

Telemetry Onboarding

Roberto Vitillo - Firefox Data Engineering Team

Ask a question



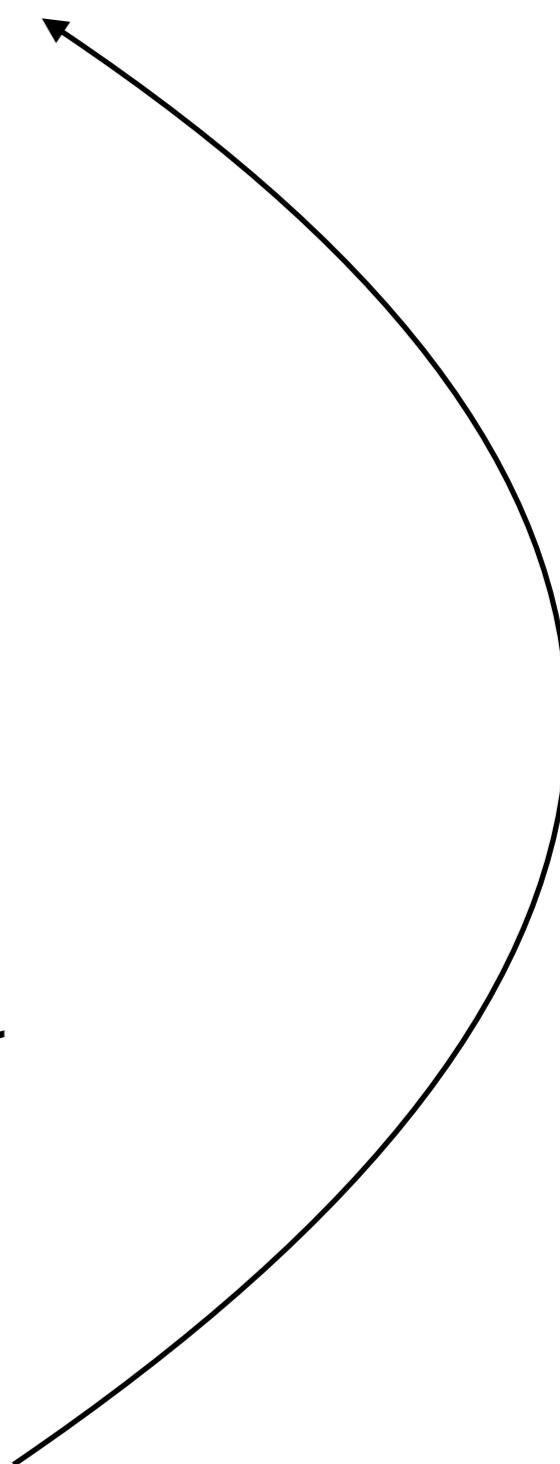
Get the data



Analyse the data



Report



TELEMETRY 101

A dark grey sidebar with white icons and text, listing settings categories: General, Search, Content, Applications, Privacy, Security, Sync, and Advanced.

Advanced

General Data Choices Network Update Certificates

Enable Firefox Health Report **(Base Telemetry)**
Helps you understand your browser performance and shares data with Mozilla about your browser health [Learn More](#)

Share additional data (i.e., Telemetry) **(Extended Telemetry)**
Shares performance, usage, hardware and customization data about your browser with Mozilla to help us make Firefox better [Learn More](#)

Enable Crash Reporter
Firefox submits crash reports to help Mozilla make your browser more stable and secure [Learn More](#)



```
{  
  type: <string>, // "main", "activation", "deletion", "saved-session", ...  
  id: <UUID>, // a UUID that identifies this ping  
  creationDate: <ISO date>, // the date the ping was generated  
  version: <number>, // the version of the ping format, currently 4  
  
  application: {  
    architecture: <string>, // build architecture, e.g. x86  
    buildId: <string>, // "20141126041045"  
    name: <string>, // "Firefox"  
    version: <string>, // "35.0"  
    displayVersion: <string>, // "35.0b3"  
    vendor: <string>, // "Mozilla"  
    platformVersion: <string>, // "35.0"  
    xpcomAbi: <string>, // e.g. "x86-msvc"  
    channel: <string>, // "beta"  
  },  
  
  clientId: <UUID>, // optional  
  environment: { ... }, // optional, not all pings contain the environment  
  payload: { ... }, // the actual payload data for this ping type  
}
```

about:telemetry

This page shows the information about performance, hardware, usage and customisations collected by Telemetry. This information is submitted to Mozilla to help improve Firefox Developer Edition.

FHR data upload is **enabled**. [Change](#)
Extended Telemetry recording is **enabled**. [Change](#)

Ping data source:
 Current ping data
 Archived ping data

Ping data display:
 Structured
 Raw JSON

Show subsession data

Payload
Parent Payload [▼](#)

General Data Click to toggle section

Environment Data Click to toggle section

Session Information Click to toggle section

Scalars Click to toggle section

Keyed Scalars (No data collected)

Histograms Click to toggle section

Keyed Histograms Click to toggle section

Simple Measurements Click to toggle section

Need help?

Real-time



Hindsight Admin (0.0.4)

ttrink@mozilla.com

Plugins **Plugin Deployment** Grammar Tester Dashboards

Heka Analysis Plugin Configuration

```
filename = 'example.lua'
message_matcher = 'TRUE'
preserve_data = false
ticker_interval = 60
```

Heka Analysis Plugin

```
require "circular_buffer"
cb = circular_buffer.new(60, 1, 60)
cb:set_header(1, "messages")

function process_message()
    cb:add(read_message("Timestamp"), 1, 1)
    return 0
end

function timer_event(ns, shutdown)
    inject_payload("cbuf", "messages per minute", cb)
end
```

Run Matcher Test Plugin Deploy Plugin

Matched Messages

▼ Messages

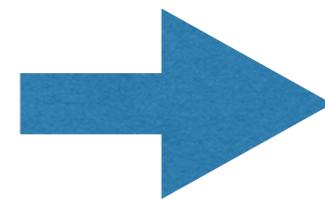
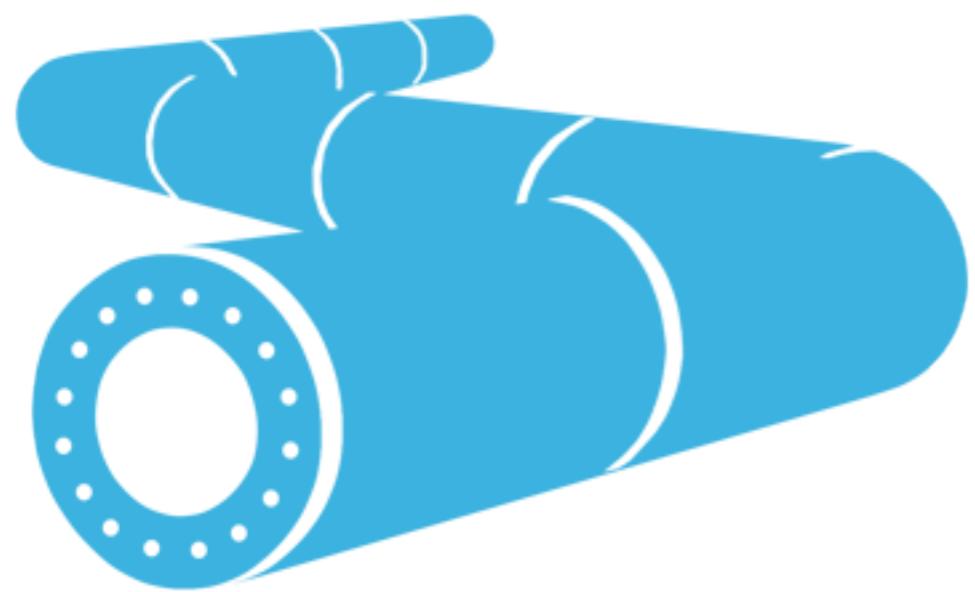
- ▶ Uuid = 4548ab665f364c0da76431dc3f66c329
- ▶ Uuid = 23848f99be7d4e99ac588f8ecad9075c
- ▶ Uuid = 56b83b1bc66a4b20a3bee9bbcccc9928
- ▶ Uuid = 220e9cb4112445ada12e9f2778f7878e
- ▶ Uuid = 7bba2e4c65b440dba157316be8ee2f96

Debug Output

Injected Messages

► Messages

<https://pipeline-cep.prod.mozaws.net/>



Amazon RDS



Amazon **Redshift**



Ask a question



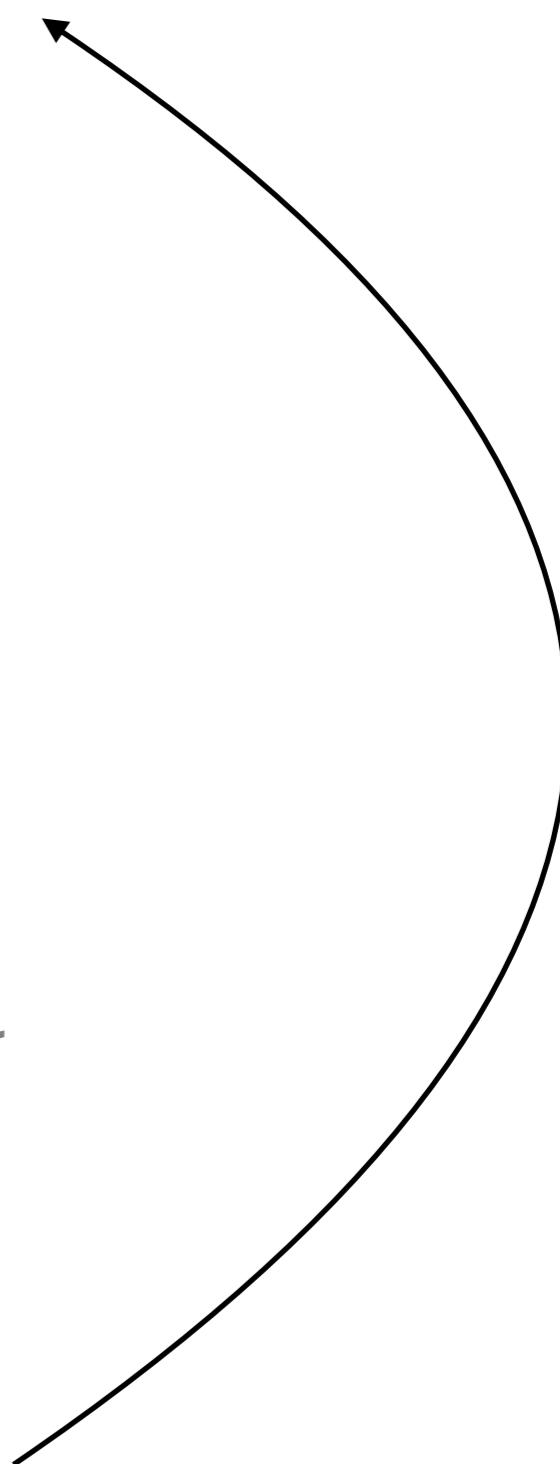
Get the data



Analyse the data



Report



Step 1: Ask a question



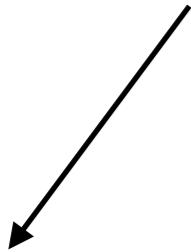
Step 2: Get the data

Data exists

~~Data exists~~

Step 2: Get the data

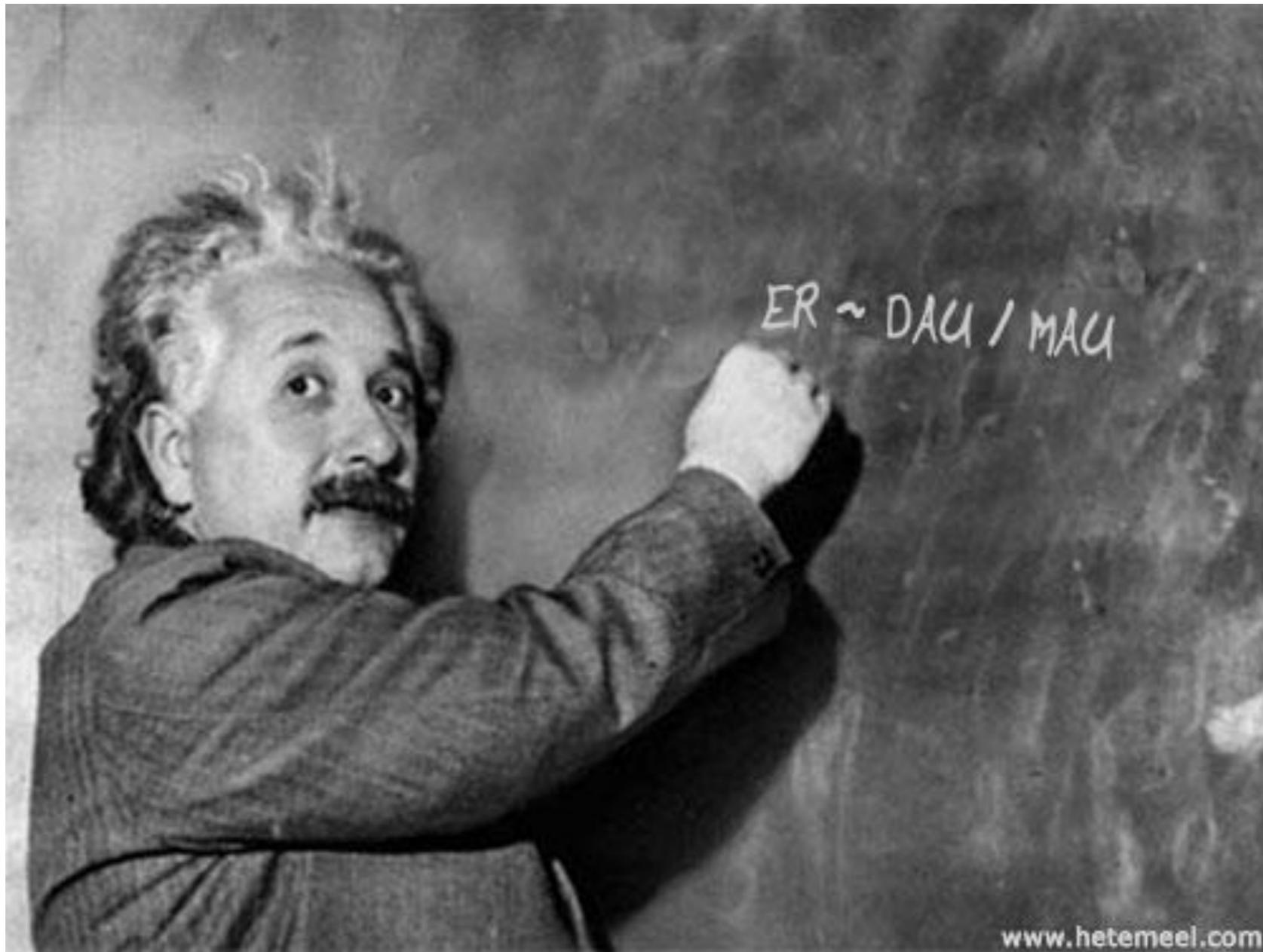
Data exists



Data ~~exists~~

dashboard

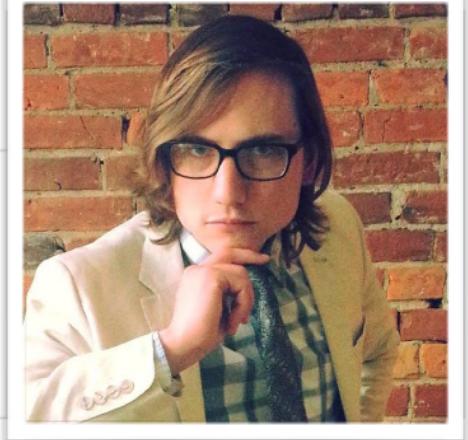
KPIs



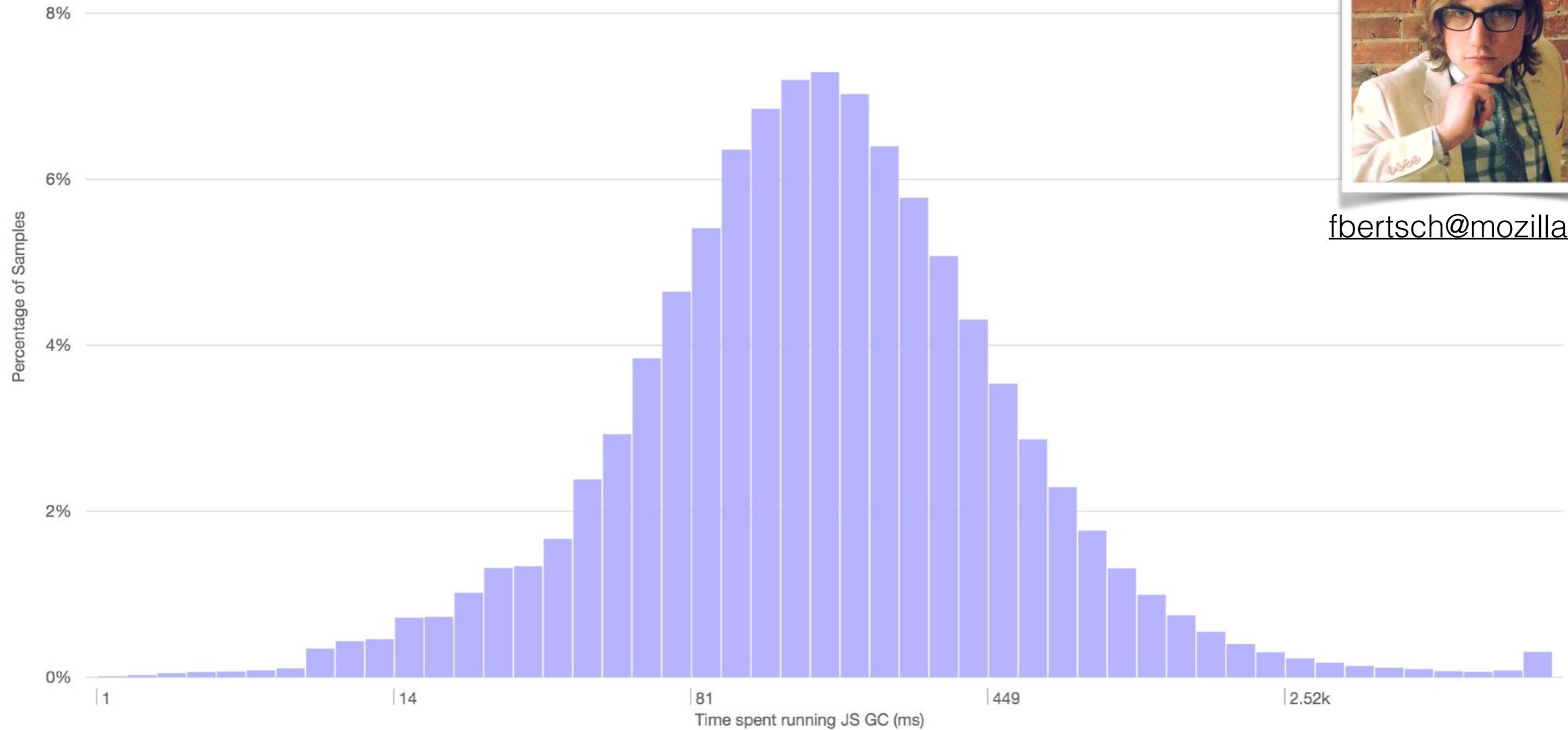
<https://metrics.services.mozilla.com/>

Performance

Need help?

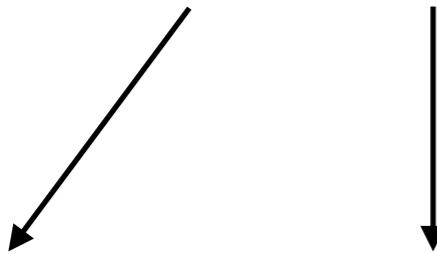


fbertsch@mozilla.com



Step 2: Get the data

Data exists



Data exists

dashboard **s.t.m.o**

Need help?



bimsland@mozilla.com

Events by Date & Action

No description

Created By Arik Fraimovich Updated at 7 months ago Created at 7 months ago

Search schema... Execute Format Query Data Source re:dash

cohort2 (9)
cohort_example (15)
dashboards (1408)
data_sources (5)
events (605319)
mytable (0)
persons (1)
queries (2015)
query_results (45474)
testable (0)
visualizations (2586)
widgets (1010)

```
1 select created_at::date, action, case when action = 'view' then count(0) /100 else count(0) end
2 from events
3 where created_at > current_date - 30 and action in ('fork', 'edit_name', 'subscribe', 'view')
4 group by 1,2
5 order by 1,2 desc
```

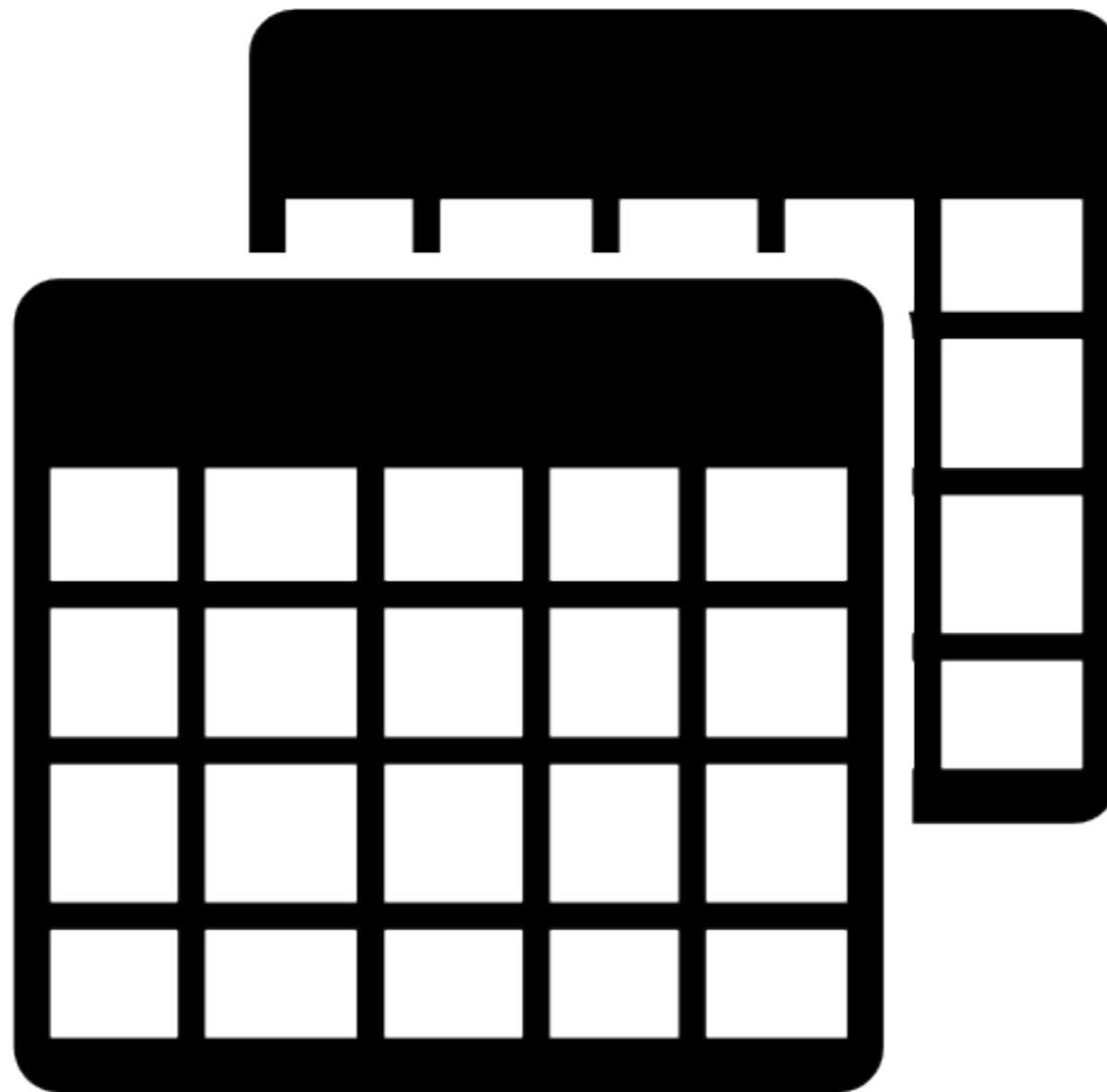
Runtime 5s Rows 86 Refresh Schedule Never Last update 3 minutes ago Download Dataset

TABLE CHART × CHART × + NEW VISUALIZATION

CREATED_AT	ACTION	COUNT
29/05/16	view	5
29/05/16	edit_name	3
30/05/16	view	22
30/05/16	subscribe	1

<https://sql.telemetry.mozilla.org/>

Datasets

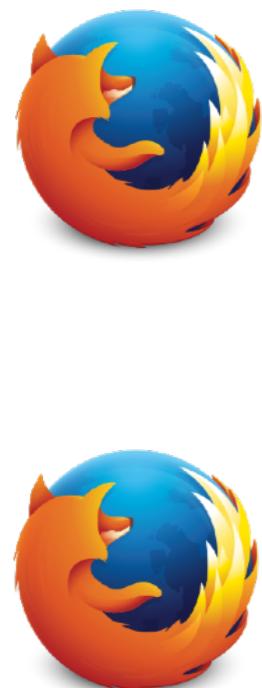


https://github.com/mozilla/telemetry-batch-view/blob/master/docs/choosing_a_dataset.md

Need help?



mreid@mozilla.com



12/12/2016
21/12/2016
01/12/2016

main_summary

Client ID	A	B
“X”	FALSE	42
“Y”	TRUE	null
“X”	FALSE	40

<https://github.com/mozilla/telemetry-batch-view/blob/master/docs/MainSummary.md>

Need help?



longitudinal

rvitillo@mozilla.com

Client ID	A	B
“X”	[FALSE, FALSE, TRUE]	[42, 40, 41]
“Y”	[TRUE]	[null]

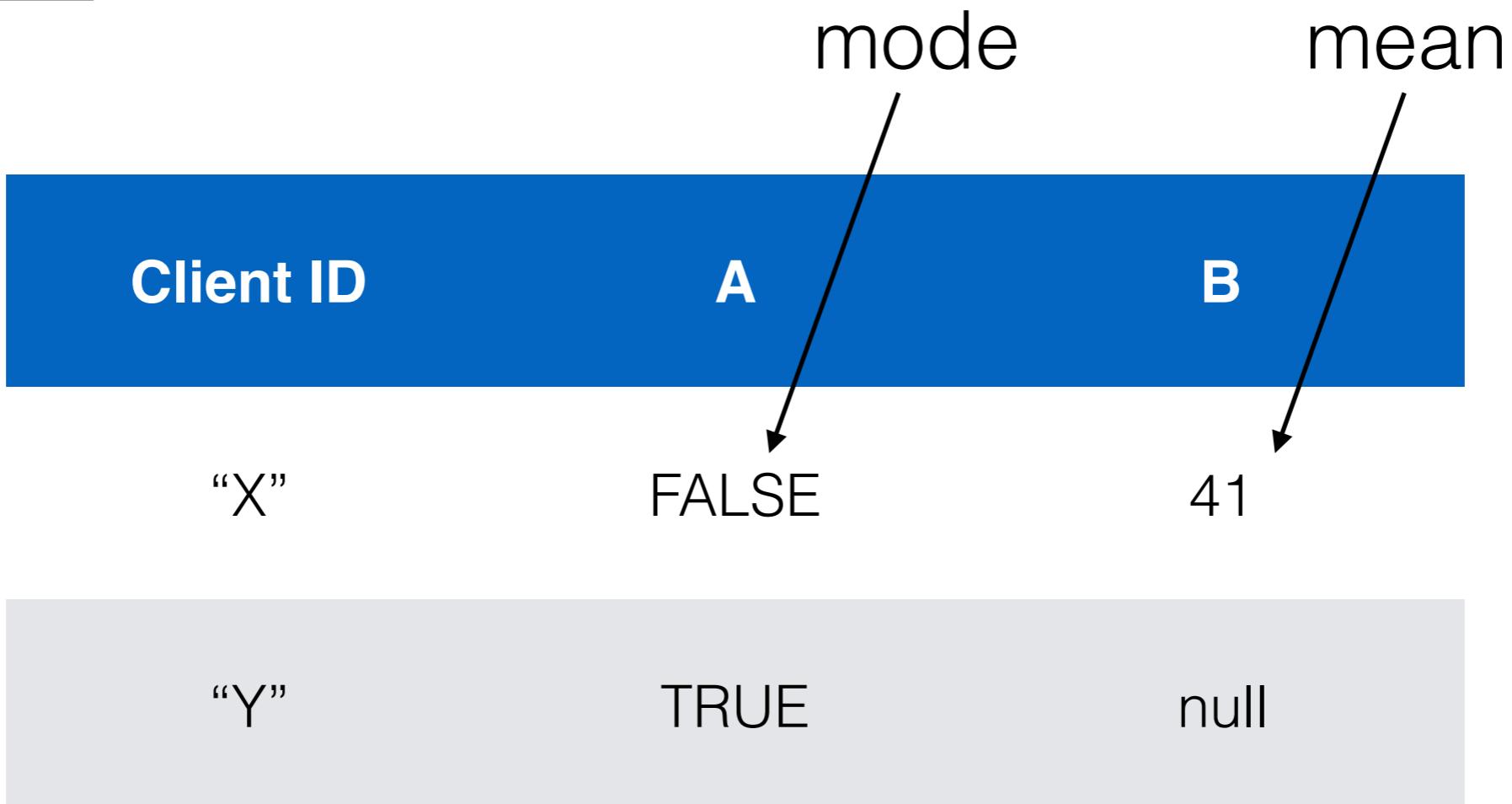
<https://github.com.mozilla/telemetry-batch-view/blob/master/docs/Longitudinal.md>

Need help?



rhaber@mozilla.com

cross_sectional



Need help?



client_count

Clients

A

{"3"}

FALSE

{"1", "2"}

TRUE

Need help?



mdoglio@mozilla.com

crash_aggregates

#crashes

A

B

42

GPU X

OSX

12

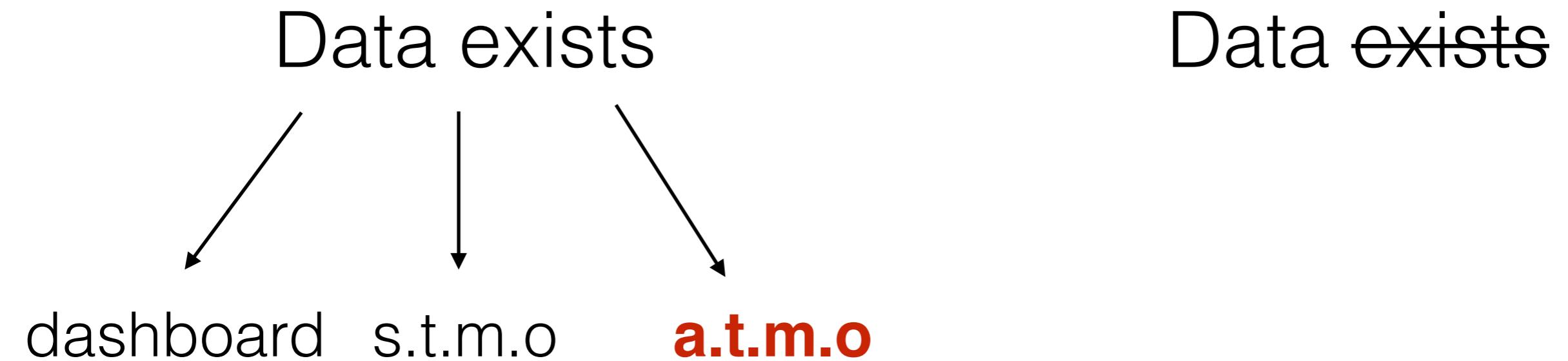
GPU Y

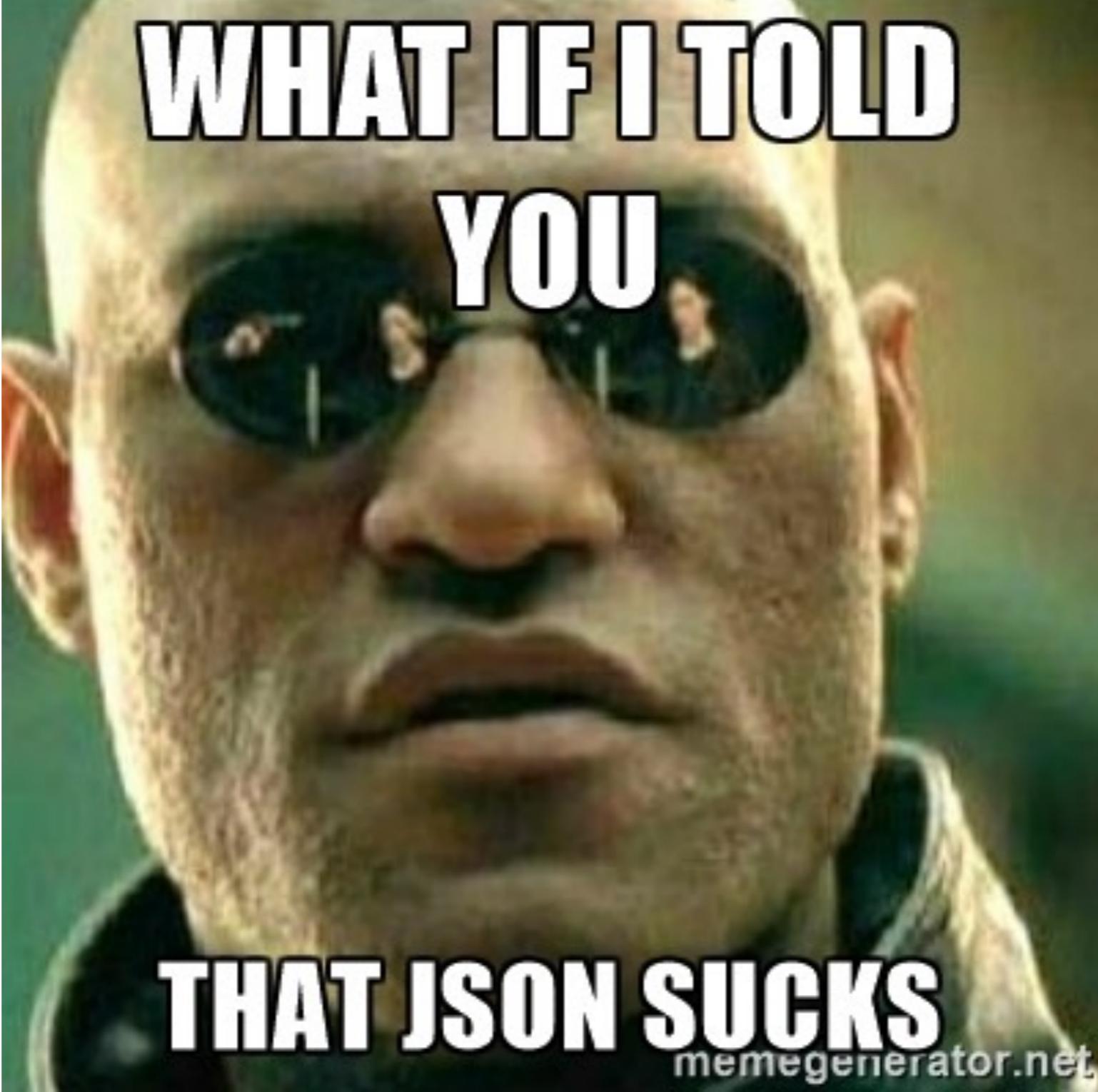
WIN

<https://github.com.mozilla/telemetry-batch-view/blob/master/docs/CrashAggregateView.md>



Step 2: Get the data

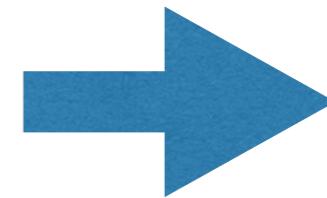
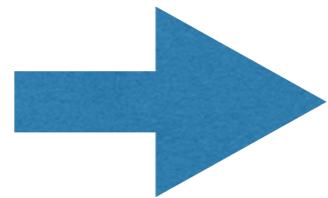
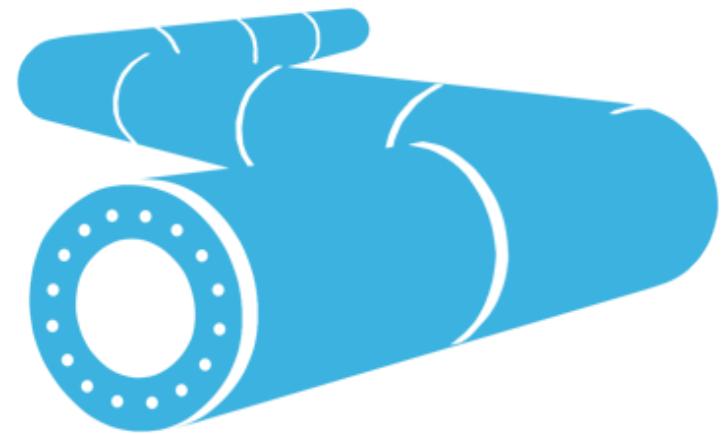




**WHAT IF I TOLD
YOU**

THAT JSON SUCKS

memegenerator.net



Spark

Need help?



jannis@mozilla.com

Telemetry Analysis Service Documentation Tutorial

New Spark cluster Launch a new EMR cluster

Cluster identifier

rvitillo-telemetry-analysis

A brief description of the cluster's purpose, visible in the AWS management console.

Cluster size

1



Number of workers to use in the cluster (1 is recommended for testing or development).

Public SSH key

Browse...

Upload your SSH **public key**, not private key! This will generally be found in places like
`~/.ssh/id_rsa.pub`.

EMR release version

5.0.0



Different EMR versions have different versions of software like Hadoop, Spark, etc

▶ Launch

Cancel

<https://atmo-prod.herokuapp.com/>

Need help?



Let's fetch 10% of Telemetry submissions for a given submission date...

```
In [3]: Dataset.from_source("telemetry").schema
```

```
Out[3]: [u'submissionDate',
 u'sourceName',
 u'sourceVersion',
 u'docType',
 u'appName',
 u'appUpdateChannel',
 u'appVersion',
 u'appBuildId']
```

```
In [4]: pings = Dataset.from_source("telemetry") \
    .where(docType='main') \
    .where(submissionDate="20161101") \
    .where(appUpdateChannel="nightly") \
    .records(sc, sample=0.1)
```

mdoglio@mozilla.com

... and extract only the attributes we need from the Telemetry submissions:

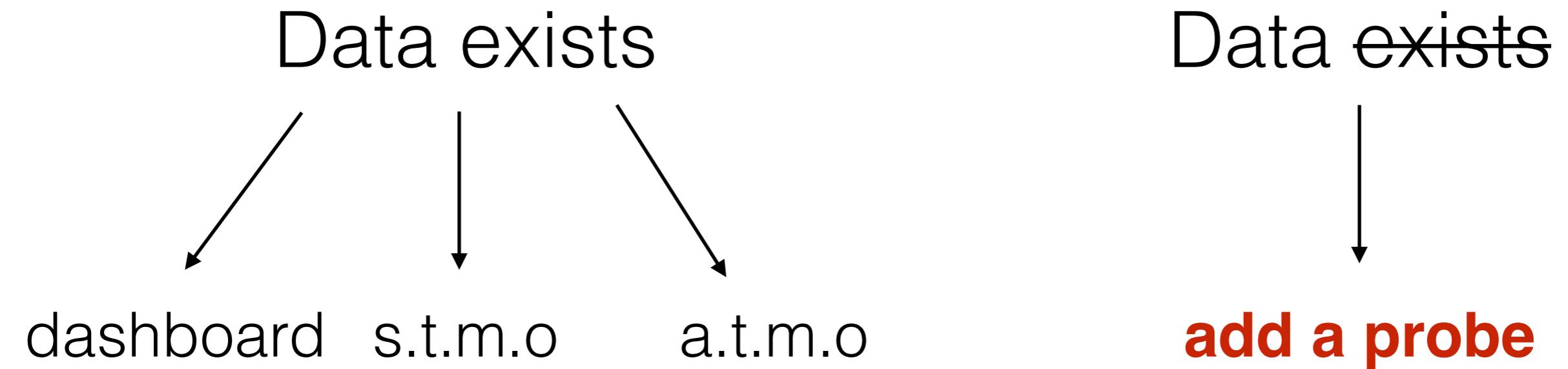
```
In [5]: subset = get_pings_properties(pings, ["clientId",
                                         "environment/system/os/name",
                                         "payload/simpleMeasurements/firstPaint"])
```

Let's filter out submissions with an invalid startup time:

```
In [6]: subset = subset.filter(lambda p: p.get("payload/simpleMeasurements/firstPaint", -1) >= 0)
```

https://wiki.mozilla.org/Telemetry/Custom_analysis_with_spark

Step 2: Get the data



New ping?

```
{  
  type: <string>, // "main", "activation", "deletion", "saved-session", ...  
  id: <UUID>, // a UUID that identifies this ping  
  creationDate: <ISO date>, // the date the ping was generated  
  version: <number>, // the version of the ping format, currently 4  
  
  application: {  
    architecture: <string>, // build architecture, e.g. x86  
    buildId: <string>, // "20141126041045"  
    name: <string>, // "Firefox"  
    version: <string>, // "35.0"  
    displayVersion: <string>, // "35.0b3"  
    vendor: <string>, // "Mozilla"  
    platformVersion: <string>, // "35.0"  
    xpcomAbi: <string>, // e.g. "x86-msvc"  
    channel: <string>, // "beta"  
  },  
  
  clientId: <UUID>, // optional  
  environment: { ... }, // optional, not all pings contain the environment  
  payload: { ... }, // the actual payload data for this ping type  
}
```



```
19 tab_open_event_count:  
20     bug_numbers:  
21         - 1271304  
22     description: >  
23         The count of tab open events per subsession, across all windows, after the  
24         session has been restored. This includes tab open events from private windows  
25         and from manual session restorations (i.e. after crashes and from about:home).  
26 expires: "55"  
27 kind: uint  
28 notification_emails:  
29     - rweiss@mozilla.com  
30 release_channel_collection: opt-out  
31
```



aplacitelli@mozilla.com

```
"TELEMETRY_TUTORIAL_PROBE": {  
    "alert_emails": ["rvitillo@mozilla.com"],  
    "bug_numbers": [1242013],  
    "expires_in_version": "55",  
    "kind": "exponential",  
    "high": 1000,  
    "n_buckets": 50,  
    "description": "Telemetry tutorial probe (ms)"  
},
```

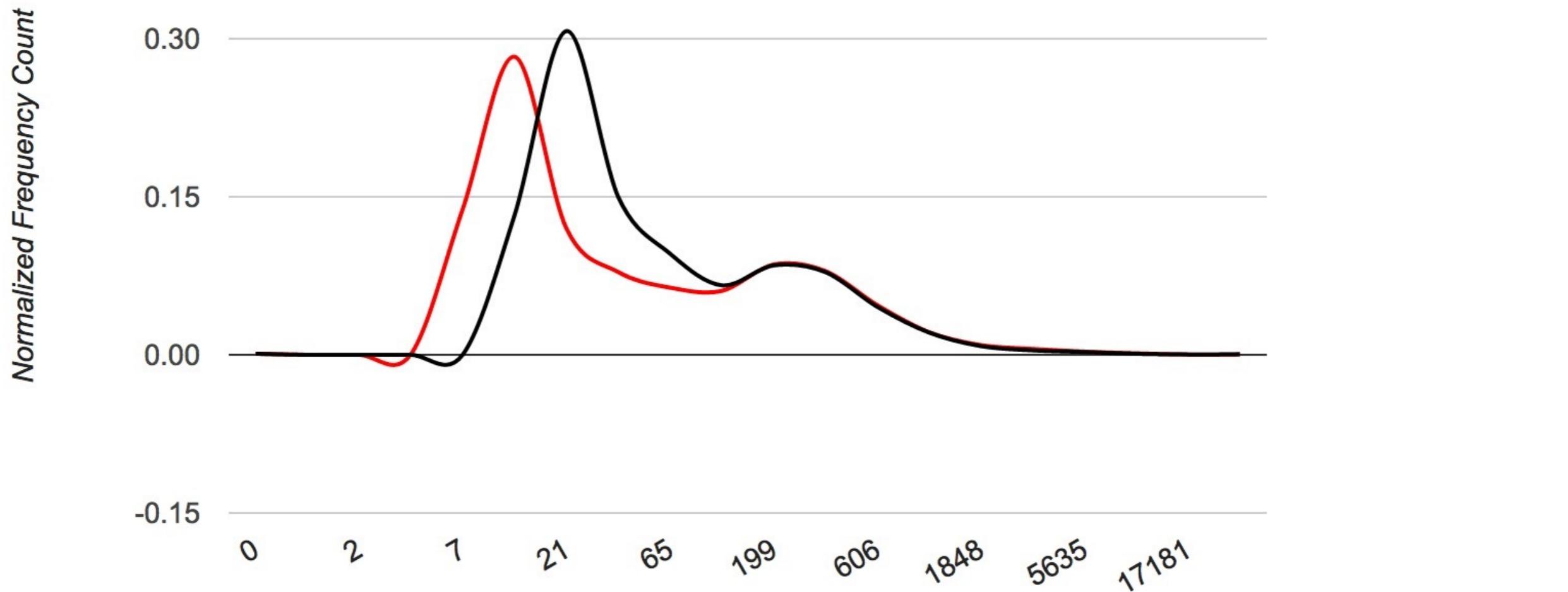
[https://developer.mozilla.org/en-US/docs/Mozilla/Performance/
Adding_a_new_Telemetry_probe](https://developer.mozilla.org/en-US/docs/Mozilla/Performance/Adding_a_new_Telemetry_probe)

Need help?



FX_SESSION_RESTORE_STARTUP_ONLOAD_INITIAL_WINDOW_MS

chutten@mozilla.com



Session restore: Time it takes to finish restoration once we have first opened a window (ms)

<http://alerts.telemetry.mozilla.org>



Owner: Benjamin Smedberg

Peers: Chenxia Liu, François Marier, Ally Naagktgeboren, Rebecca Weiss

<https://www.mozilla.org/en-US/privacy/>

https://wiki.mozilla.org/Firefox/Data_Collection

Ask a question



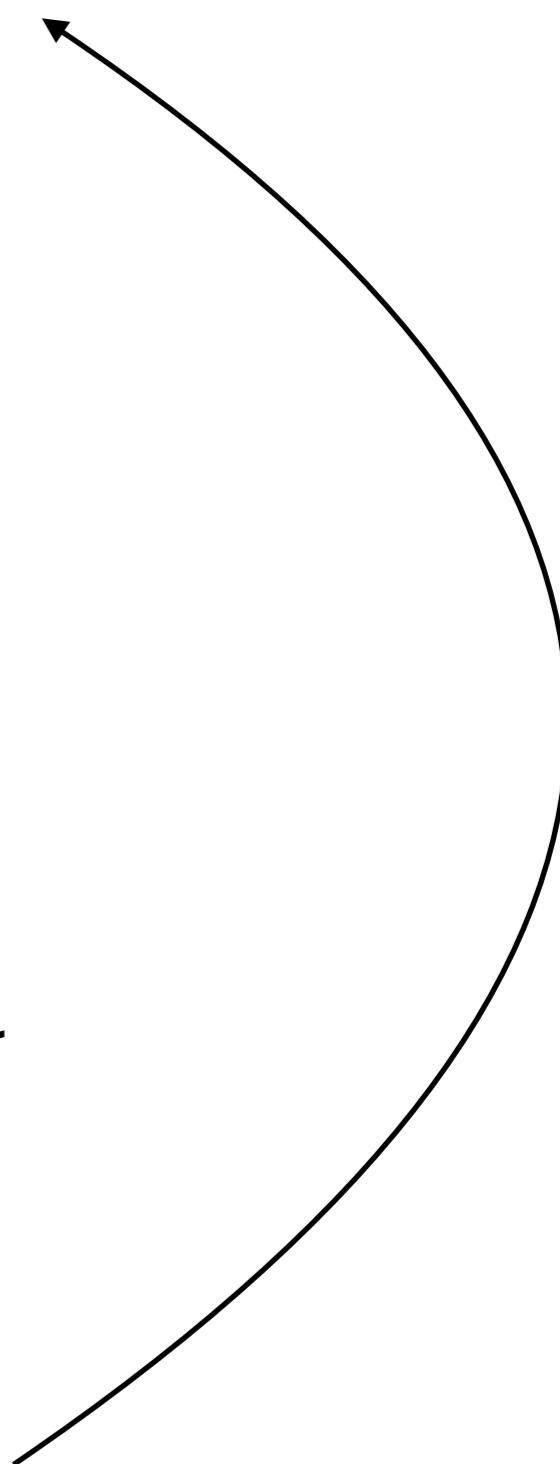
Get the data



Analyse the data



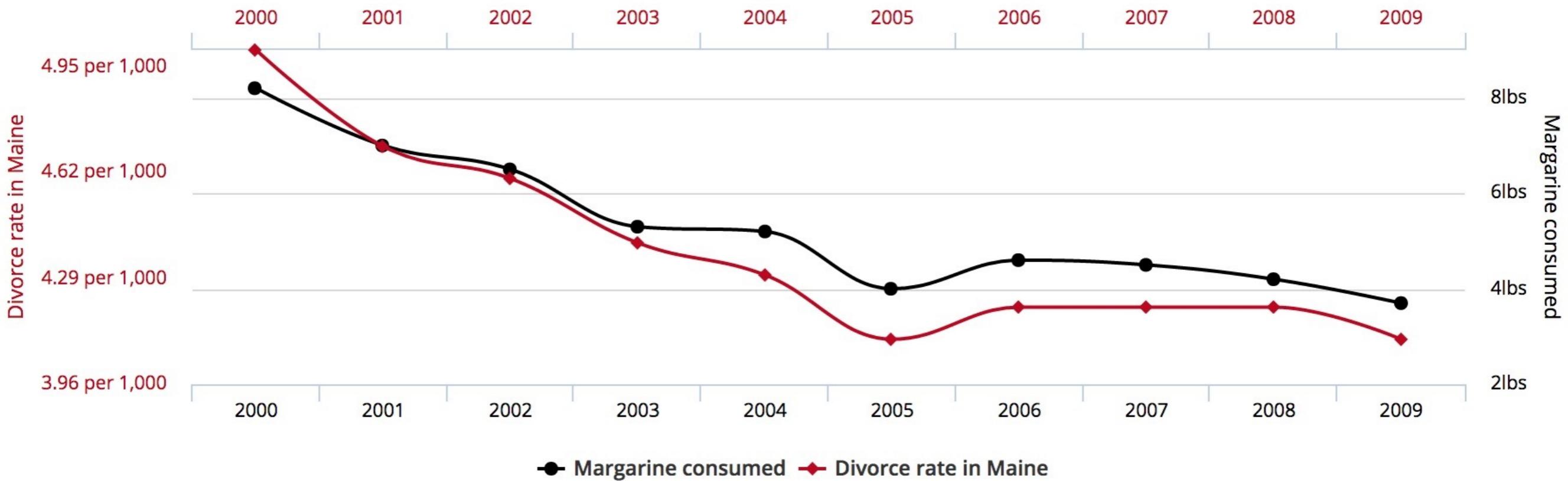
Report



Divorce rate in Maine correlates with Per capita consumption of margarine



Correlation: 99.26% ($r=0.992558$)



STATISTICS DONE WRONG

THE WOEFULLY COMPLETE GUIDE

ALEX REINHART



**PEER REVIEW
MATTERS**

Ask a question



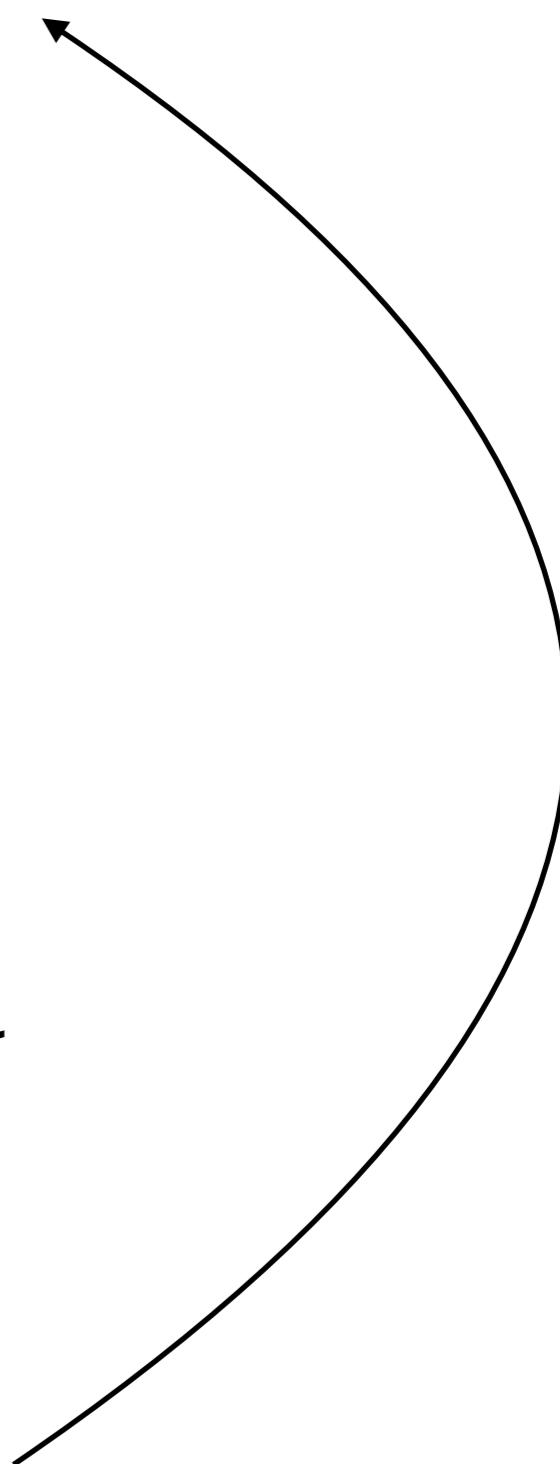
Get the data



Analyse the data



Report



#datapipeline
#telemetry

<https://wiki.mozilla.org/CloudServices/DataPipeline>

<https://wiki.mozilla.org/Telemetry>