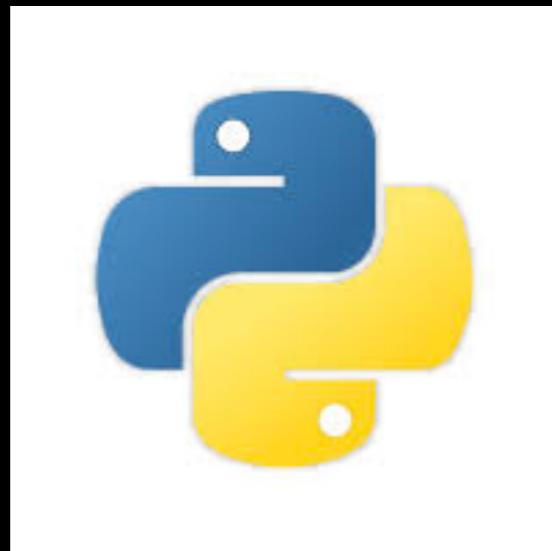
Java

Kotlin

Go, Perl, Bash, Powershell, Babel, Lua, IDL, Forth, OCaml, Clojure, etc.

<https://github.com/jupyter/jupyter/wiki/Jupyter-kernels>

Mostly Python and R



Python



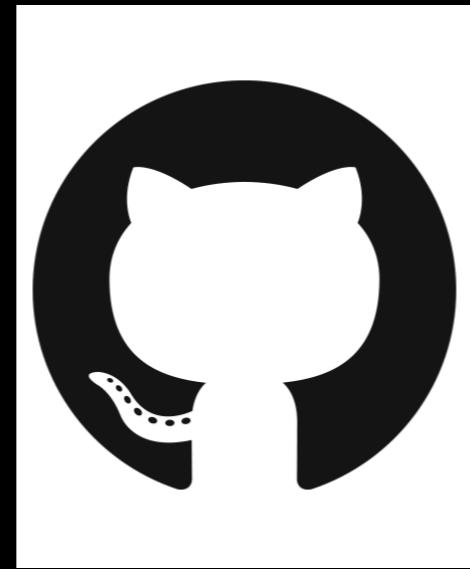
R



Format



Share File .ipynb



**Share .ipynb
on Github**



Convert to PDF

Run Notebook



CCNYhub.org

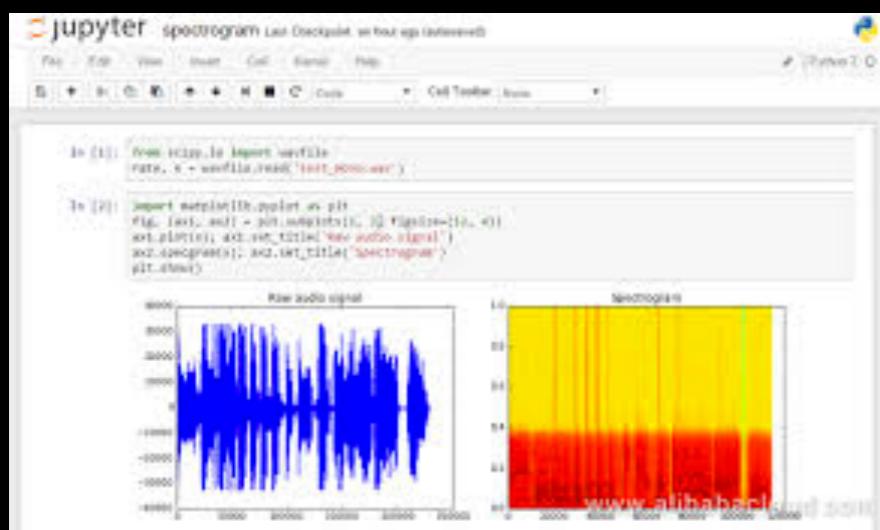
Locally on your
computer

Public Demo

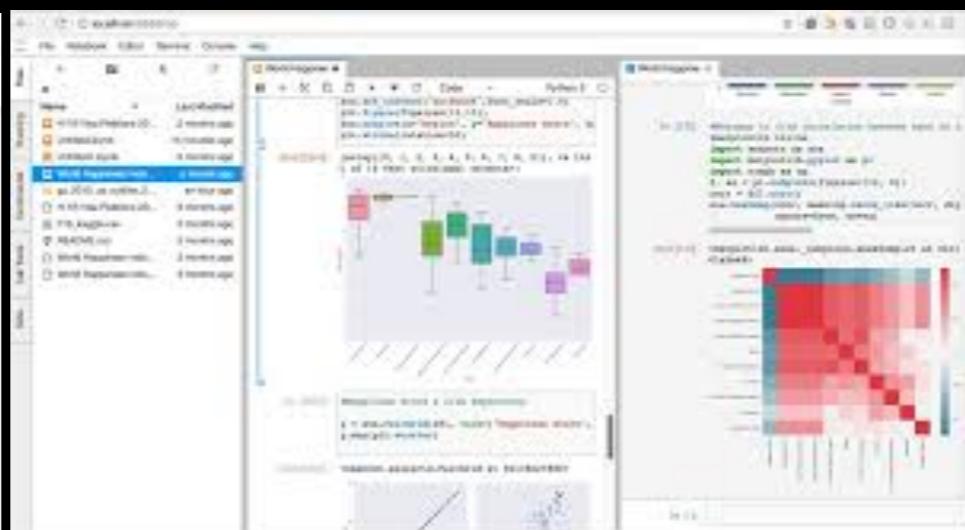
Commercial
Cloud

Local
Cloud

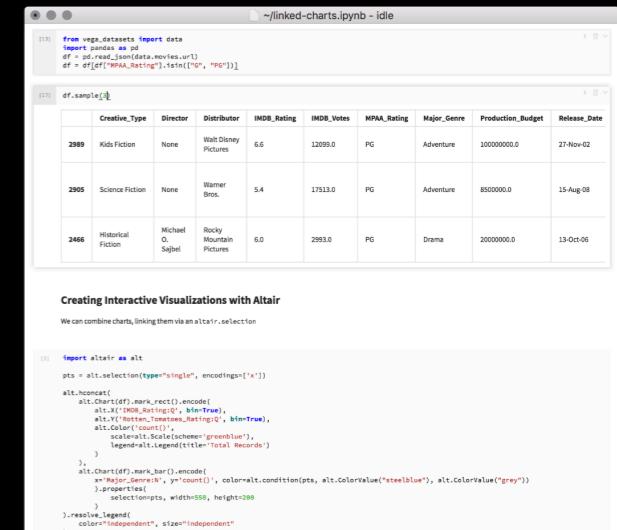
Notebooks vs Lab Interface



Jupyter Notebook



JupyterLab



Nteract Notebook

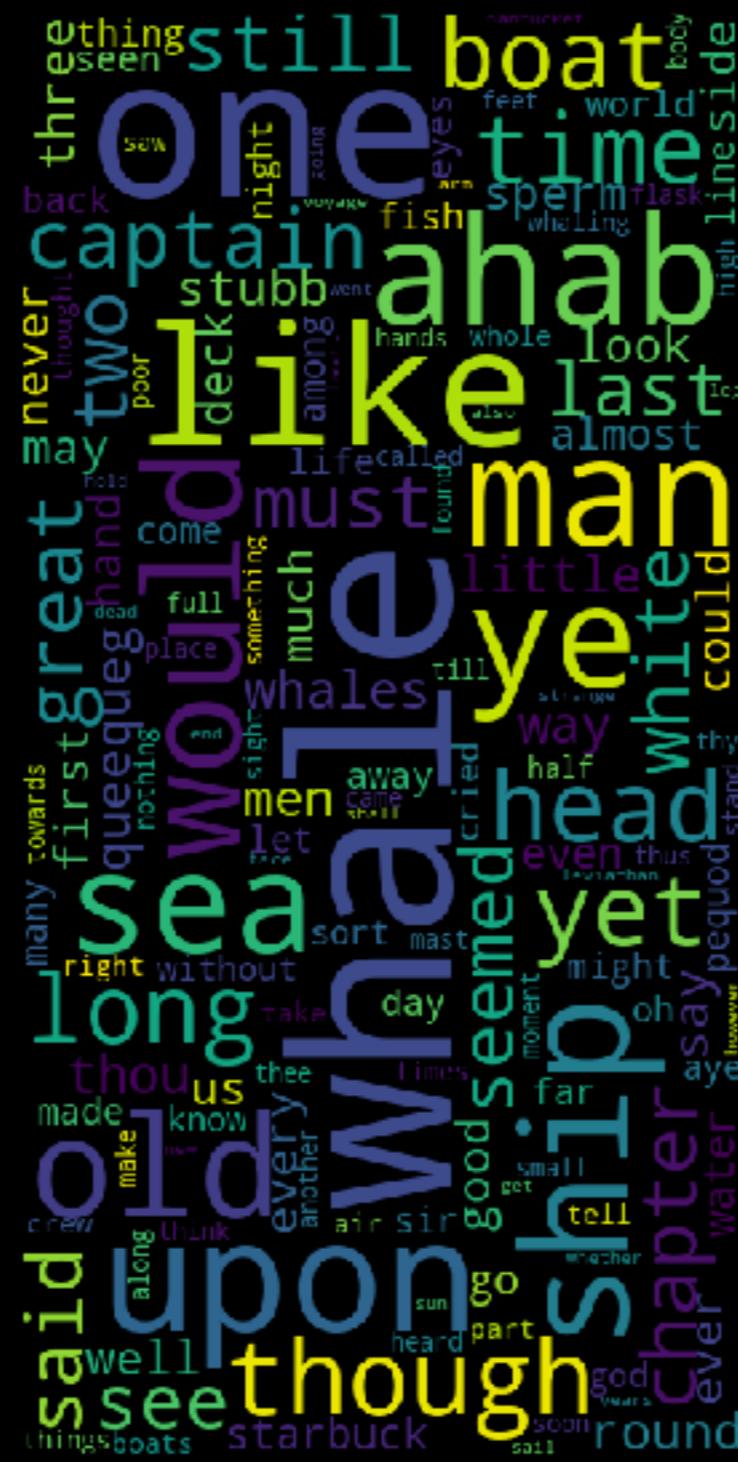
Large Supportive Community



Data Science +
Humanities?

What can we do?

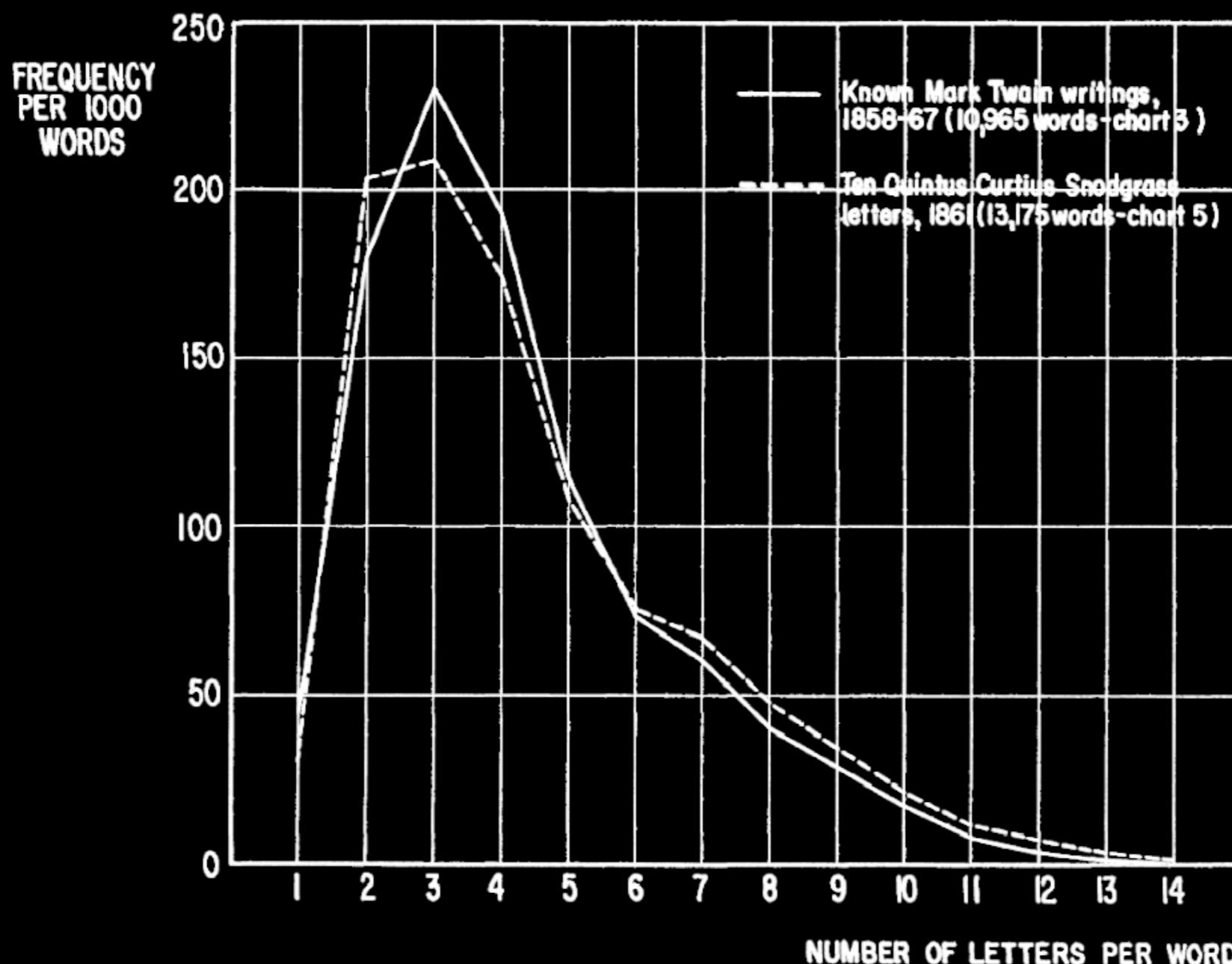
Count/Measure Stuff



- **Text Length**
 - **Word Distributions**
 - **Frequency of contractions**
 - **Numbers of features**
 - **in an image**
 - **Frequency of Characters**
 - **Frequency of Place Names**

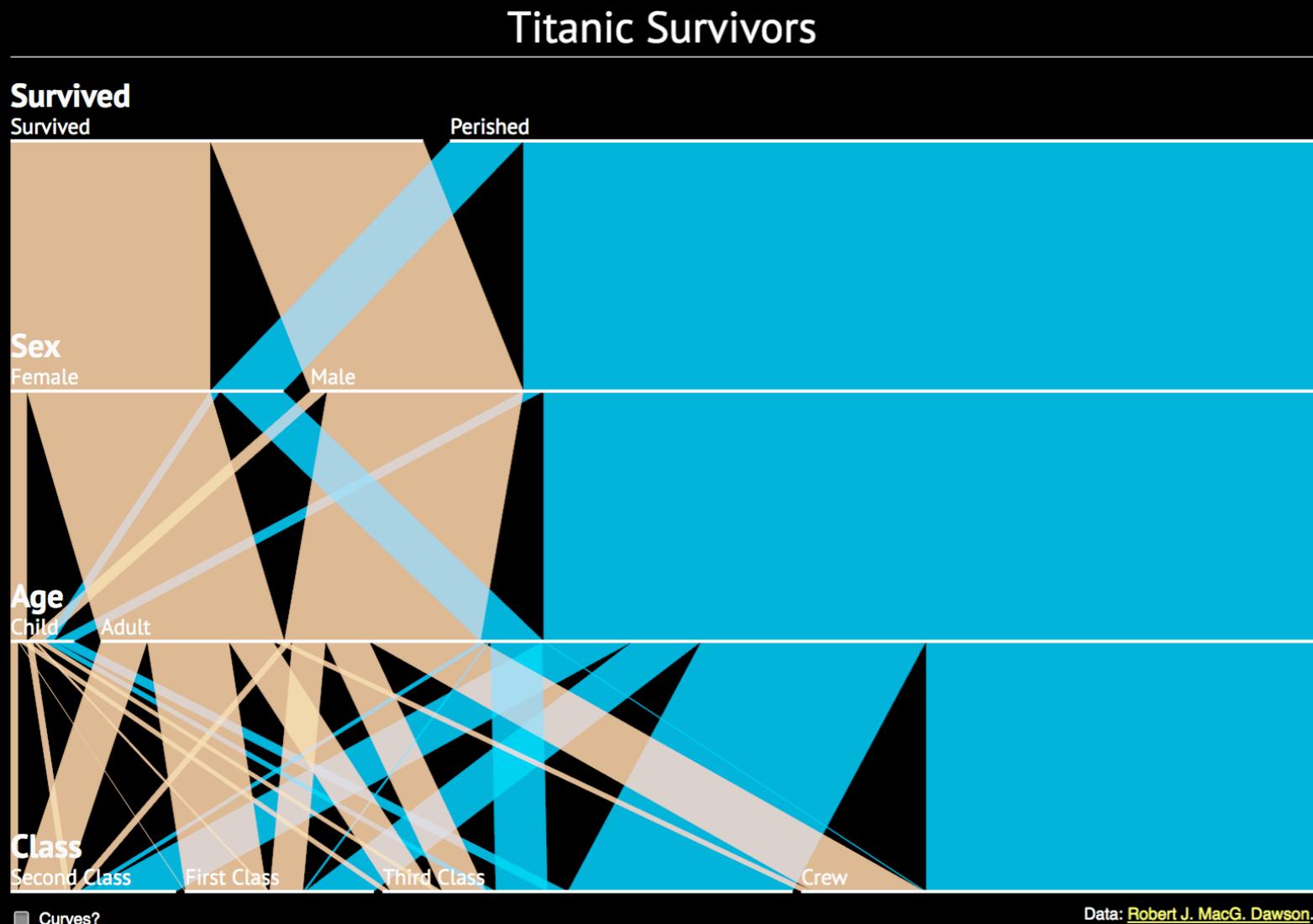
MARK TWAIN AND THE QUINTUS CURTIUS SNODGRASS
LETTERS: A STATISTICAL TEST OF AUTHORSHIP

CLAUDE S. BRINEGAR



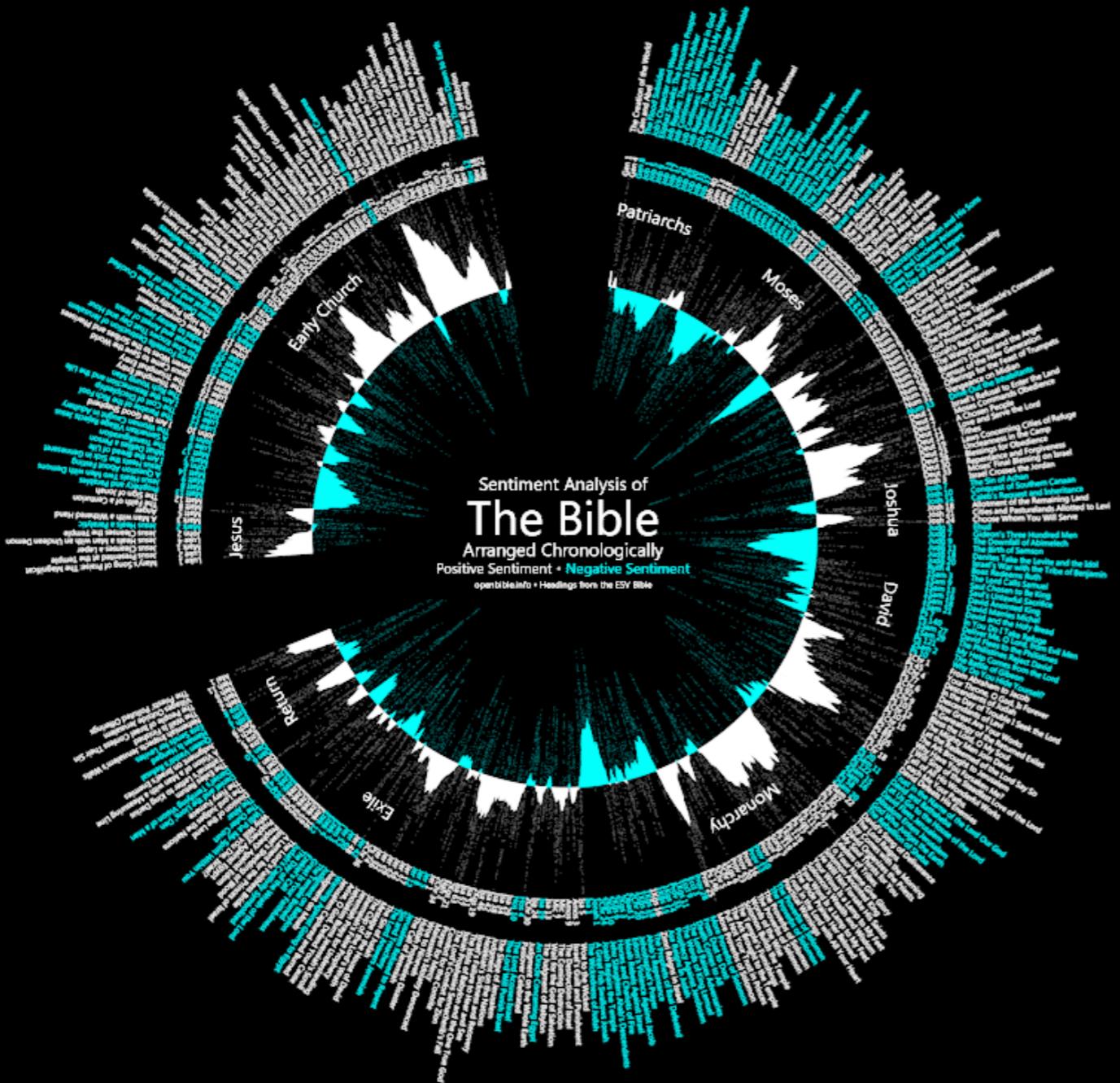
$$X^2 = 127.2, df=12$$

Relate quantities



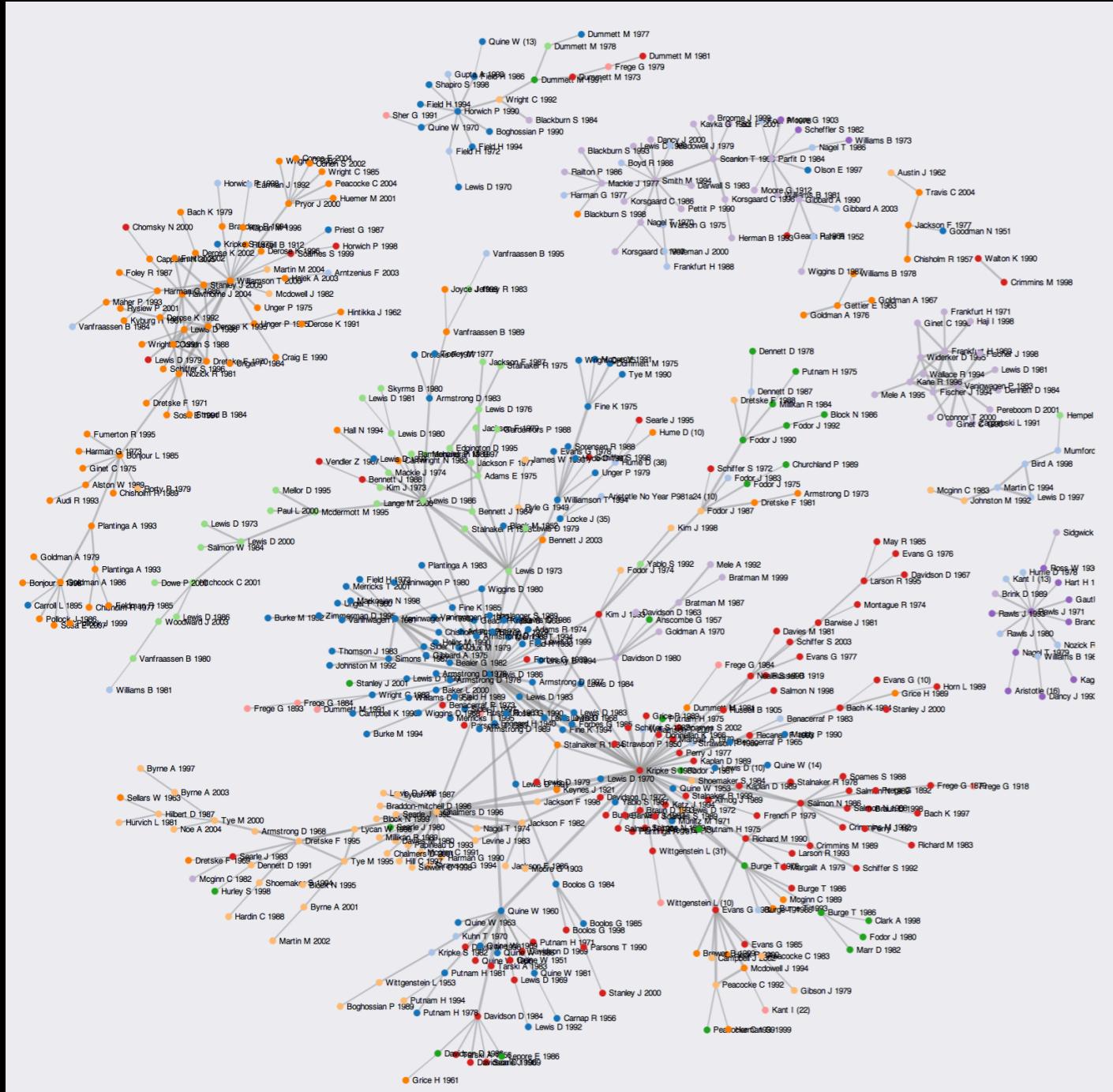
- **Age, Gender vs Probability of Death (Titanic)**
- **Sentence length vs word length**
- **Image entropy over time**

Recognize (some ML)



- Sentiment Analysis
- Object Recognition in images
- Clustering of like things

Network Analysis



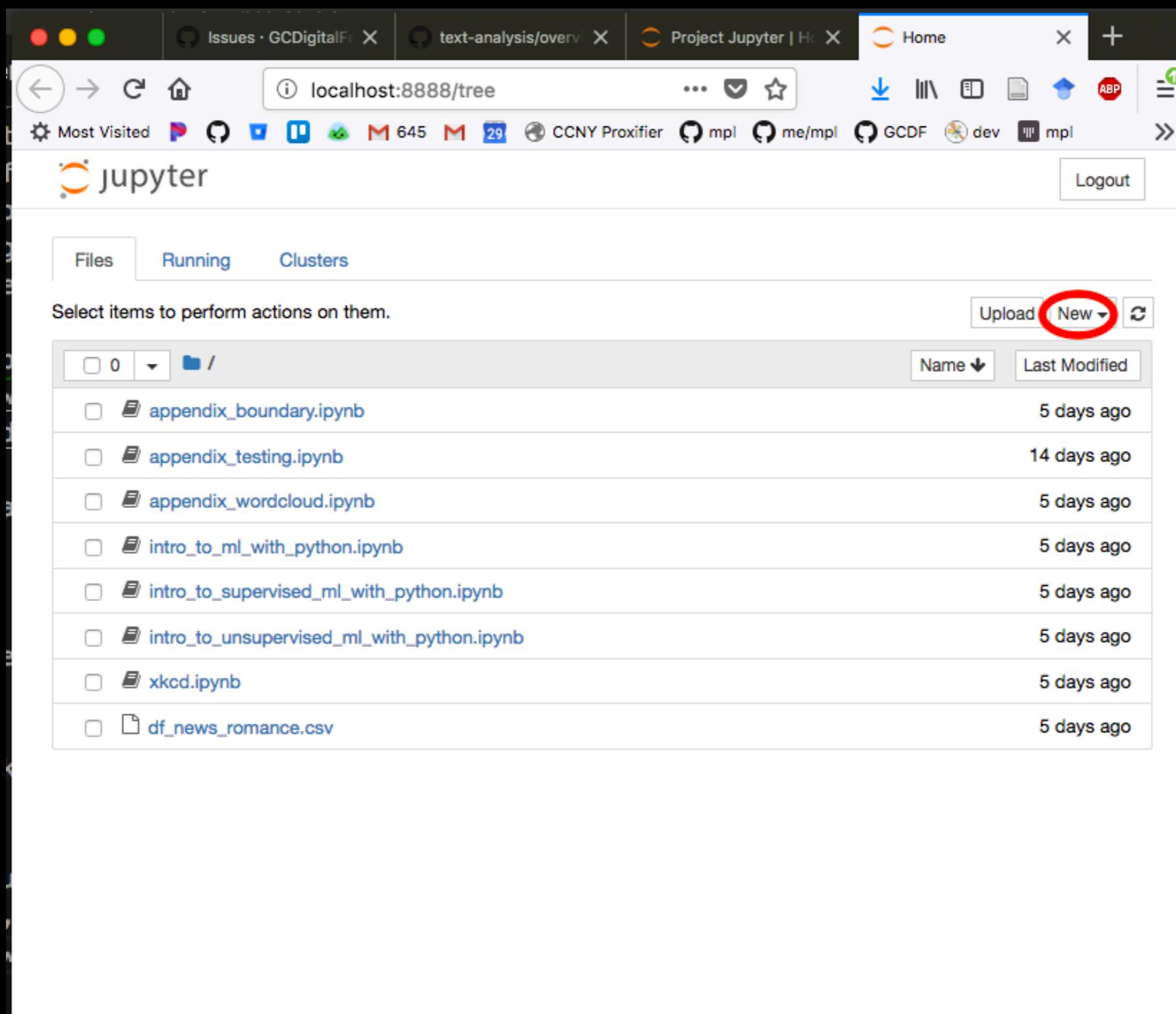
Philosophy Co-citation network

- Co-citation networks
- Co-occurrence of characters
- Similarity networks of actors by writing style
- Similarity of paintings by features

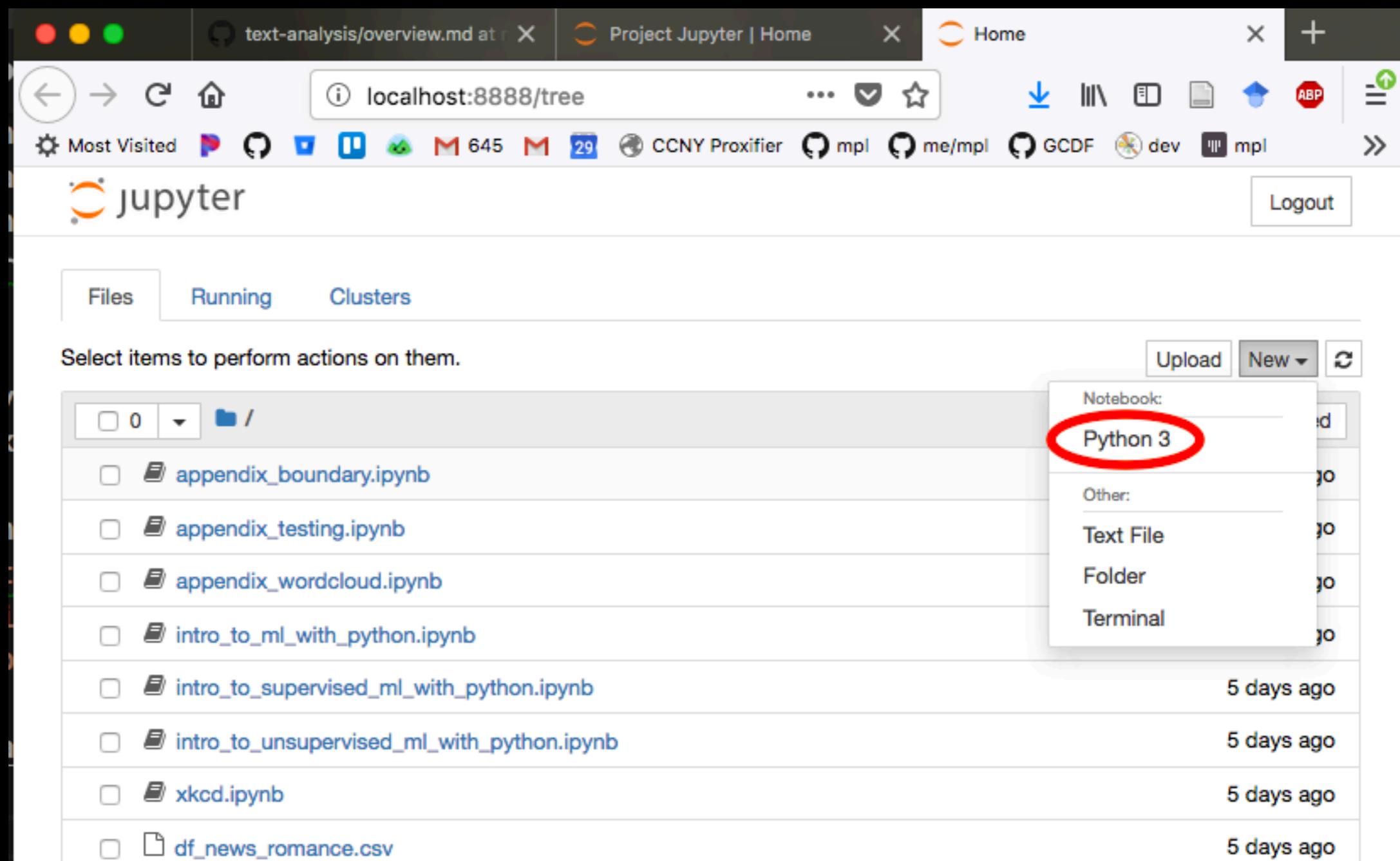
Tour of Notebook

[DHRI-Curriculum / install](#)
forked from [GCDigitalFellows/installdri.github.io](#)

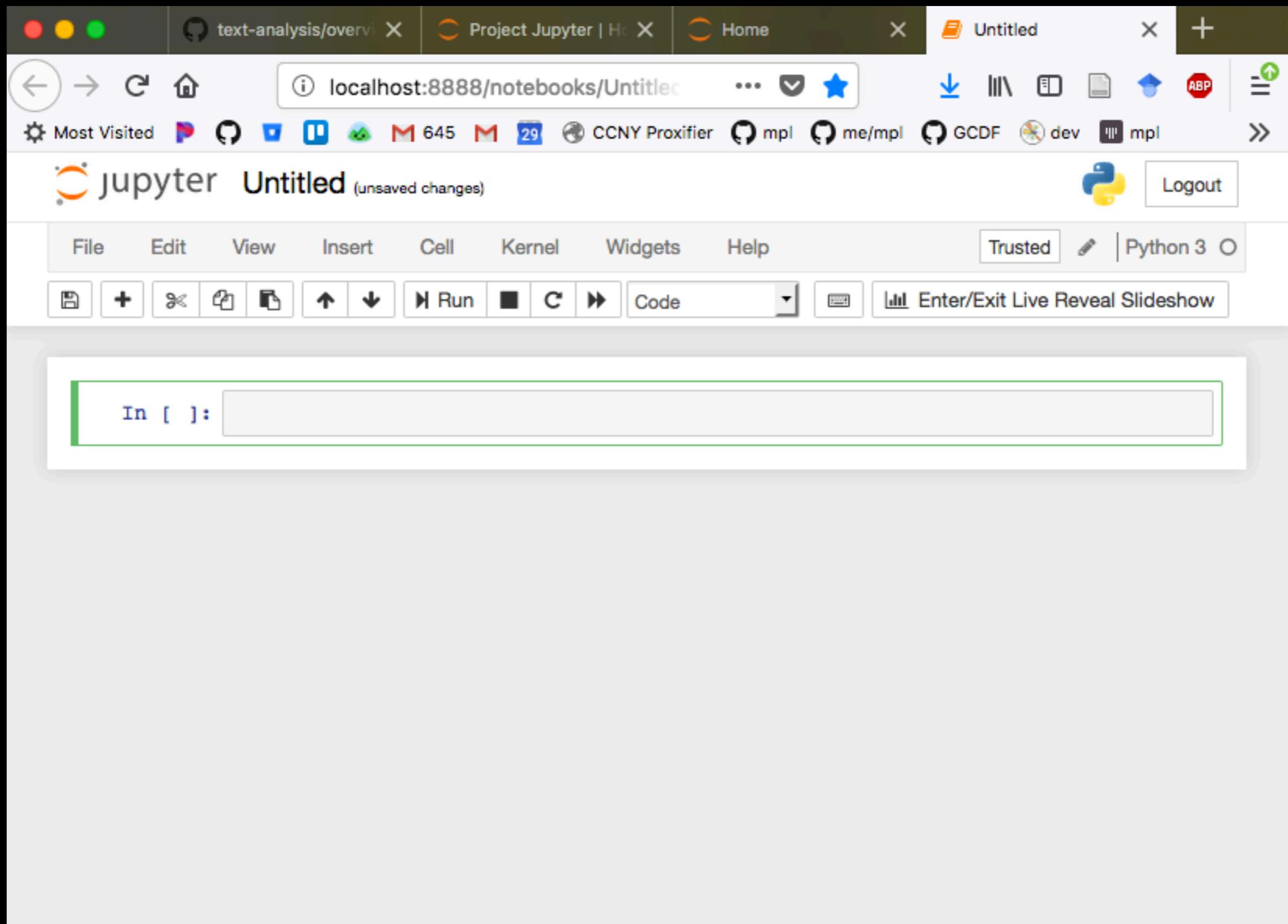
Jupyter Tree



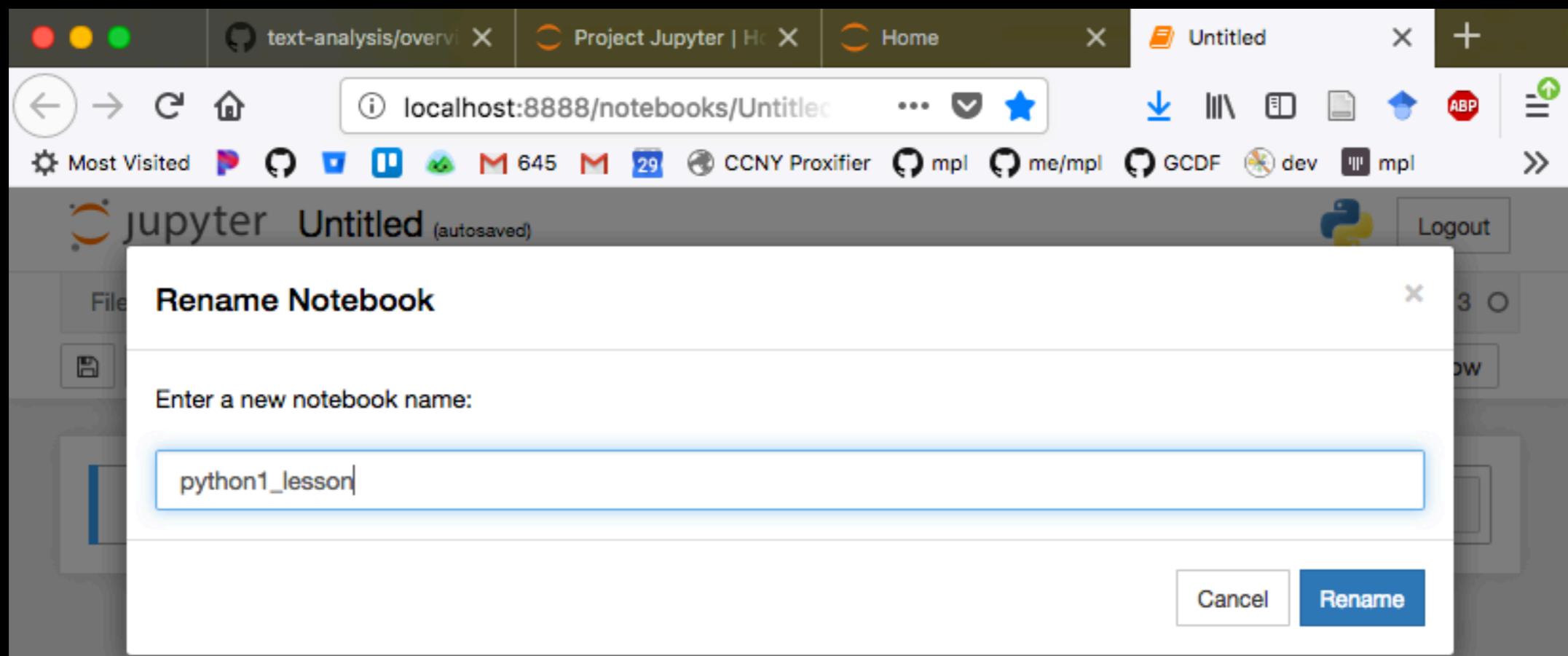
New Python 3



New Untitled Notebook

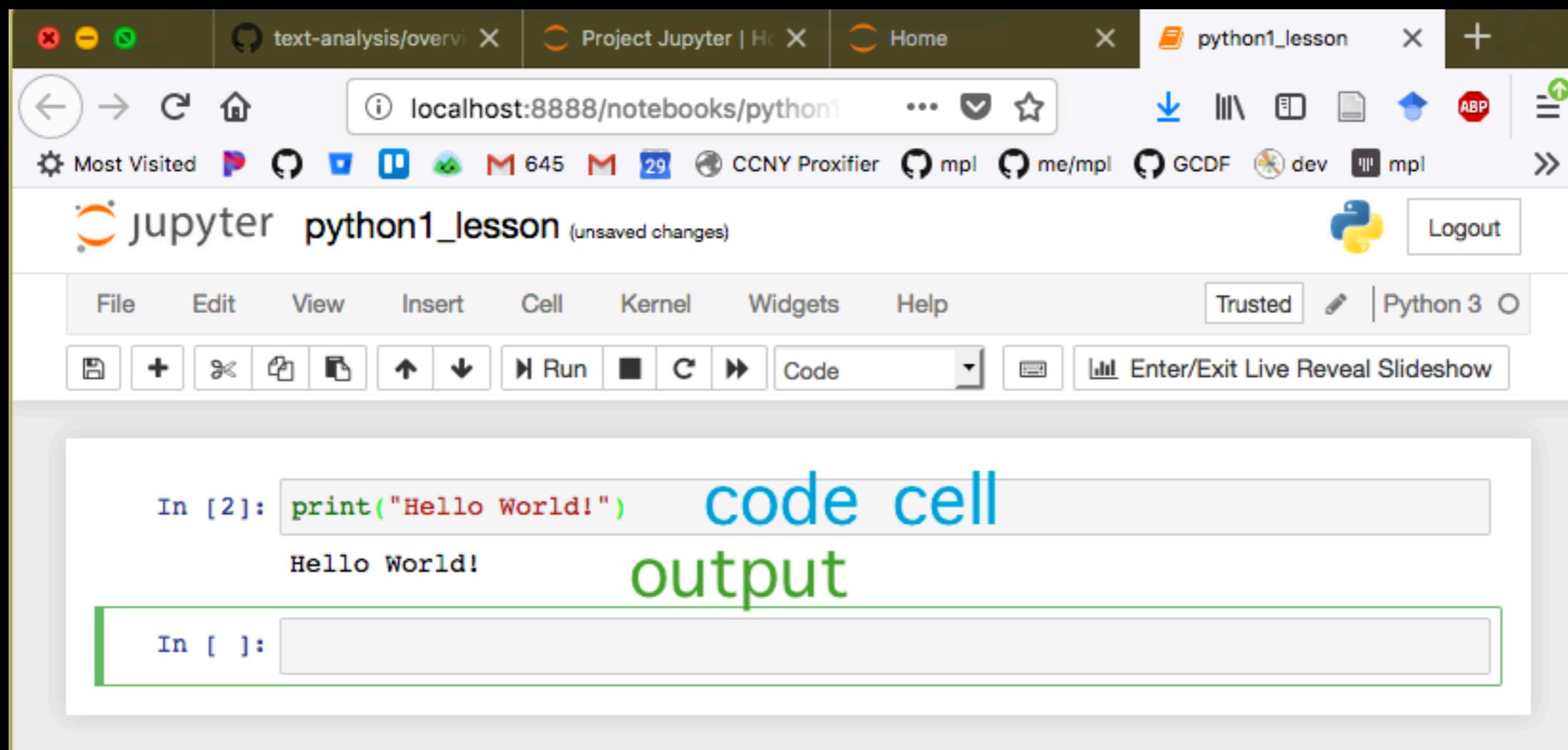


Rename



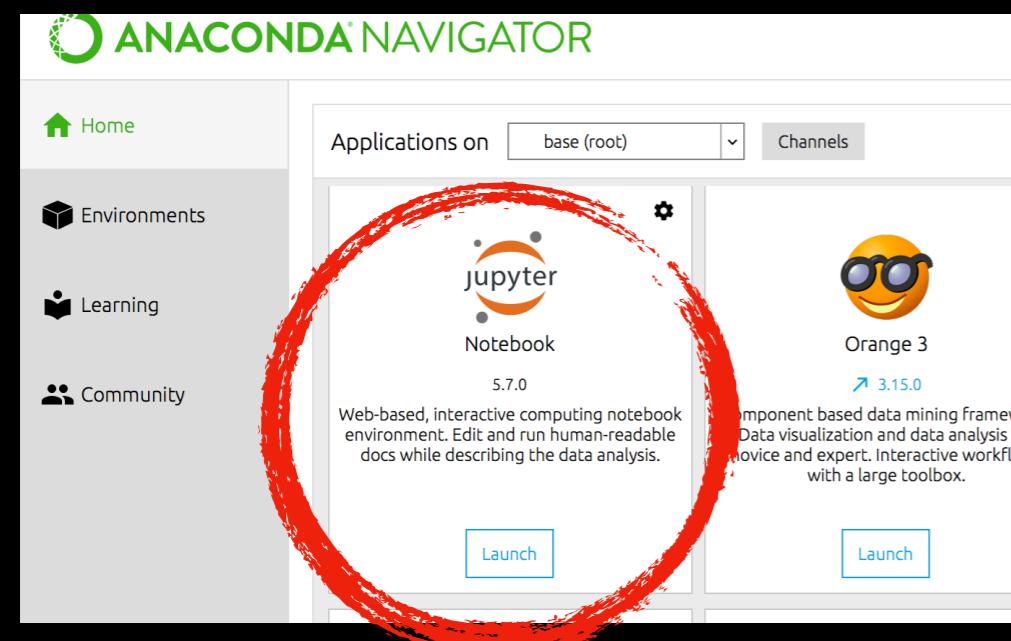
Evaluate-> See output

<shift> + <enter>



Lets Get Started!

While we talk



Install
Anaconda
Open Notebook

Use your account at

ccnyhub.org

OR

OR



colab.research.google.com

OR

