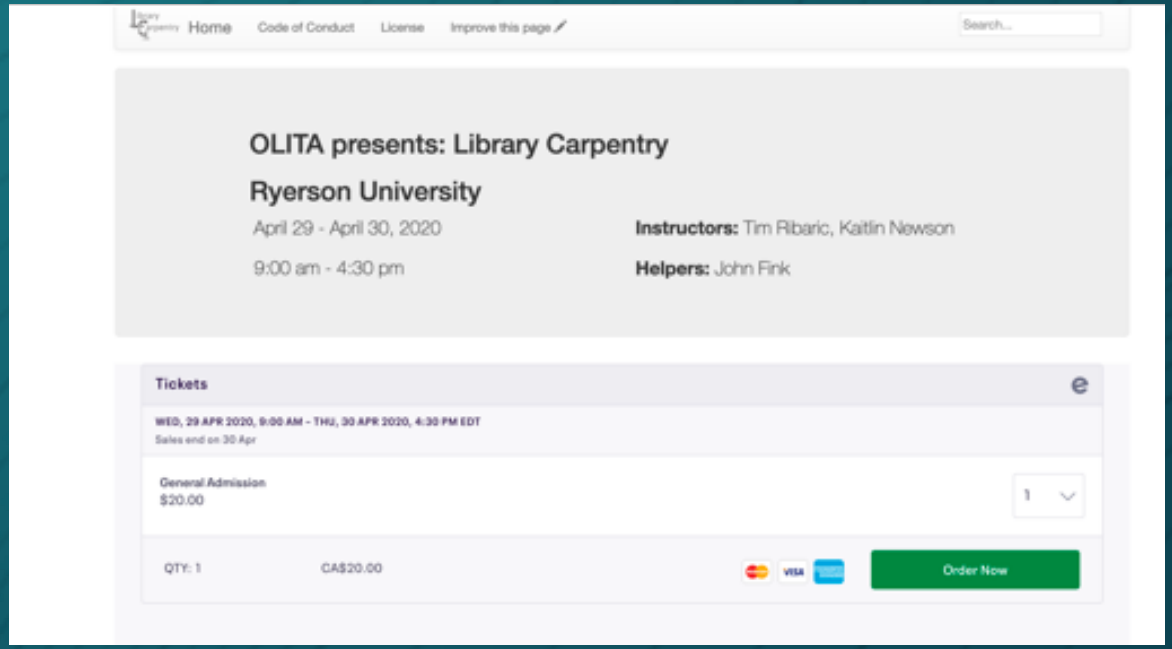
To Jupyter and Beyond

Presented by Tim Ribaric and Daniel Brett

First, a commercial!



Agenda

What is a computational notebook?

Typical Use Cases

How to get and use Jupyter?

Live Demo

Questions / Comments?

What is a Computational Notebook?

Have you heard this phrase before?

What a Computational Notebook Is

First introduced in 1988 with Wolfram Mathematica, followed closely by Maple in 1989

You can think of them as an analogue to the notebook a scientist would keep track of their experiments and finds with

How it works

It is a virtual notebook environment used for something referred to as **Literate programming**

Basically it pairs word processing with two things called a shell and a kernel

What it looks like

A webpage with a bunch of text boxes on it!

Text boxes are referred to as cells

Cell come in two types:

- Code
- Markdown

jupyter exploring data (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Markdown

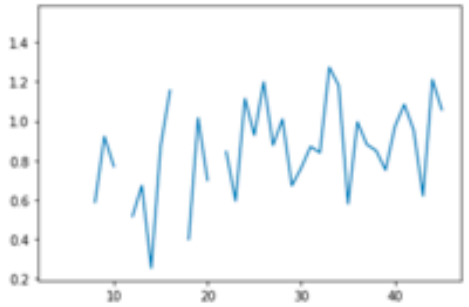
Exploring the Data

The following block loads the data

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

total_gi = float(open("alltime.txt","r").readline())

df = pd.read_csv("grief_series.csv",header=None,names=["Month", "GI"])
df["Month"] = pd.to_datetime(df["Month"])
```

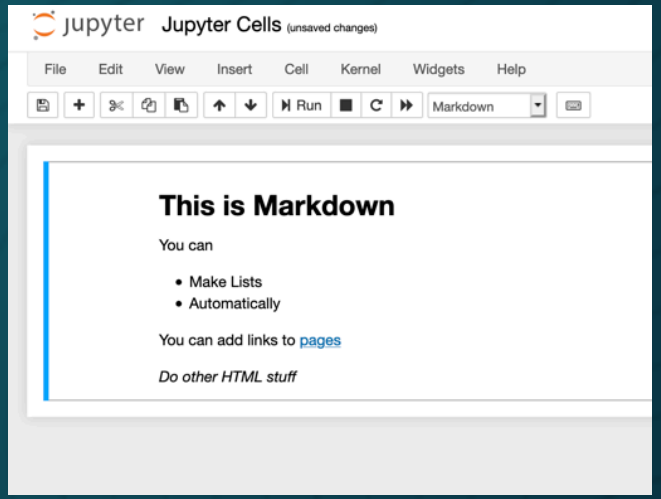
Let's now plot the data into a graph. More information here

In [9]:
```
data_series = list(np.log10(df["GI"]))

plt.plot(data_series)
plt.show()
```



Cells - Markdown

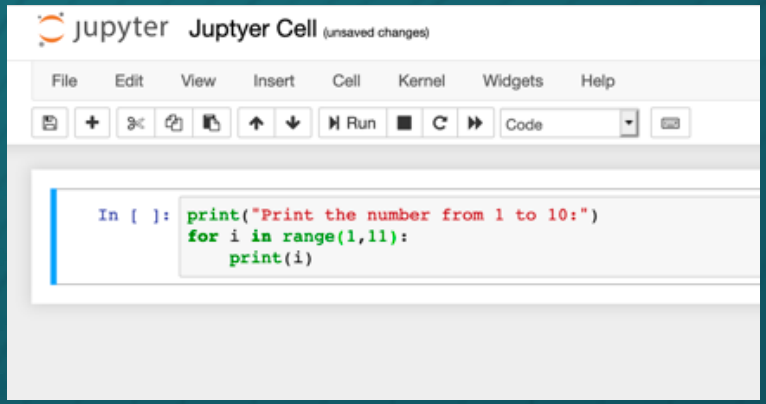
A syntax that makes HTML

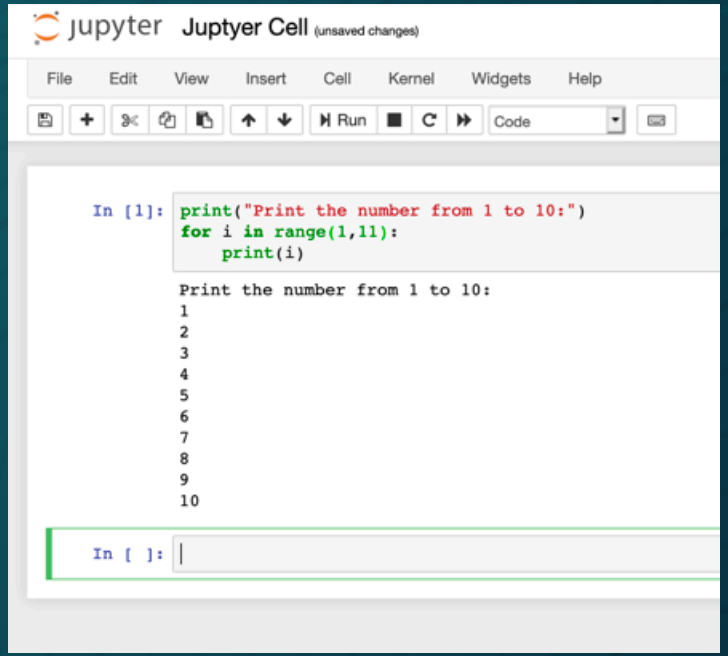


Cells - Code

This is the computer code

Simple!

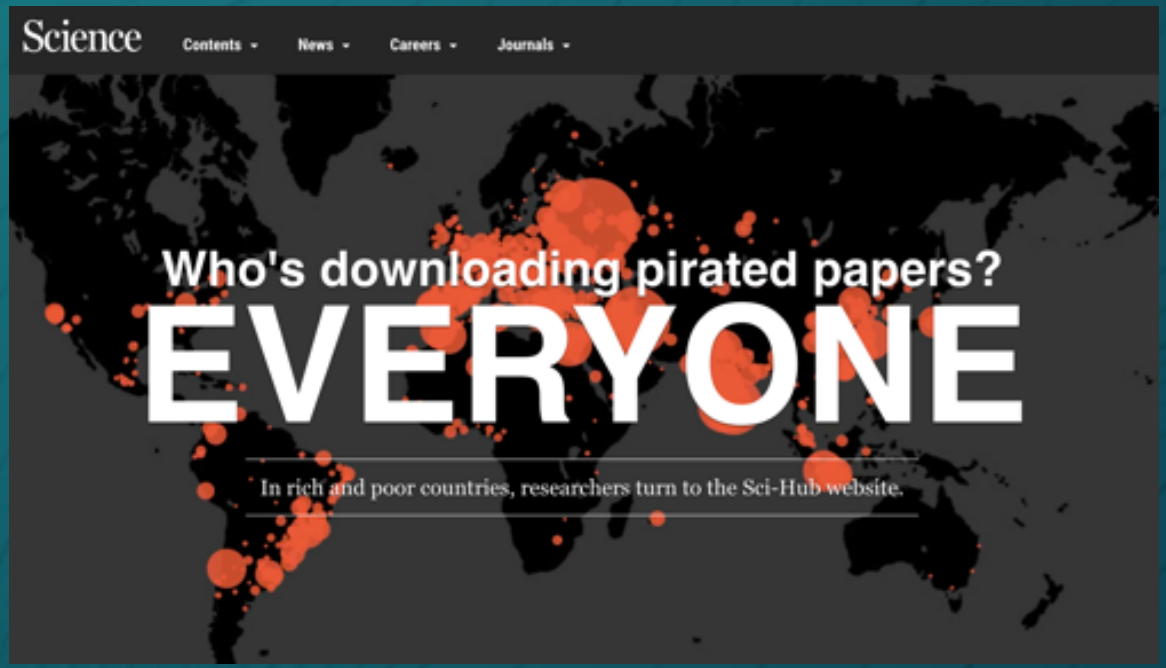




That's pretty much it!

How are they being used today?

Repeatable Research



SCI-HUB

Repeatable Research Cont.

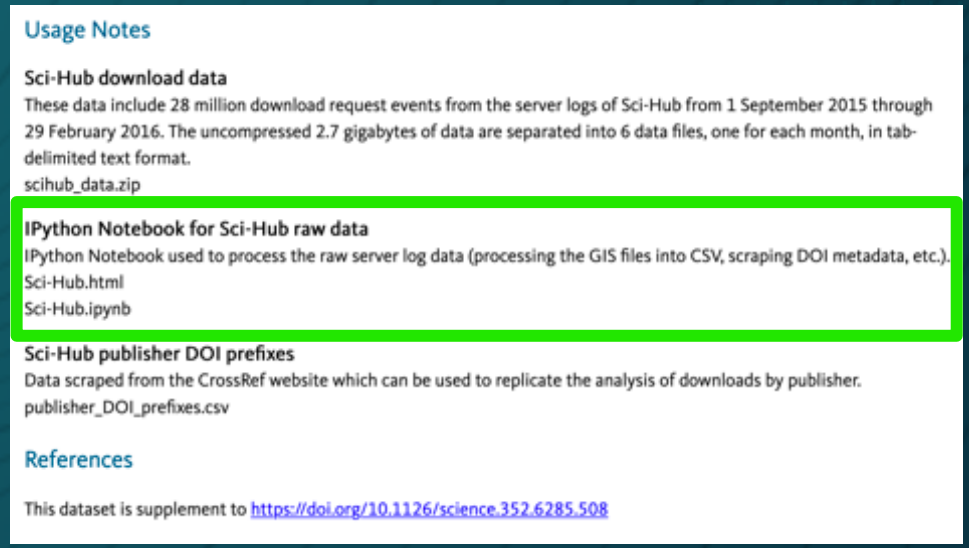
DRYAD

Data from: Who's downloading pirated papers? Everyone

Bohannon, John
Elbakyan, Alexandra
Publication date: April 22, 2017
Publisher: Dryad
https://doi.org/10.5061/dryad.q447c

Citation

Bohannon, John; Elbakyan, Alexandra (2017), Data from: Who's downloading pirated papers? Everyone, Dryad, Dataset, https://doi.org/10.5061/dryad.q447c



Usage Notes

Sci-Hub download data
These data include 28 million download request events from the server logs of Sci-Hub from 1 September 2015 through 29 February 2016. The uncompressed 2.7 gigabytes of data are separated into 6 data files, one for each month, in tab-delimited text format.
scihub_data.zip

IPython Notebook for Sci-Hub raw data
IPython Notebook used to process the raw server log data (processing the GIS files into CSV, scraping DOI metadata, etc.).
Sci-Hub.html
Sci-Hub.ipynb

Sci-Hub publisher DOI prefixes
Data scraped from the CrossRef website which can be used to replicate the analysis of downloads by publisher.
publisher_DOI_prefixes.csv

References

This dataset is supplement to https://doi.org/10.1126/science.352.6285.508

https://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone

In the DSL

TWARC tool to harvest tweets that include #datavisualization

Ran text processing on the data in a Jupyter Notebook using R

Finally, we fed the cleaned text to Voyant to produce a word cloud

Use for tutorials

Low overhead for users. (Nothing to install)

At the DSL we use Jupyter Notebooks to contain our tutorials that use Python or R code

Our Python 2.0 workshop comes with a Jupyter Notebook that learners can run through

https://brockdsl.github.io/Python_2.0_Workshop/

Brock University

Brock University Library
Digital Scholarship Lab

Netflix

Integral part of their Data Platform since 2017

“The most popular tool for working with data at Netflix” (Netflix TechBlog, 2018)

Use for creating reusable templates for frequent tasks

And much more!

https://netflixtechblog.com/notebook-innovation-591ee3221233

Kaggle

Great source of free datasets

Hub where Jupyter Notebooks containing all kinds of projects and code are shared freely

High profile coding competitions with Jupyter Notebooks as the preferred submission method.

Competitions range from Passenger Screening Algorithm Challenges to solving Santa's workshop efficiency problems

How do we get to Jupyter?

My Binder

A tool that turns a repository of Jupyter Notebooks on GitHub into a clean, web accessible interface.

Allows you to share Jupyter Notebooks with others without them needing Jupyter on their computers.

https://mybinder.org/

ArcGIS Notebooks

Allows you to build and run Notebooks that are connected to all of you Arc resources and your ArcGIS Online storage space.

Gives access to repeatable ArcGIS functions and mapping tools from within the notebook.

https://www.esri.com/arcgis-blog/products/arcgis-enterprise/analytics/introducing-arcgis-notebooks/

Syzygy – Via Compute Canada

A Jupyter environment accessible on the web that uses your existing single-sign-on system to allow access

Hosted by Compute Canada and Cybera

Only for Universities unfortunately

https://syzygy.ca/

Desktop Version via Anaconda

A Desktop version of Jupyter is available through a free distribution software called Anaconda

Package of statistical and mathematical tools useful for many things

All free

https://www.anaconda.com

Live Demo

Questions / Comments?

Credits

dsl@brocku.ca

http://brocku.ca/library/dsl

Presentation template by SlidesCarnival

Tim Ribaric

tribaric@brocku.ca

@elibtronic

Daniel Brett

dbrett@brocku.ca