



# The Beautiful Mind of a Timeline

Yet Another Data Science Talk, Filled With Buzzwords

OSDFCon 2019



Kristinn Guðjónsson

Johan Berggren

# Current State



Data Overflow

“ Ain’t nobody got time  
for all that data...”

---

**Kristinn Guðjónsson**

*Security Engineer, Google*

Is there any hope?

# Data Science



Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data



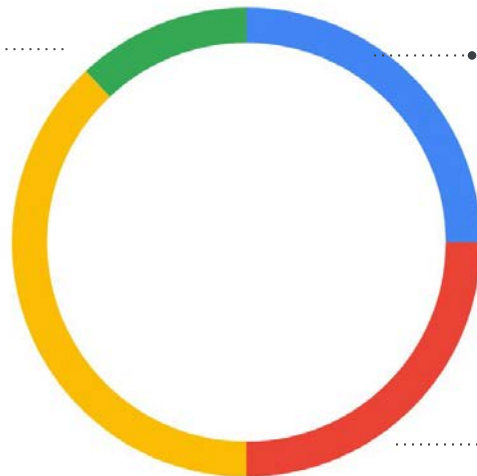
48%  
Something

## Some Random Statistics

19%  
Lorem ipsum

26%  
Foo

29%  
Foobar



# Data Scientist



# Forensicator





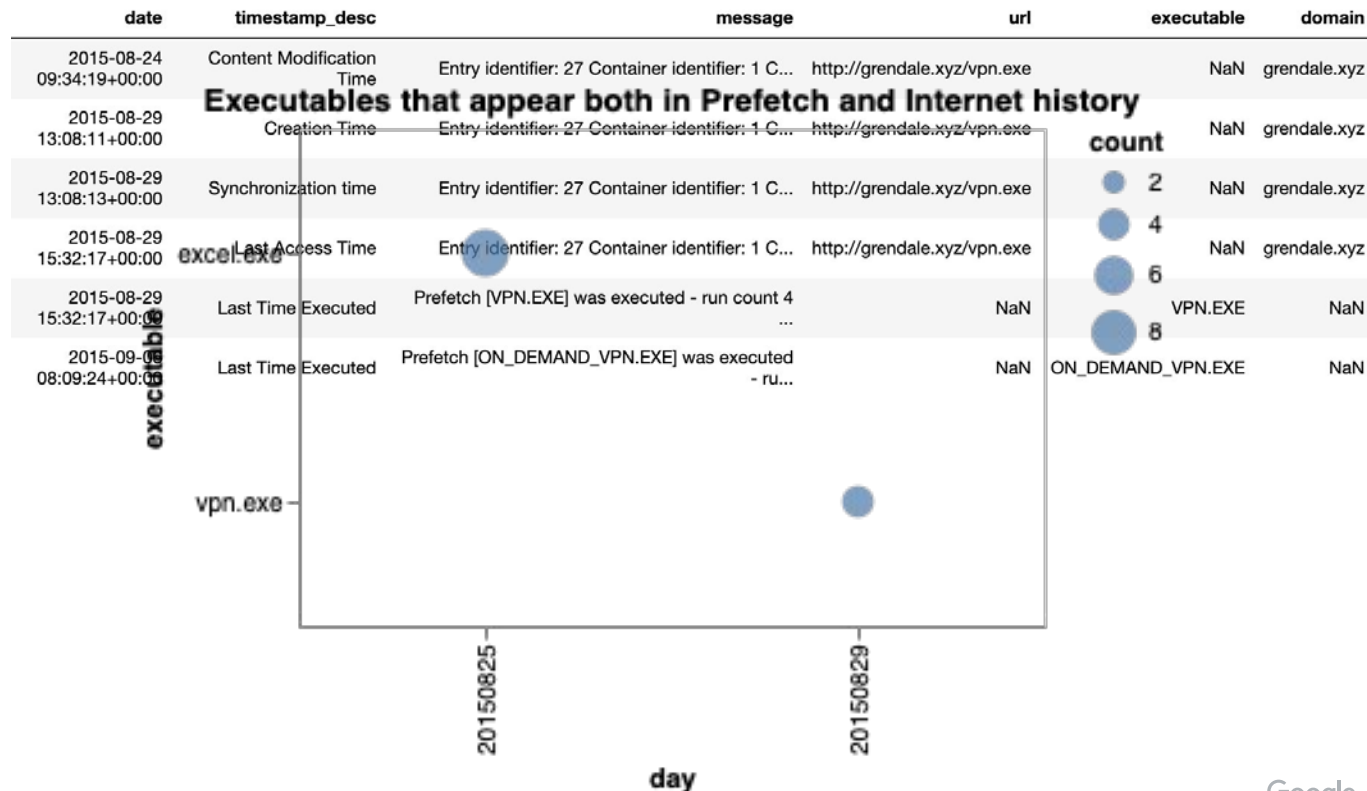
# Anything in Common With Forensics?



**...extract knowledge and insights from structured and unstructured data...**

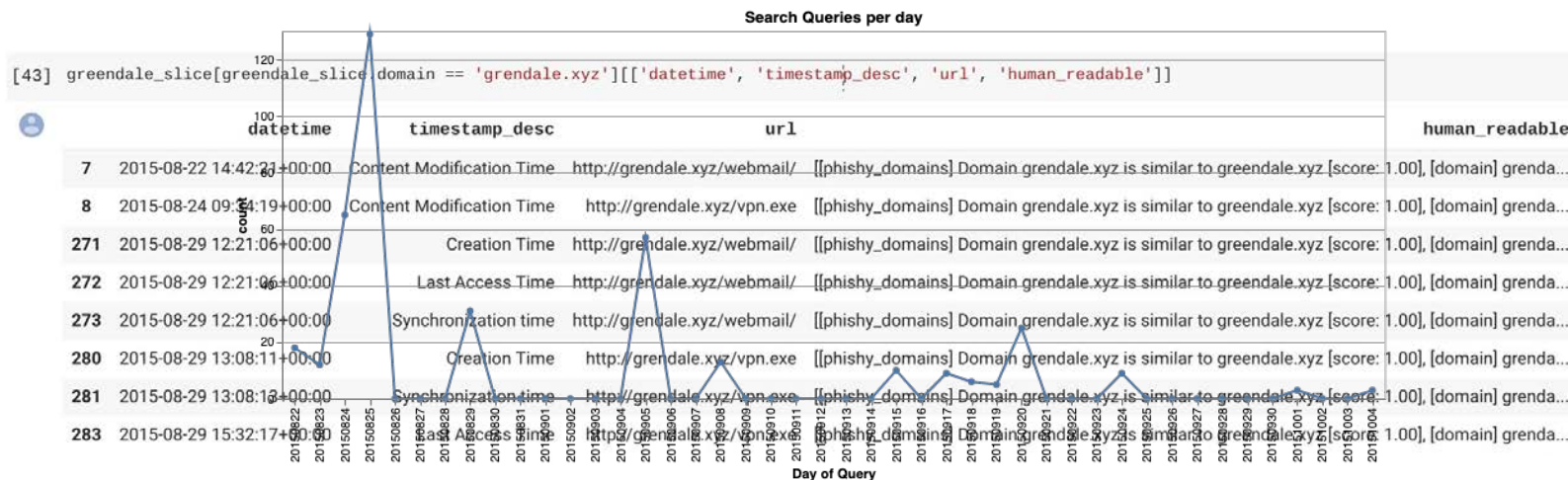
# What Can We Gain From Data Science?

- 1 Tools
- 2 Methodology
- 3 Visualization
- 4 Aggregation



# IPython Notebooks - Jupyter/Colab

- 1 Interactive “notebooks”, mix of text/code
- 2 Standalone: jupyter, or collaborative: colab
- 3 Uses standard Python



# Can I Play?

README.md

## Timesketch

build passing

pypi v20190207

Open in Colab

launch binder

Tweet

# Live Demo

You will now experience a *live demo* that will most likely fail, as all live demos are doomed to do.

Notebook available at: <https://bit.ly/32iK4xR>

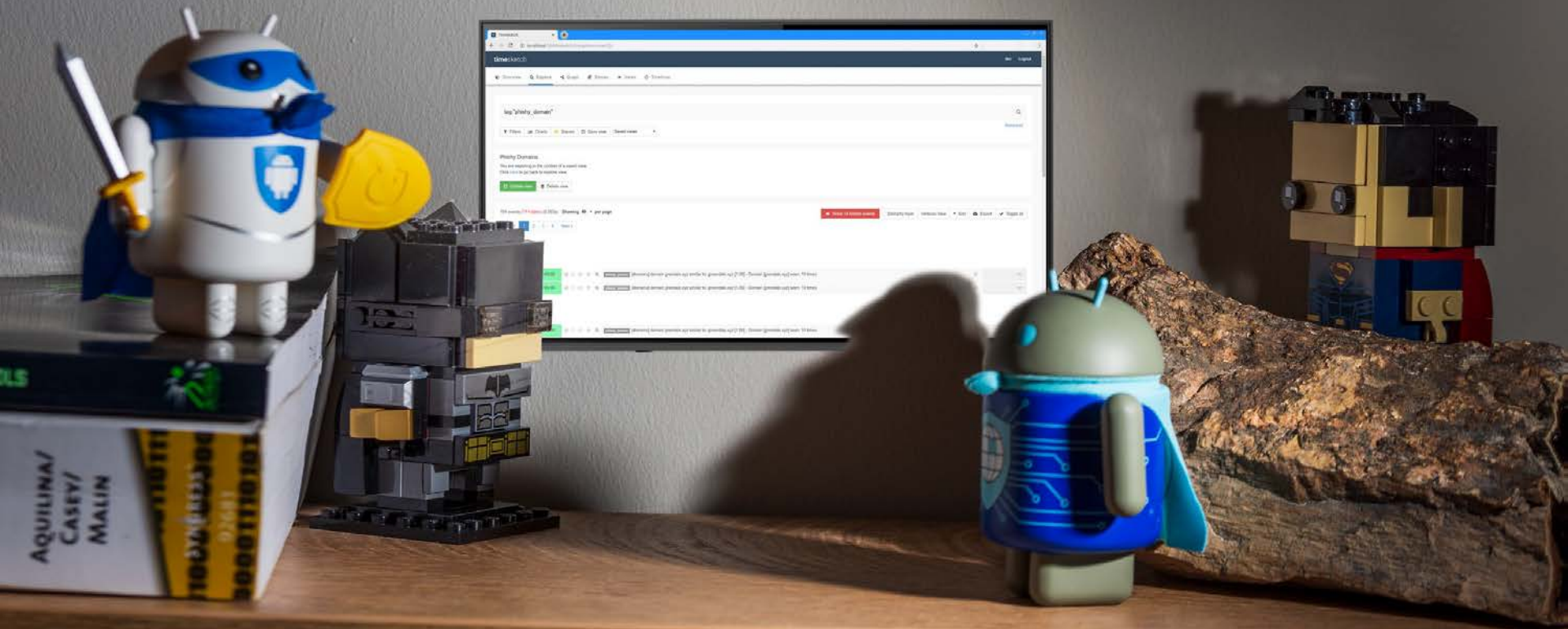
<https://github.com/google/timesketch/blob/master/notebooks/OSDFCon%20Demo.ipynb>

# Summary

- 1 Don't need to become statistician
- 2 Know how to extract features and count
- 3 Great way to “get to know your data”
- 4 Sometimes learning from other fields can benefit you



# Research → Reality





Plaso

Input

160+ parsers

Output

Millions of events



```
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{00A0C9B4D50C}] (default): [REG_SZ] Microsoft COM+ Services Meta Data,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0C5672F9-3EDC-4b24-95B5-A6C54C0B79AD}\InprocServer32] (default): [REG_EXPAND_SZ] %SystemRoot%\System32\wiaaut.dll ThreadingModel: [REG_SZ] Apartment,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{083863F1-70DE-11D0-BD40-00A0C911CE86}\Instance\{70E102B0-5556-11CE-97C0-00AA0055595A}] CLSID: [REG_SZ] {70E102B0-5556-11CE-97C0-00AA0055595A} FilterData: [REG_BINARY] FriendlyName: [REG_SZ] Video Renderer,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{055CB2D7-2969-45CD-914B-76890722F112}\ProgID] (default): [REG_SZ] BDATuner.MPEG2Component.1,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0D0E47ED-7220-411f-8F81-1118095DA5E7}\InProcServer32] (default): [REG_EXPAND_SZ] %SystemRoot%\System32\provsvc.dll ThreadingModel: [REG_SZ] Both,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{12D51199-0DB5-46FE-A120-47A3D7D937CC}] (default): [REG_SZ] DVD: Pluggable Protocol,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0CFDD070-581A-11D2-9EE6-006008039E37}] (default): [REG_SZ] General,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0D6417E3-866C-4959-9816-4BF5B85CDAD7}] (default): [REG_SZ] MsasrUI Class,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{093FF999-1EA0-4079-9525-9614C3504B74}\Programmable] Value: No values stored in key.,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0B124F8F-91F0-11D1-B8B5-006008059382}\InProcServer32] (default): [REG_EXPAND_SZ] %SystemRoot%\System32\appwiz.cpl ThreadingModel: [REG_SZ] Apartment,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{0BE35202-8F91-11CE-9DE3-00AA004BB851}] (default): [REG_SZ] Picture Property Page,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{067B4B81-B1EC-489f-B111-940EBDC44EBE}\VersionIndependentProgID] (default): [REG_SZ] WMDMCESP.WMDMCESP,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
2009-07-14T04:53:38.728440+00:00,Content Modification Time,REG,UNKNOWN,[HKEY_LOCAL_MACHINE\Software\Classes\Wow6432Node\CLSID\{14d7a407-396b-44b3-be85-5199a0f0f80a}] (default): [REG_SZ] Media Foundation DShow Source Resolver,winreg/winreg_default,OS:/vagrant/example_data/Greendale/1/evtx_registry/student-pc1/config/SOFTWARE,-
```

sed | grep | awk



timesketch

Digital Forensic Timeline Analysis

1

Search

2

Collaboration

[Overview](#) [Explore](#) [Stories](#)

Search

Save

Views ▾

Search

2014-01-02 → 2014-05-02 + Add time range

tag:rare\_domain

working-with-domain

greendale

[Enable all](#)[Disable all](#)

4 events (0.001s)

2014-01-02T13:30:02+00:00



rare\_domain

URL: http://adventori.com/cartouche/appnexus?uid=1106782659016455919 Acc...

greendale

2014-01-02T13:30:02+00:00



rare\_domain

URL: http://adventori.com/cartouche/appnexus?uid=1106782659016455919 Acc...

greendale

81  
days

2014-03-24T17:41:57+00:00



rare\_domain

URL: http://s3-ak.buzzfeed.com/static/2014-03/user\_images/webdr04/24/13/micr...

greendale

2014-03-24T17:41:57+00:00



rare\_domain

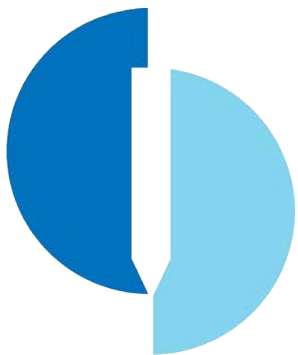
URL: http://s3-ak.buzzfeed.com/static/2014-03/user\_images/webdr04/24/13/micr...

greendale

“ Need to do it more than  
once? Automate.”

Hans Berggren - father of infamous Johan Berggren





timesketch

Digital Forensic Timeline Analysis

- 1 Search
- 2 Collaboration
- 3 **Assisted analysis**



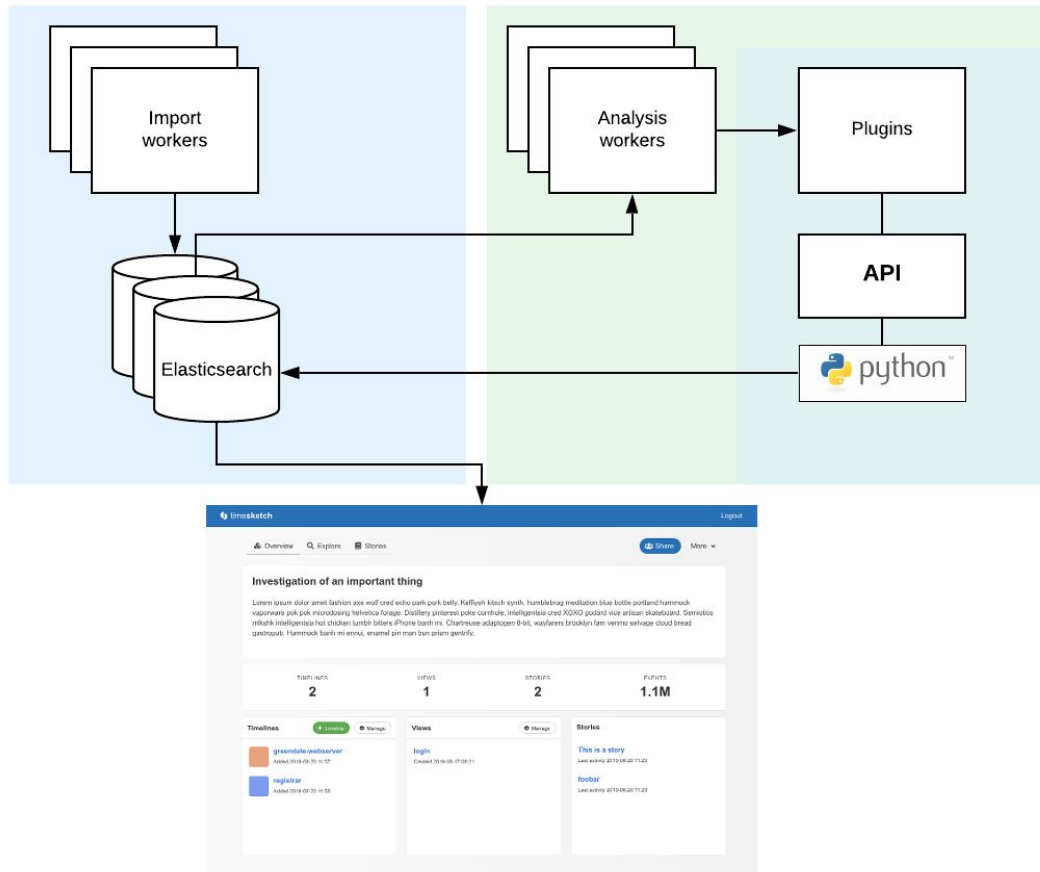
timesketch

Digital Forensic Timeline Analysis

- 1 Search
- 2 Collaboration
- 3 Assisted analysis
- 4 Encoded knowledge



Plugin system for  
running analysis on a  
stream of events.



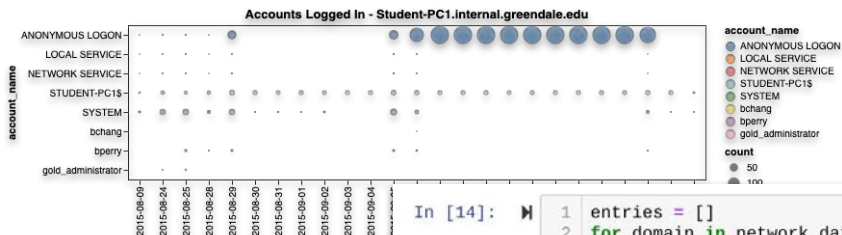


# Just make it work





# For Researchers



In [14]:

```
1 entries = []
2 for domain in network_data.domain.unique():
3     similar_domains = _get_similar_domains(domain)
4     if not similar_domains:
5         continue
6
7     print('Domain: {0:s} does have similar domains')
8     for similarities in similar_domains:
9         s_domain, s_score = similarities
10        print('    [{0:s}] - {1:0.2f}%'.format(s_domain, s_score))
11        entry = {'domain': domain, 'similar_domain': s_domain, 'score': s_score}
12        entries.append(entry)
13    print('---')
14 similar_domains = pd.DataFrame(entries)
```

```
Domain: s.yimg.com does have similar domains
[ytimg.com] - 78.91%
[i.ytimg.com] - 75.78%
---
```

```
Domain: l.yimg.com does have similar domains
[ytimg.com] - 78.91%
```

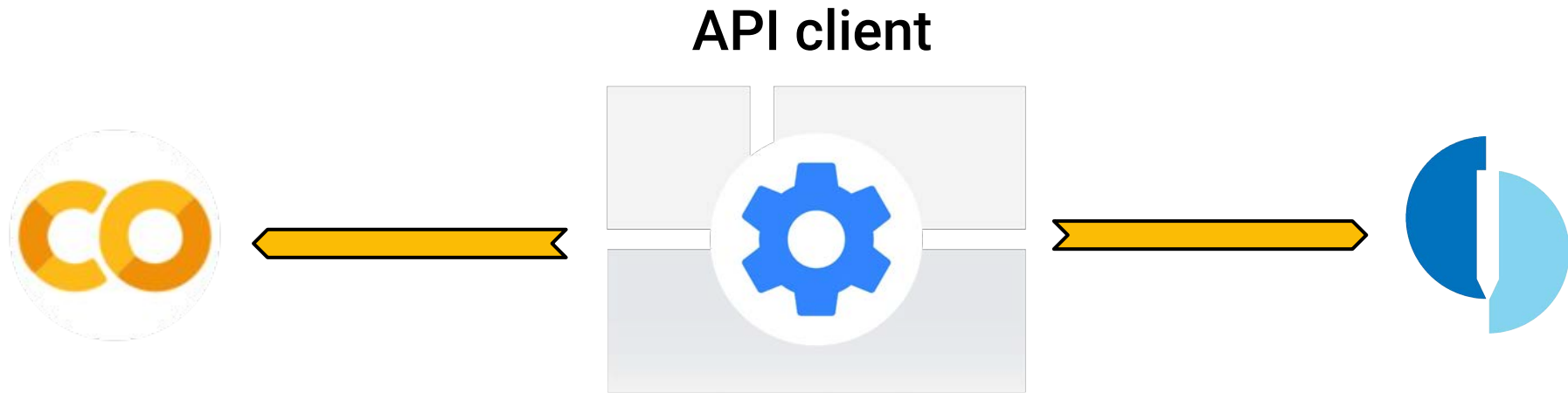
In [15]:

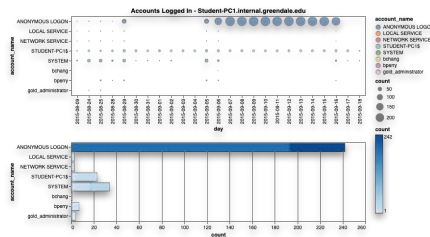
```
1 similar_domains.sort_values('score', ascending=False)
```

Out[15]:

	domain	score	watched_domain
11	greendale.xyz	1.000000	greendale.xyz
13	tse1.mm.bing.net	1.000000	bing.com
0	s.yimg.com	0.789062	ytimg.com
2	l.yimg.com	0.789062	ytimg.com
14	l2.yimg.com	0.789062	ytimg.com
15	l1.yimg.com	0.789062	ytimg.com
3	static-entertainment-neu.s-msn.com	0.765625	msn.com
4	static-finance-neu.s-msn.com	0.765625	msn.com
5	static-sports-neu.s-msn.com	0.765625	msn.com
6	static-hp-neu.s-msn.com	0.765625	msn.com
7	static-hp-eus.s-msn.com	0.765625	msn.com
8	static-news-neu.s-msn.com	0.765625	msn.com
9	img.stb.s-msn.com	0.765625	msn.com
10	img.s-msn.com	0.765625	msn.com
1	s.yimg.com	0.757812	i.ytimg.com
12	tse1.mm.bing.net	0.750000	www.bing.com

# For Researchers

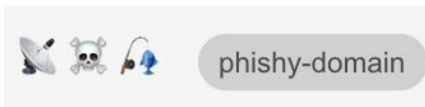
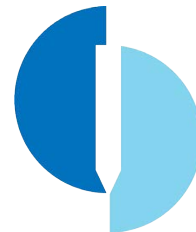




```
In [15]: M[similar_domains.sort_values('score', ascending=False)

Out[15]:
```

	domain	score	watched domain
11	greendate.xyz	1.000000	greendate.xyz
13	tsel1.mn.bing.net	1.000000	bing.com
0	s.yimg.com	0.789062	yimg.com
2	l.yimg.com	0.789062	yimg.com
14	l2.yimg.com	0.789062	yimg.com
15	l1.yimg.com	0.789062	yimg.com
3	static-entertainment-nru.s-mn.com	0.765625	mn.com
4	static-finance-nru.s-mn.com	0.765625	mn.com
5	static-sports-nru.s-mn.com	0.765625	mn.com
6	static-tp-nru.s-mn.com	0.765625	mn.com
7	static-tp-eus.s-mn.com	0.765625	mn.com
8	static-revee-nru.s-mn.com	0.765625	mn.com
9	img.sib.s-mn.com	0.765625	mn.com
10	img.s-mn.com	0.765625	mn.com
1	s.yimg.com	0.757812	l.yimg.com
12	tsel1.mn.bing.net	0.750000	www.bing.com



\$ 12t\_scaffolder.py

== Starting the scaffolder ==  
Gathering required information.

Path to the project root: <PATH TO TIMESKETCH SOURCE>

Path [.] set as the project path.

Name of the module to be generated. This can be something like "foobar sqlite" or "event analytics".

This will be used for class name generation and file name prefixes.

Module Name: demo\_analyzer

About to create a new feature branch to store newly generated code.

Creating feature branch: demo\_analyzer inside .

Switching to feature branch demo\_analyzer

Ready to generate files? [Y/n]:

```
$ 12t_scaffolder.py
```

```
== Starting the scaffolder ==
```

```
Gathering required information.
```

```
Path to the project root: <PATH TO TIMESKETCH SOURCE>
```

```
Path [.] set as the project path.
```

```
Name of the module to be generated. This can be something like "foobar sqlite" or  
"event analytics".
```

```
This will be used for class name generation and file name prefixes.
```

```
Module Name: demo_analyzer
```

```
About to create a new feature branch to store newly generated code.
```

```
Creating feature branch: demo_analyzer inside .
```

```
Switching to feature branch demo_analyzer
```

```
Ready to generate files? [Y/n]: Y
```

```
File: ./timesketch/lib/analyzers/demo_analyzer.py written to disk.
```

```
File: ./timesketch/lib/analyzers/demo_analyzer_test.py written to disk.
```

```
def run(self):
    """Entry point for the analyzer.

    Returns:
        String with summary of the analyzer result
    """
    # TODO: Add Elasticsearch query to get the events you need.
    query = ''

    # Generator of events based on your query.
    events = self.event_stream(query_string=query)

    # TODO: Add analyzer logic.
    # event.add_star()
    # event.add_comment('comment')
    for event in events:
        pass

    # TODO: Return a summary from the analyzer.
    return 'String to be returned'
```

# For Analysts

The screenshot shows the 'timesketch' web interface. At the top is a blue header with the 'timesketch' logo on the left and a 'Logout' link on the right. Below the header is a navigation bar with three tabs: 'Overview' (selected), 'Explore', and 'Stories'. To the right of these tabs are a 'Share' button and a 'More' dropdown menu. The main content area is titled 'Demo of analyzers'. Below this title is a row of four summary cards: 'TIMELINES' with a count of 2, 'VIEWS' with a count of 3, 'STORIES' with a count of 2, and 'EVENTS' with a count of 46.8K. A red arrow points from the 'VIEWS' card down to the 'Views' section below. The 'Views' section has a 'Manage' button. Below the summary cards are three columns: 'Timelines' (with a '+ Timeline' and 'Manage' button) showing two items: 'working-with-domain' (added 2019-10-15 11:34) and 'greendale' (added 2019-10-15 11:34); 'Views' (with a 'Manage' button) showing three items: '[phishy\_domains] Phishy Domains' (created 2019-09-30 10:22), 'Rare domains' (created 2019-10-11 16:42), and 'Domains' (created 2019-10-11 14:54); and 'Stories' showing two items: 'test' (last activity 2019-10-07 12:12) and 'Another story' (last activity 2019-10-14 16:23).

timesketch Logout

Overview Explore Stories Share More

Demo of analyzers

TIMELINES 2 VIEWS 3 STORIES 2 EVENTS 46.8K

Timelines + Timeline Manage

- working-with-domain Added 2019-10-15 11:34
- greendale Added 2019-10-15 11:34

Views Manage

- [phishy\_domains] Phishy Domains Created 2019-09-30 10:22
- Rare domains Created 2019-10-11 16:42
- Domains Created 2019-10-11 14:54

Stories

- test Last activity 2019-10-07 12:12
- Another story Last activity 2019-10-14 16:23

# For Analysts

4 events (0.004s)

2015-08-22T14:42:21+00:00



phishy-domain

URL: http://grendale.xyz/webmail/ Access count: 1 Sync count: 0 Filenam...

grendale

2015-08-22T14:42:21+00:00



phishy-domain

URL: http://grendale.xyz/webmail/ Access count: 1 Sync count: 0 Filenam...

grendale

1  
days

2015-08-24T09:34:19+00:00



phishy-domain

URL: http://grendale.xyz/vpn.exe Access count: 4 Sync count: 0 Filename...

grendale

2015-08-24T09:34:19+00:00



phishy-domain

URL: http://grendale.xyz/vpn.exe Access count: 4 Sync count: 0 Filename...

grendale



# Summary

“Be that unicorn...”

---

**Kristinn Guðjónsson**

*True believer of forensic unicorns*



# Any Questions?



(if we don't have time to answer, tackle us in the hallway)