SOCIAL MEDIA ANALYTICS DURING CRISIS — SESSION I

DATA ACQUISITION AND SIMPLE ANALYSIS

AGENDA

SESSION I
DATA ACQUISITION
1:15PM-3:00PM

1. Jupyter Notebooks
2. Social media data sources
3. Parsing Twitter data
4. Frequency analysis

COFFEE BREAK
3:00PM-3:20PM

SESSION II
ADVANCED ANALYSIS
3:20PM-4:30PM

5. Location analysis
6. Sentiment analysis
7. Other content analysis
8. Network analysis

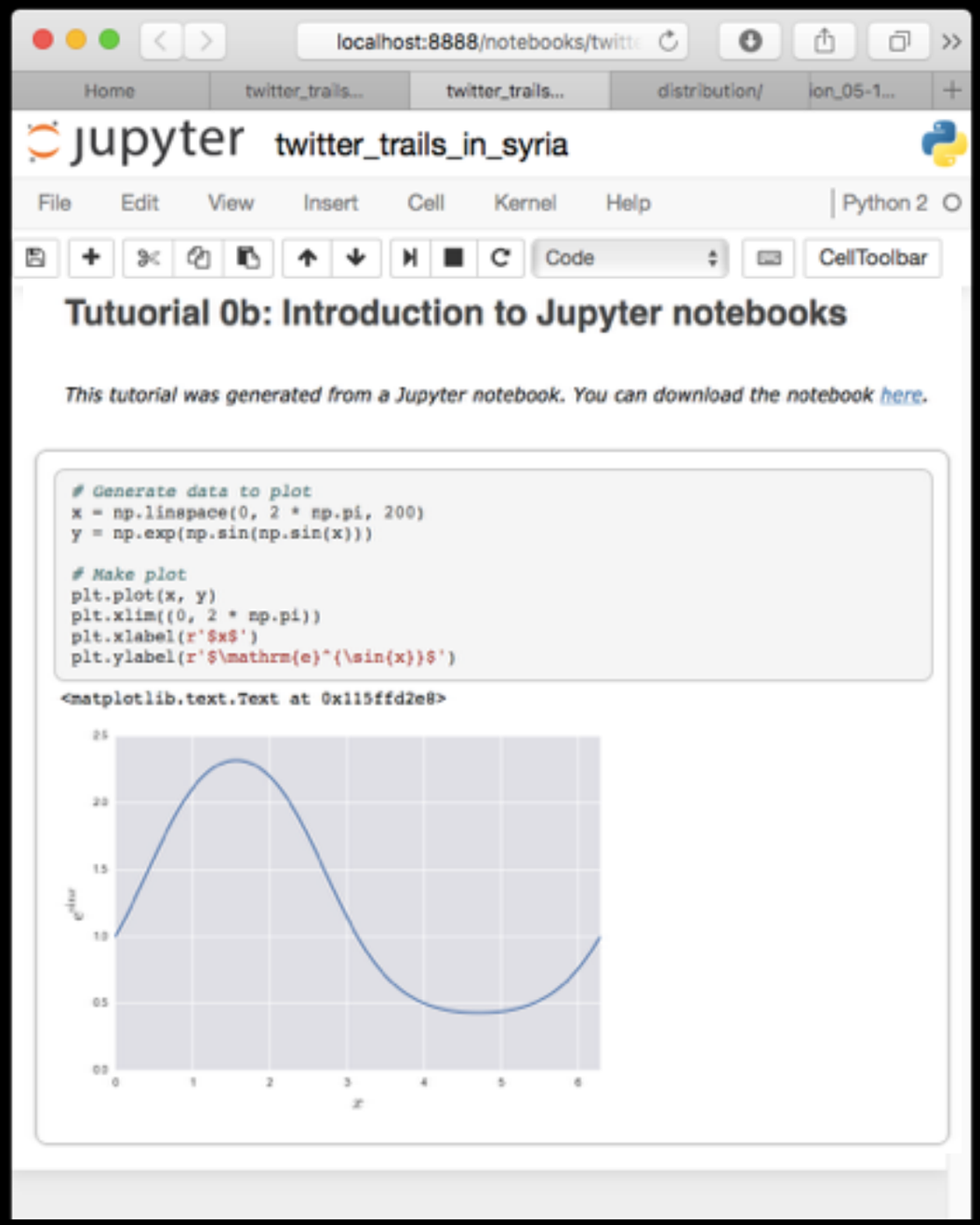
JUPYTER NOTEBOOKS

JUPYTER NOTEBOOKS

WHAT IS THE JUPYTER NOTEBOOK?

An interactive , browser-based development environment for scientific computer and data science.

Meant to be language-agnostic and support reproducible research.



JUPYTER NOTEBOOKS

WHAT IS THE JUPYTER NOTEBOOK?

An interactive , browser-based development environment for scientific computer and data science.

Meant to be language-agnostic and support reproducible research.

JUPYTER NOTEBOOKS

CODE

jupyter twitter_trails_in_syria (autosaved)

File Edit View Insert Cell Kernel Help Python 2

Code CellToolbar

```
for (user, locs) in userMap.items():
    if user not in plottableUsers:
        continue

    lats = map(lambda x: x[1], locs)
    lons = map(lambda x: x[0], locs)

    (x, y) = bmap(lons, lats)

    userCount += 1

    style = "%ss-" % (colorList[userCount % len(colorList)])
    bmap.plot(x, y, style, linewidth=2.0, markersize=20)

plt.show()
```

RESULTS

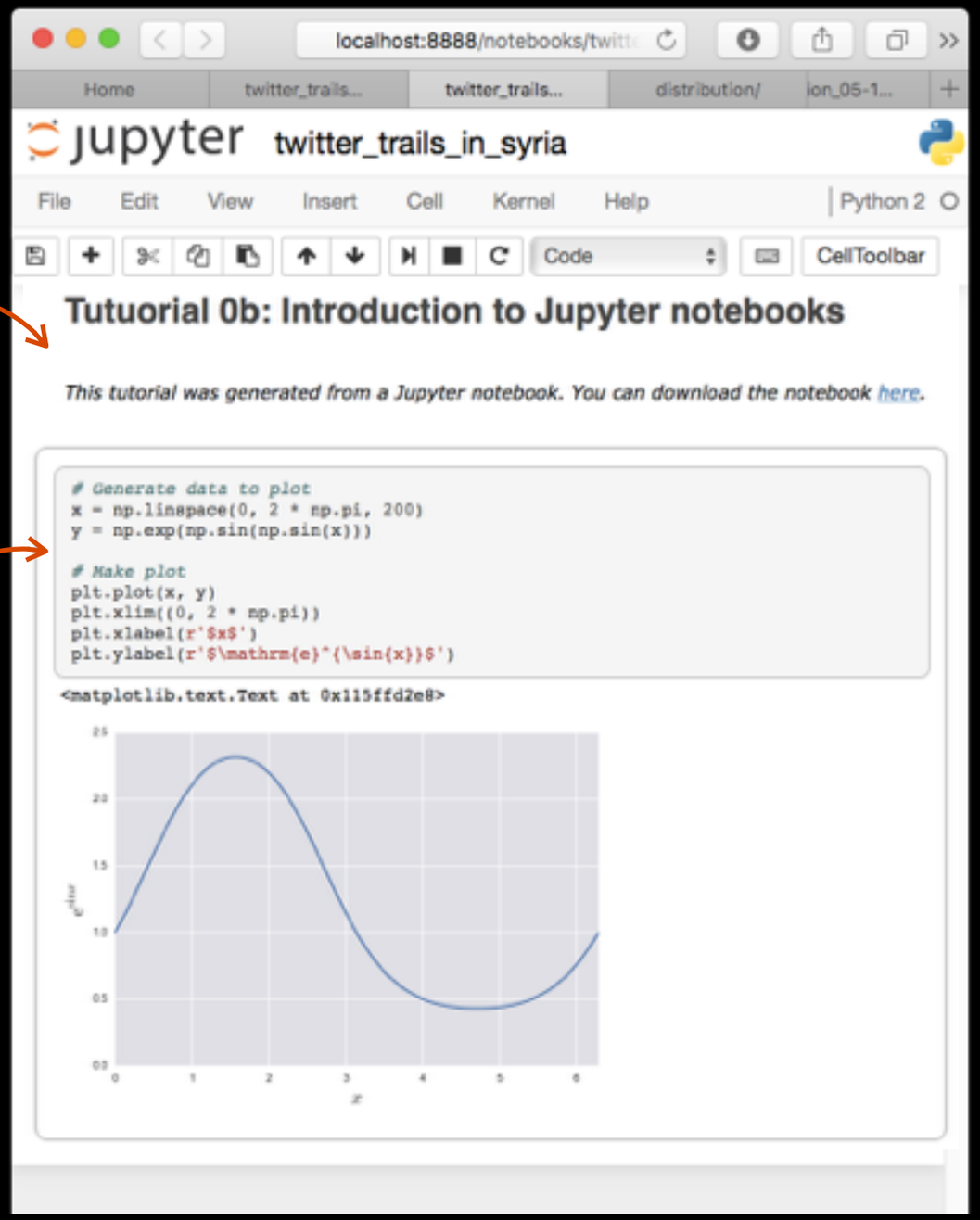
JUPYTER NOTEBOOKS

WHAT IS THE JUPYTER NOTEBOOK?

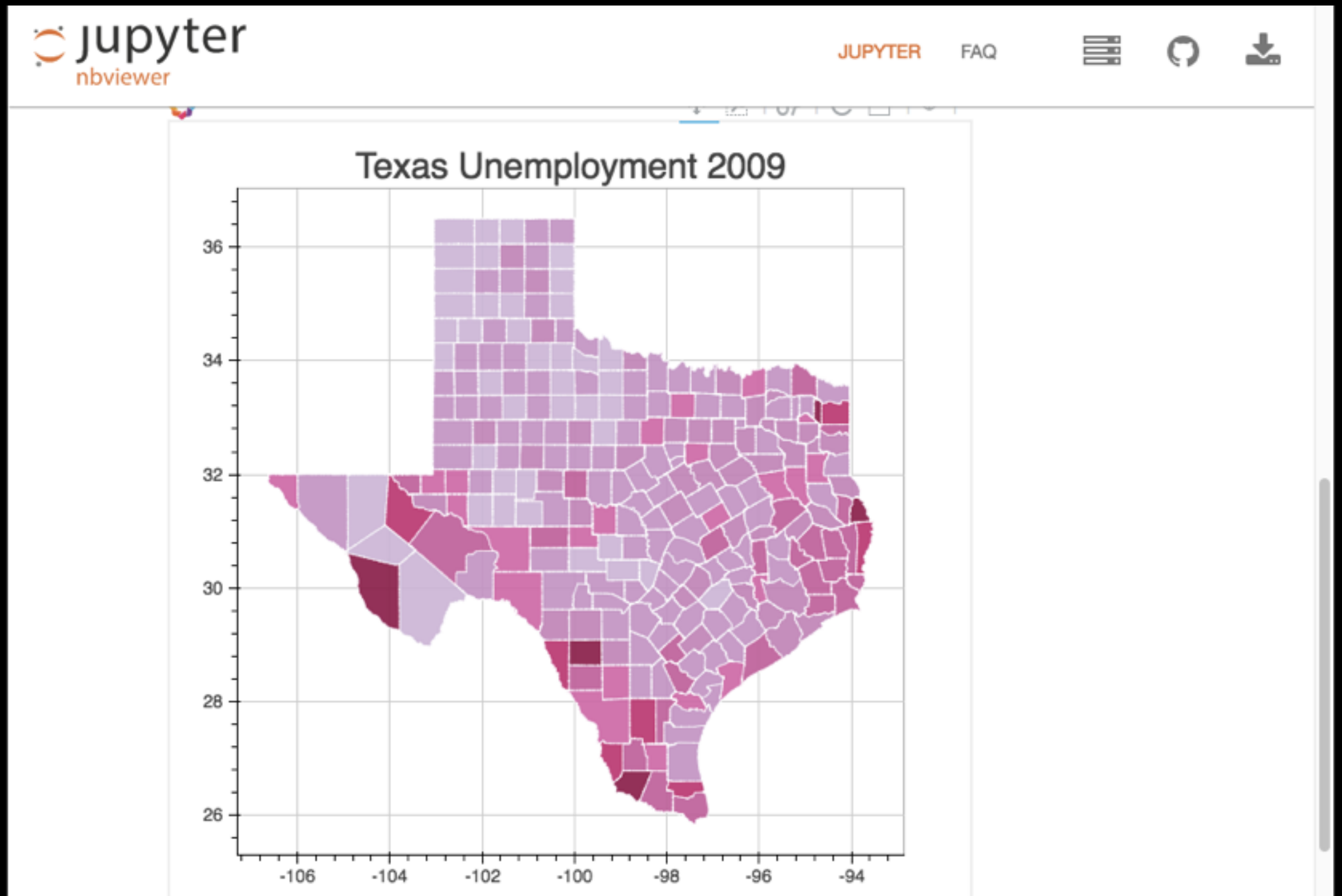
Notebooks are divided into cells:

Markup cells

Code cells



JUPYTER NOTEBOOKS



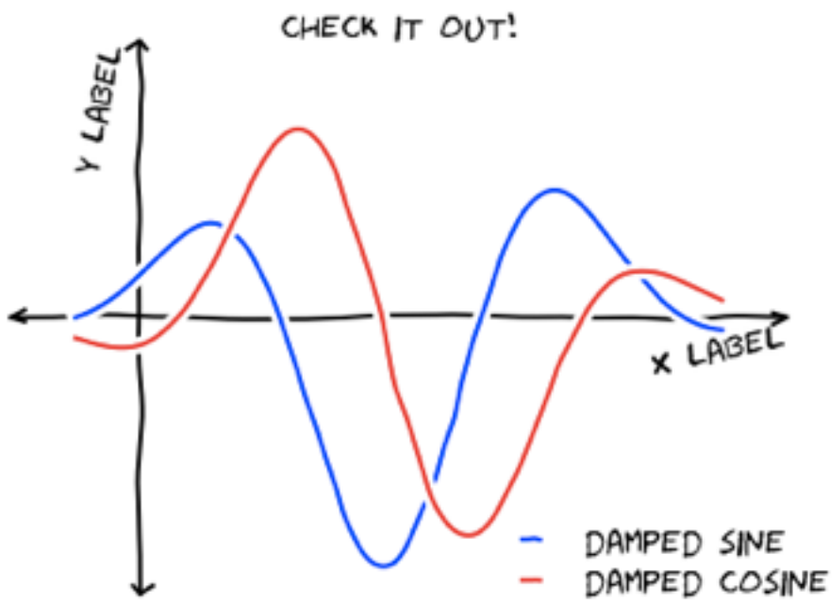
JUPYTER NOTEBOOKS

jupyter nbviewer

JUPYTER FAQ

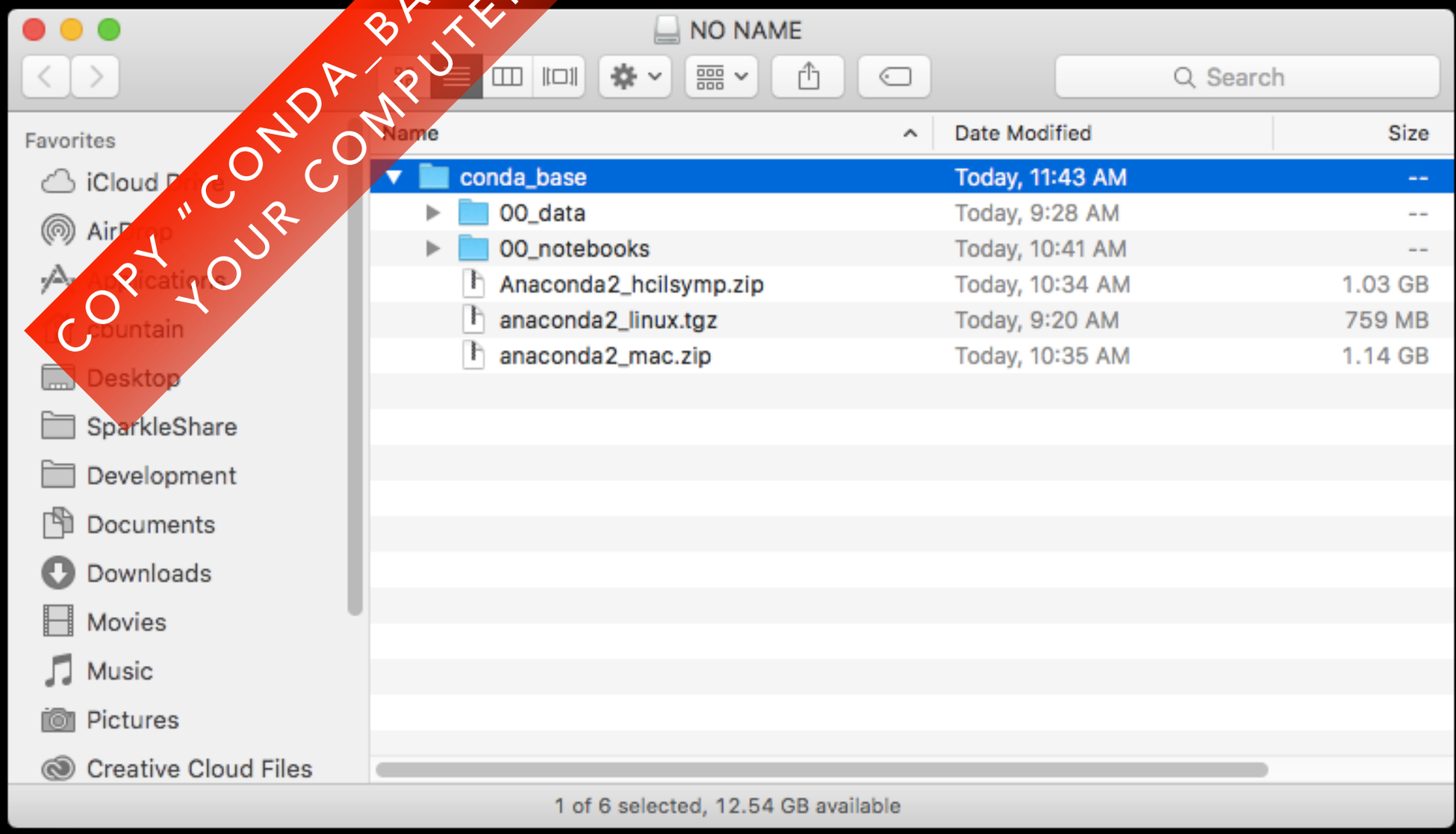
```
ax.plot(x, -np.cos(x) * np.exp(-0.1 * (x - 5) ** 2), 'r', lw=1, label='damped cosine')

ax.set_title('check it out!')
ax.set_xlabel('x label')
ax.set_ylabel('y label')

ax.legend(loc='lower right')

ax.set_xlim(0, 10)
ax.set_ylim(-1.0, 1.0)

#XKCDify the axes -- this operates in-place
XKCDify(ax, xaxis_loc=0.0, yaxis_loc=1.0,
        xaxis_arrow='+-', yaxis_arrow='+-',
        expand_axes=True)
```

Out[5]: <matplotlib.axes.AxesSubplot at 0x2fecbd0>



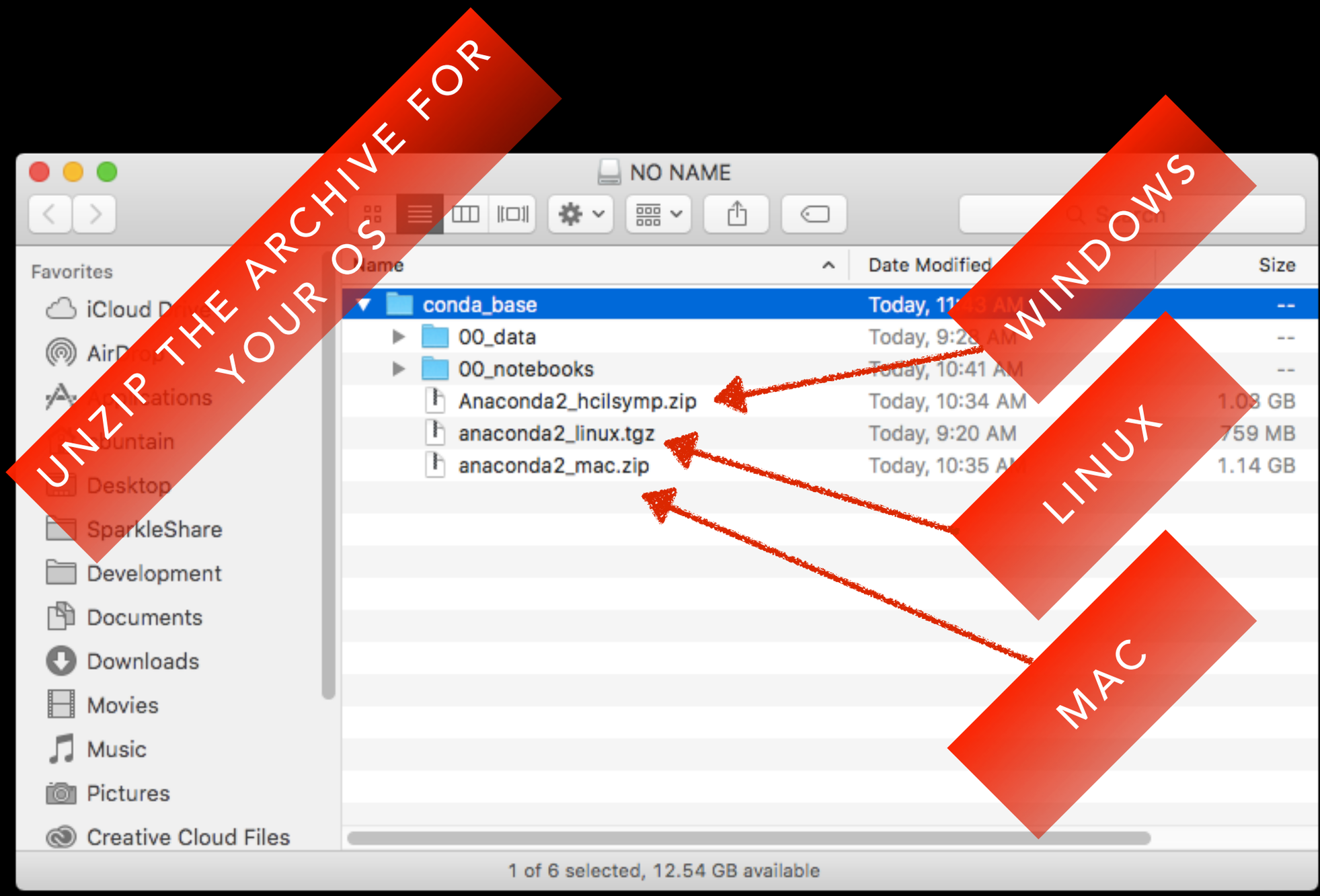
JUPYTER NOTEBOOKS

TUTORIAL FILES



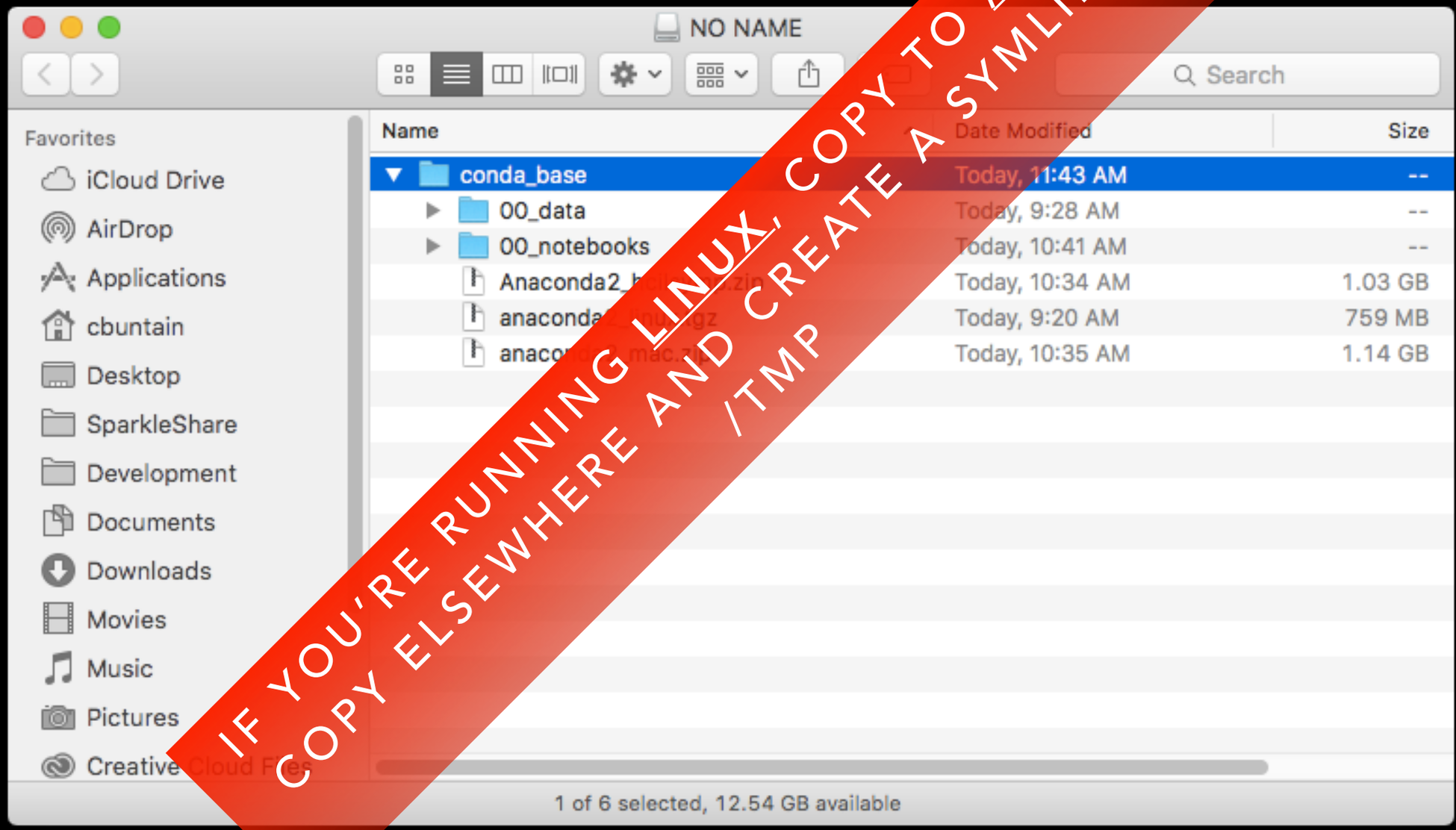
JUPYTER NOTEBOOKS

TUTORIAL FILES



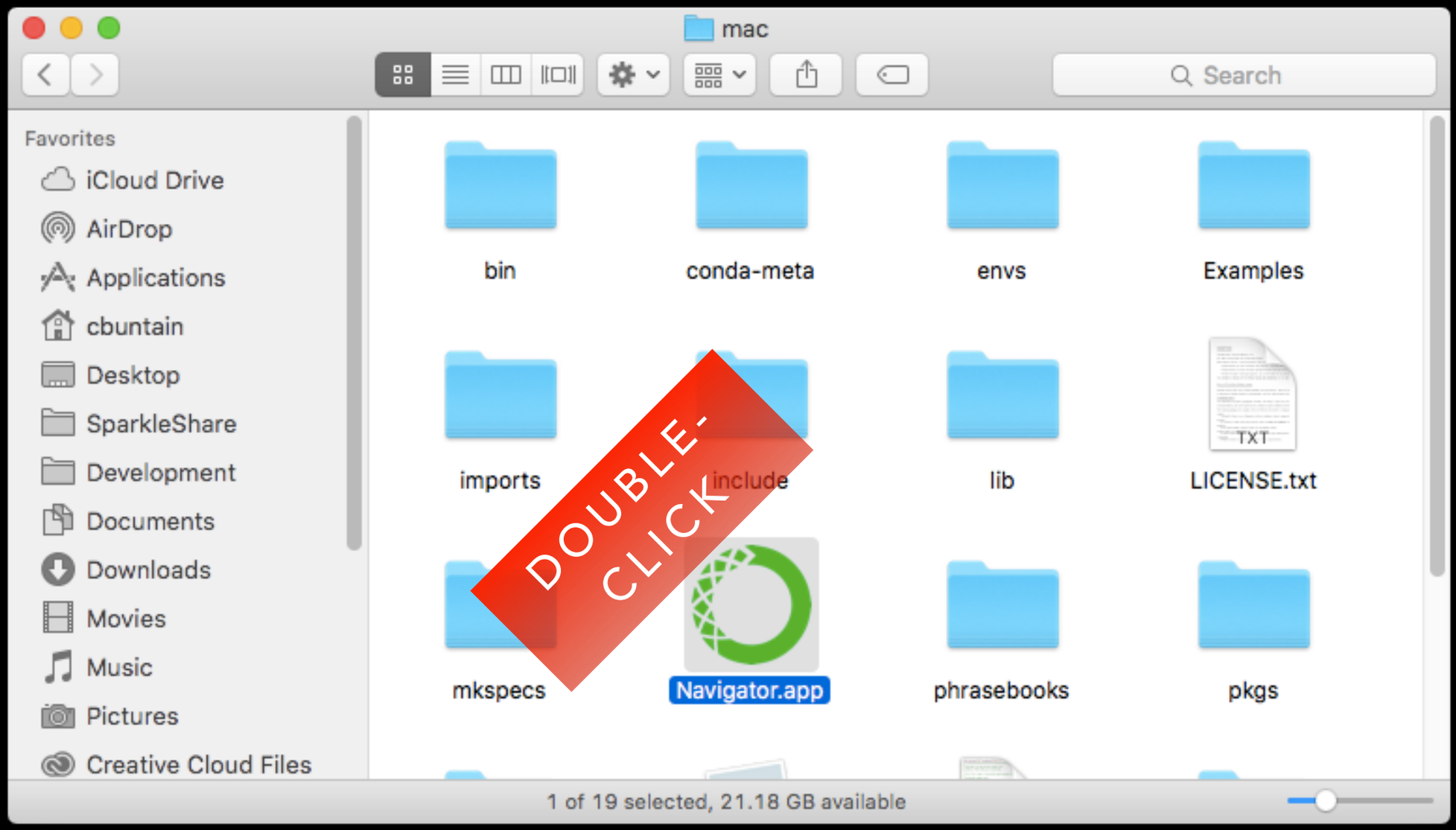
JUPYTER NOTEBOOKS

TUTORIAL FILES



JUPYTER NOTEBOOKS

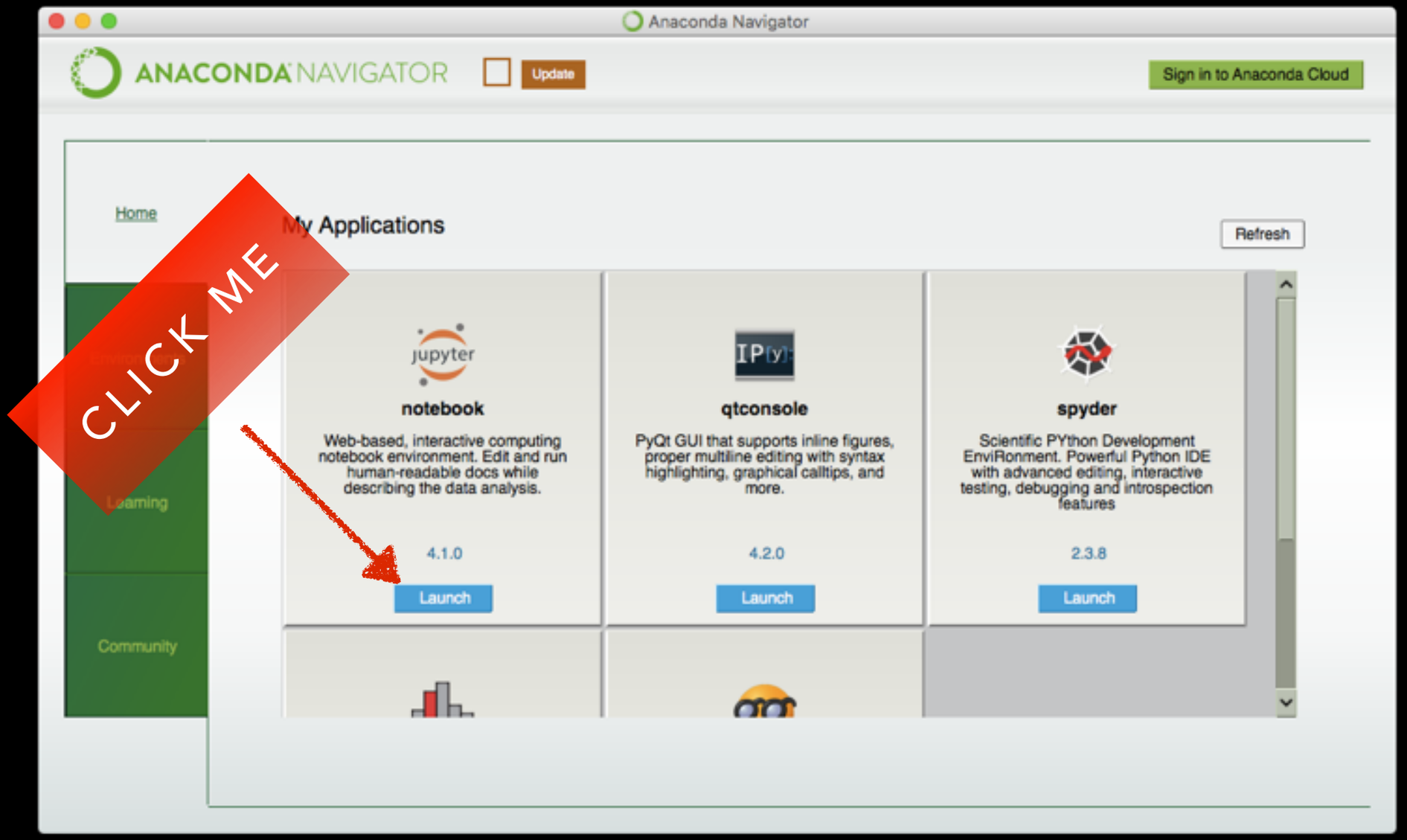
STARTING THE NOTEBOOK



Mac OSX

JUPYTER NOTEBOOKS

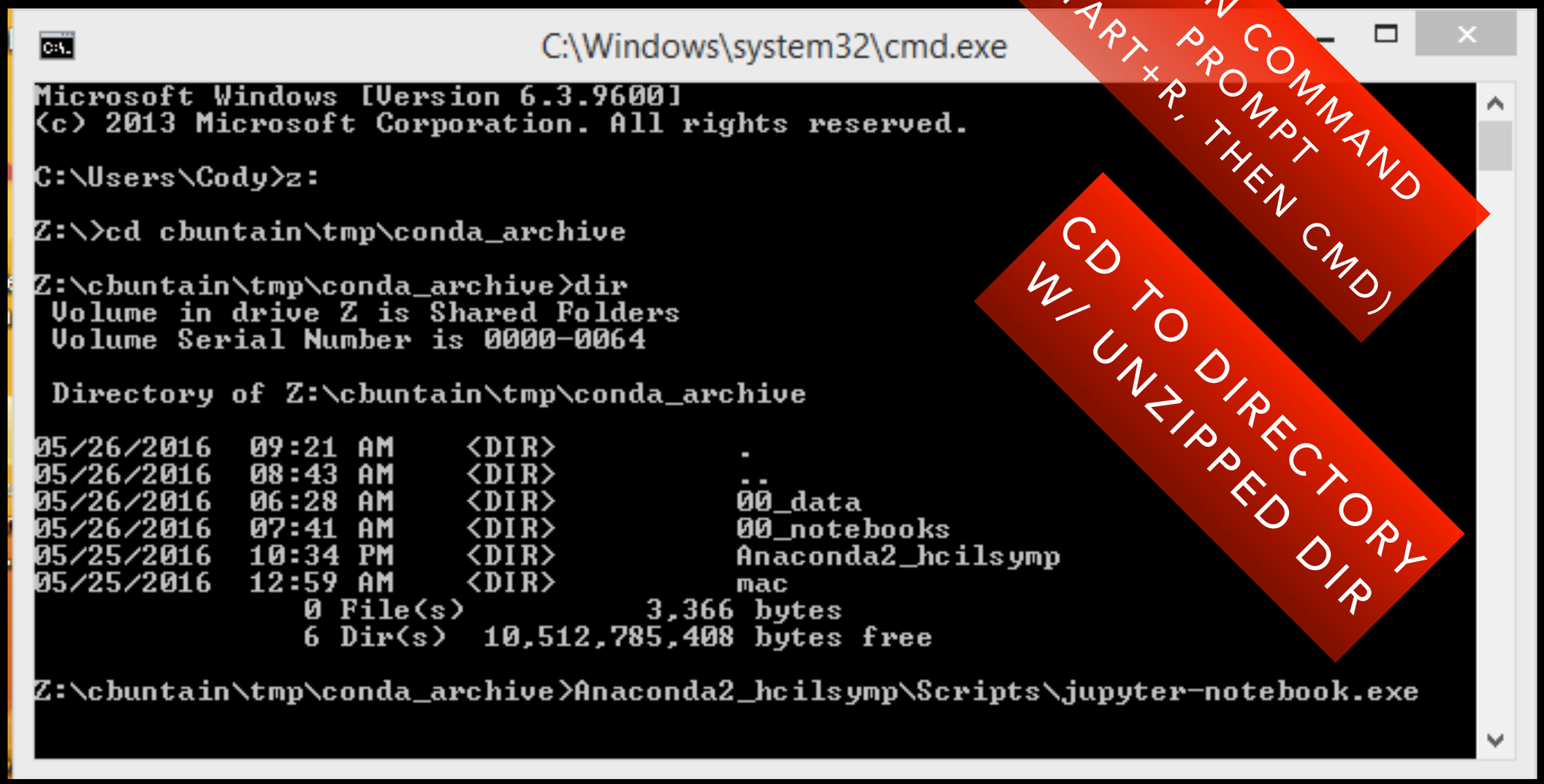
STARTING THE NOTEBOOK



Mac OSX

JUPYTER NOTEBOOKS

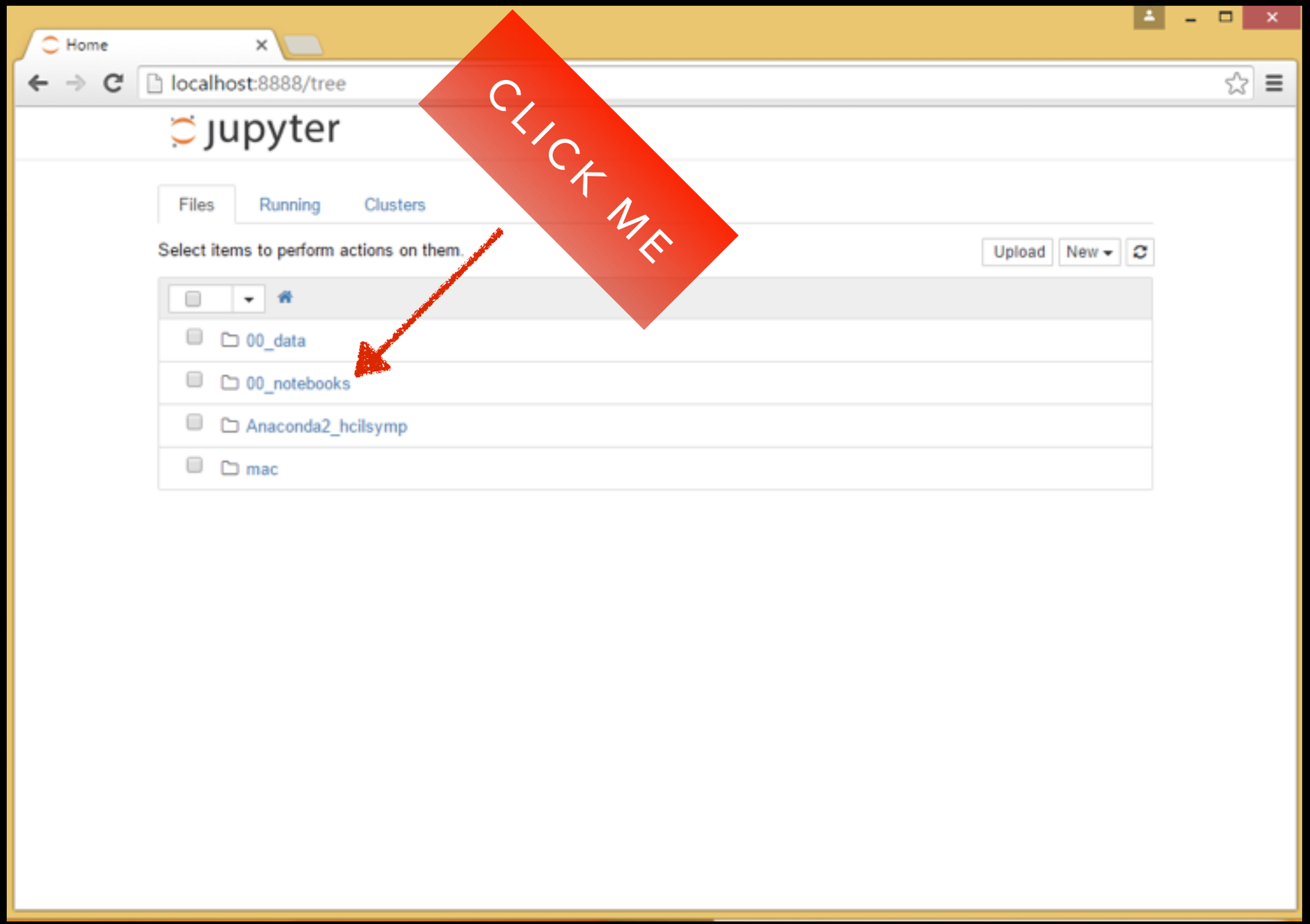
STARTING THE NOTEBOOK



Windows

JUPYTER NOTEBOOKS

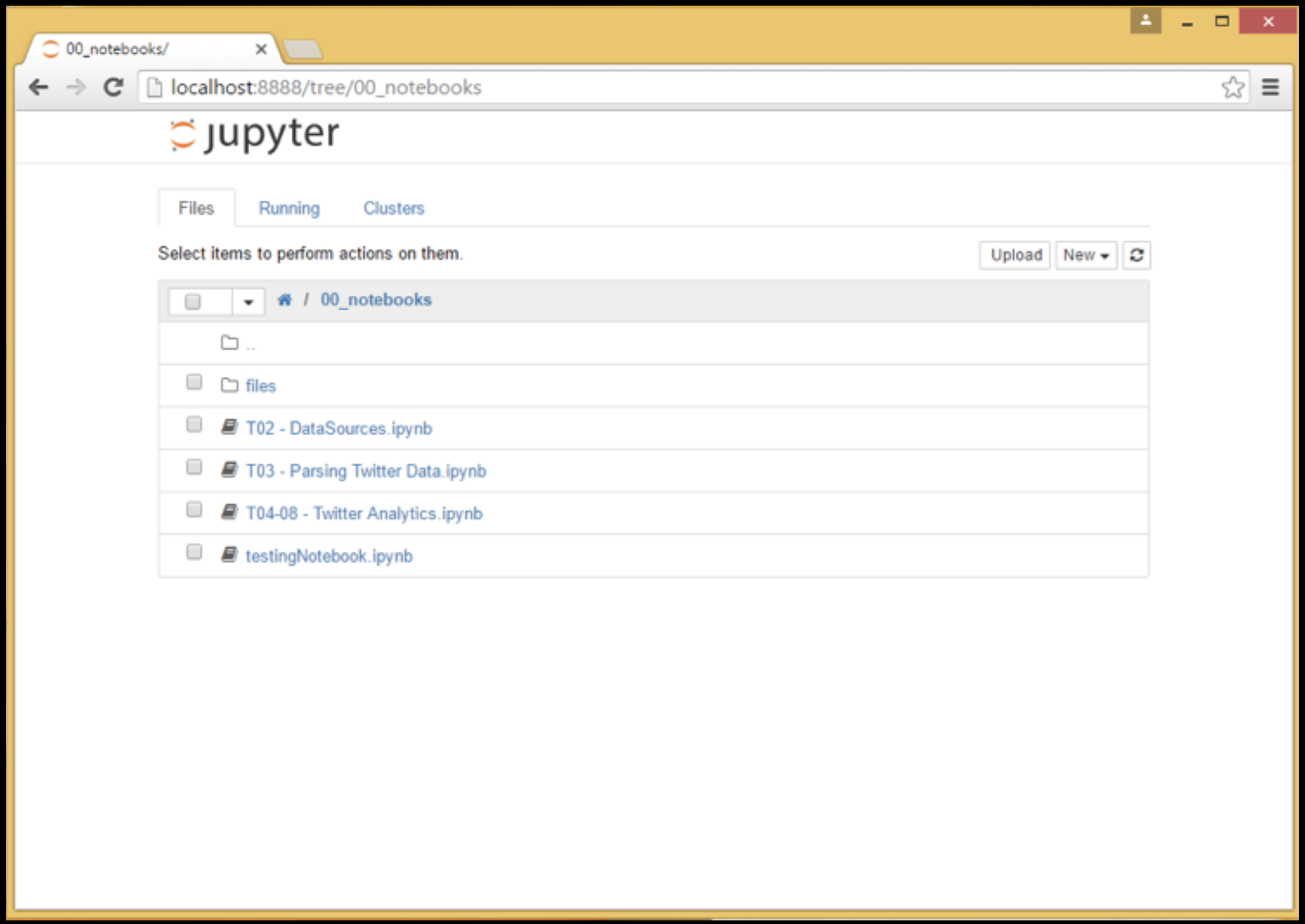
STARTING THE NOTEBOOK



Windows

JUPYTER NOTEBOOKS

STARTING THE NOTEBOOK



Windows

JUPYTER NOTEBOOKS

STARTING THE NOTEBOOK

.-(/tmp/conda_base)——————————————————————————(user@system)-
`--> **/tmp/conda_base/anaconda2_linux/bin/jupyter-notebook**

—OR—

bash$ **cd /tmp/conda_base/**
bash$ **/tmp/conda_base/anaconda2_linux/bin/jupyter-notebook**

Linux

DATA SOURCES

DATA ARCHIVES

https://archive.org/details/twitterstream

Goes back to 2011

~40 gigabytes per month of tweets

DATA ARCHIVES

https://archive.org/details/

2015_reddit_comments_corpus

Jason Baumgartner's reddit archive

October 2007 to May 2015

Contains ~1.65 billion comments

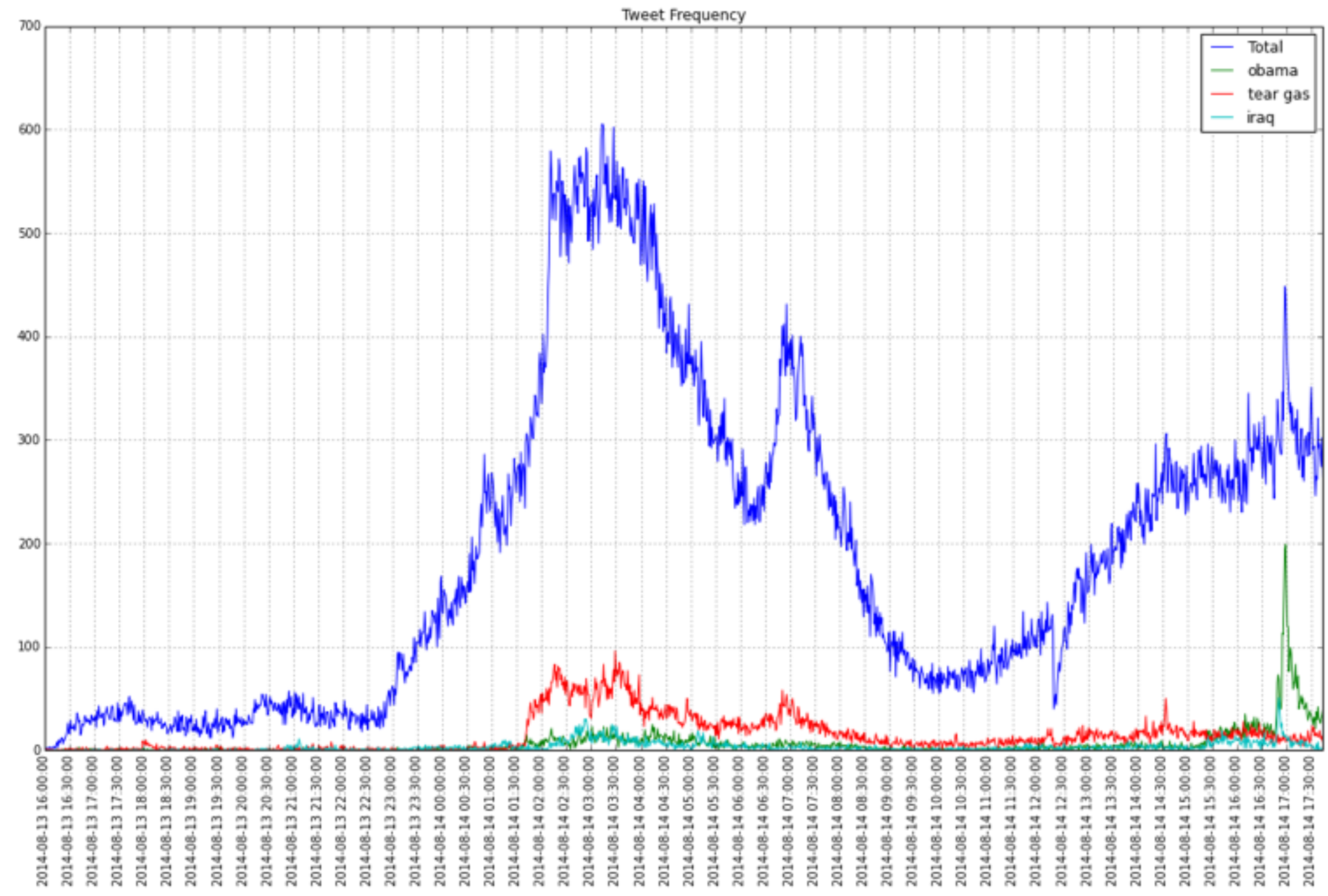
DATA ARCHIVES

Few Facebook data archives
that are publicly available

PARSING TWITTER DATA


```
{
  "contributors": null,
  "truncated": false,
  "text": "@megturney After seeing your tweets about #Ferguson I decided to look it up. The situation is horrible!",
  "in_reply_to_status_id": null,
  "id": 499922742045868000,
  "favorite_count": 0,
  "source": "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>",
  "retweeted": false,
  "coordinates": {
    "type": "Point",
    "coordinates": [
      -79.11194229,
      43.07537783
    ]
  },
  "entities": {
    "symbols": [],
    "user_mentions": [
      {
        "indices": [
          0,
          10
        ],
        "screen_name": "megturney",
        "id": 23711472,
        "name": "Meg Turney",
        "id_str": "23711472"
      }
    ],
    "hashtags": [
      {
        "indices": [
          42,
          51
```

FREQUENCY ANALYSIS




```
Top Twenty Hashtags:
        #ferguson 209701
        #mikebrown 17824
        #mediablackout 5322
        #gaza 4497
        #michaelbrown 2541
        #dontshoot 1968
        #anonymous 1836
        #stl 1607
        #palestine 1542
        #prayforferguson 1525
        #justiceformikebrown 1322
        #opferguson 1160
        #myawhite 995
        #usa 956
        #policestate 906
        #fergusonshooting 875
        #tcot 805
        #inners 773
        #iraq 736
        #fergusonriot 656
```

```
en 282138
es 3759
und 1882
de 1133
tr 795
fr 623
et 476
sk 463
tl 330
in 306
ar 282
it 241
```