

# **Device-specific research and data stream events\***

Adrian Mackenzie  
Sociology, Lancaster University, UK

12 June 2014

\* “SELECT count(type) as event\_count FROM [githubarchive:github.timeline]”

**207 million events**

In any practical going-on with numbers, what matters is that they can be *made* to work, and *making them work* is a politics. Yet is a politics that completely evades conventional foundationist analysis (Verran, 2001: 88)

'momentary apprehensions of depth' (Anne-Lise Riles, 2001)

# The datastream

<http://api.github.com/events>

# Simple event?

```
{  
1.  "id": "2111998059",  
2.  "type": "WatchEvent",  
3.  "actor": {  
4.    "id": 1459103,  
5.    "login": "mmemetea",  
6.    "gravatar_id": "4532d1e4885f579ca7d9aa8748418817",  
7.    "url": "https://api.github.com/users/mmemetea",  
8.    "avatar_url": "https://avatars.githubusercontent.com/u/1459103?"  
9.  },  
10.  "repo": {  
11.    "id": 14802742,  
12.    "name": "OpenSensorsIO/azondi",  
13.    "url": "https://api.github.com/repos/OpenSensorsIO/azondi"  
14.  },  
15.  "payload": {  
16.    "action": "started"  
17.  },  
18.  "public": true,  
19.  "created_at": "2014-05-23T08:40:56Z",  
20.  "org": {  
21.    "id": 5497318,  
22.    "login": "OpenSensorsIO",  
23.    "gravatar_id": "1e0218942846ec8ef59f5d679dbca782",  
24.    "url": "https://api.github.com/orgs/OpenSensorsIO",  
25.    "avatar_url": "https://avatars.githubusercontent.com/u/5497318?"  
  }  
},
```

# Event depth?

```
1. {
2.   "id": "2111998057",
3.   "type": "IssueCommentEvent",
4.   "actor": {
5.     "id": 179265,
6.     "login": "FroMage",
7.     "gravatar_id": "b8932676ae05ecdd084d3839d1f49da8",
8.     "url": "https://api.github.com/users/FroMage",
9.     "avatar_url": "https://avatars.githubusercontent.com/u/179265?"
10.    },
11.   "repo": {
12.     "id": 1287859,
13.     "name": "ceylon/ceylon-compiler",
14.     "url": "https://api.github.com/repos/ceylon/ceylon-compiler"
15.   },
16.   "payload": {
17.     "action": "created",
18.     "issue": {
19.       "url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490",
20.       "labels_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/labels{/name}",
21.       "comments_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/comments",
22.       "events_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/events",
23.       "html_url": "https://github.com/ceylon/ceylon-compiler/issues/1490",
24.       "id": 23820110,
25.       "number": 1490,
26.       "title": "CLI plugins: support distribution",
```

```
26.  
27. "user": {  
28.   "login": "FroMage",  
29.   "id": 179265,  
30.   "avatar_url": "https://avatars.githubusercontent.com/u/179265?",  
31.   "gravatar_id": "b8932676ae05ecdd084d3839d1f49da8",  
32.   "url": "https://api.github.com/users/FroMage",  
33.   "html_url": "https://github.com/FroMage",  
34.   "followers_url": "https://api.github.com/users/FroMage/followers",  
35.   "following_url": "https://api.github.com/users/FroMage/following{/other_user}",  
36.   "gists_url": "https://api.github.com/users/FroMage/gists{/gist_id}",  
37.   "starred_url": "https://api.github.com/users/FroMage/starred{/owner}{/repo}",  
38.   "subscriptions_url": "https://api.github.com/users/FroMage/subscriptions",  
39.   "organizations_url": "https://api.github.com/users/FroMage/orgs",  
40.   "repos_url": "https://api.github.com/users/FroMage/repos",  
41.   "events_url": "https://api.github.com/users/FroMage/events{/privacy}",  
42.   "received_events_url": "https://api.github.com/users/FroMage/received_events",  
43.   "type": "User",  
44.   "site_admin": false  
45. },  
46. "labels": [  
47. {  
48.   "url": "https://api.github.com/repos/ceylon/ceylon-compiler/labels/FEATURE",  
49.   "name": "FEATURE",  
50.   "color": "02e10c"  
51. },
```

```

52.
53. {
54.   "url": "https://api.github.com/repos/ceylon/ceylon-compiler/labels/IN+PROGRESS",
55.   "name": "IN PROGRESS",
56.   "color": "02d7e1"
57. },
58. {
59.   "url": "https://api.github.com/repos/ceylon/ceylon-compiler/labels/tools",
60.   "name": "tools",
61.   "color": "e102d8"
62. },
63. "state": "open",
64. "assignee": null,
65. "milestone": {
66.   "url": "https://api.github.com/repos/ceylon/ceylon-compiler/milestones/7",
67.   "labels_url": "https://api.github.com/repos/ceylon/ceylon-compiler/milestones/7/labels",
68.   "id": 217273,
69.   "number": 7,
70.   "title": "1.1",
71.   "description": "",
72.   "creator": {
73.     "login": "tombentley",
74.     "id": 879487,
75.     "avatar_url": "https://avatars.githubusercontent.com/u/879487?",
76.     "gravatar_id": "19f6b3877a3b4d15467481f7eeb0f635",
77.     "url": "https://api.github.com/users/tombentley",
78.     "html_url": "https://github.com/tombentley",
79.     "followers_url": "https://api.github.com/users/tombentley/followers",
80.     "following_url": "https://api.github.com/users/tombentley/following{/other_user}",
81.     "gists_url": "https://api.github.com/users/tombentley/gists{/gist_id}",
82.     "starred_url": "https://api.github.com/users/tombentley/starred{/owner}{/repo}"}
83.   },
84.   "subscriptions_url": "https://api.github.com/users/tombentley/subscriptions",
85.   "organizations_url": "https://api.github.com/users/tombentley/orgs",
86.   "repos_url": "https://api.github.com/users/tombentley/repos",
87.   "events_url": "https://api.github.com/users/tombentley/events{/privacy}",
88.   "received_events_url": "https://api.github.com/users/tombentley/received_events",
89.   "type": "User",
90.   "site_admin": false
91. },
92.   "open_issues": 21,
93.   "closed_issues": 175,
94.   "state": "open",
95.   "created_at": "2012-11-22T11:43:26Z",
96.   "updated_at": "2014-05-22T16:34:51Z",
97.   "due_on": null
98. },
99.   "comments": 4,
100.  "created_at": "2013-12-05T21:43:00Z",
101.  "updated_at": "2014-05-23T08:40:56Z",
102.  "closed_at": null,
103.  "body": "We need to find a way to support distribution/installation of script plugins. One possible strategy would be to have a 'script' type of file, which can be run by the compiler. This would allow us to have a more modular approach to handling scripts. Another possibility is to have a separate plugin system that can be loaded at runtime. I'm not sure which direction we should take, but I think it's important to have a clear plan for how we want to handle scripts in the future.",
104.  "comment": {
105.    "url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/comments/43984368",
106.    "html_url": "https://github.com/ceylon/ceylon-compiler/issues/1490#issuecomment-43984368",
107.    "issue_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490",
108.    "id": 43984368,
109.    "user": {
110.      "login": "FroMage",
111.      "id": 179265,
112.      "avatar_url": "https://avatars.githubusercontent.com/u/179265?",
113.      "gravatar_id": "b8932676ae05ecdd084d3839d1f49da8",
114.      "url": "https://api.github.com/users/FroMage",
115.      "html_url": "https://github.com/FroMage",
116.      "followers_url": "https://api.github.com/users/FroMage/followers",
117.      "following_url": "https://api.github.com/users/FroMage/following{/other_user}",
118.      "gists_url": "https://api.github.com/users/FroMage/gists{/gist_id}",
119.      "starred_url": "https://api.github.com/users/FroMage/starred{/owner}{/repo}",
120.      "subscriptions_url": "https://api.github.com/users/FroMage/subscriptions",
121.      "organizations_url": "https://api.github.com/users/FroMage/orgs",
122.      "repos_url": "https://api.github.com/users/FroMage/repos",
123.      "events_url": "https://api.github.com/users/FroMage/events{/privacy}",
124.      "received_events_url": "https://api.github.com/users/FroMage/received_events",
125.      "type": "User",
126.      "site_admin": false
127.    },
128.    "created_at": "2014-05-23T08:40:56Z",
129.    "updated_at": "2014-05-23T08:40:56Z",
130.    "body": "We could put the scripts in the '.car', but then where would the JS compiler put it?"
131.  }
132. },
133.   "public": true,
134.   "created_at": "2014-05-23T08:40:56Z",
135.   "org": {
136.     "id": 579261,
137.     "login": "ceylon",
138.     "gravatar_id": "a38479e9dc888f68fb6911d4ce05d7cc",
139.     "url": "https://api.github.com/orgs/ceylon",
140.     "avatar_url": "https://avatars.githubusercontent.com/u/579261?"
141.   }
142. }

```

## **Work to:**

render analytically productive the **formatted, dynamic** character of digital networked **data**; (Marres & Weltevrede, 2013, 327)

render traffic between the **object** and **process** of social research analytically productive. ... the **formats and life cycles of online data** may lend structure to the analytic objects and findings of social research. (Marres and Weltevrede, 2013, p. 233)

## **In order to:**

develop ethnographic sensibility in data;

momentarily apprehend depth.

# **The device/platform and its events**

# Build software better, together.

FOUNDED  
2008HOME BASE  
San FranciscoHUBBERNAUTS  
239LOVEABLE MASCOT  
The Octocat

## About

[Team](#)[Jobs](#)[Press](#)[Blog](#)

### Looking for help?

Check out our documentation or [contact support](#).

### Get in touch

Our contact form goes to real people and we'd love to hear from you. [Contact us](#).

GitHub is the best place to [share code](#) with friends, co-workers, classmates, and complete strangers. Over five million people use GitHub to build amazing things together.



With the [collaborative features of GitHub.com](#), our [desktop and mobile apps](#), and [GitHub Enterprise](#), it has never been easier for individuals and teams to write better code, faster.

Originally founded by [Tom Preston-Werner](#), [Chris Wanstrath](#), and [PJ Hyett](#) to simplify sharing code, GitHub has grown into the largest code host in the world.





Search or type a command

Explore Gist Blog Help



rian39

+ X E



About GitHub

Our Team

Press

Jobs

There are **6.1M** people collaborating right now across **13.2M** repositories on GitHub. Developers from all around the world are building amazing things together. Their story is our story.

---

[GitHub: Innovations That Mattered in 2013](#)

GovLoop, December 12, 2013

[GitHub's Tom Preston-Werner: How We Went Mainstream](#)

ReadWriteWeb November 18, 2013

[press@github.com](mailto:press@github.com)

[GitHub Fact Sheet](#)

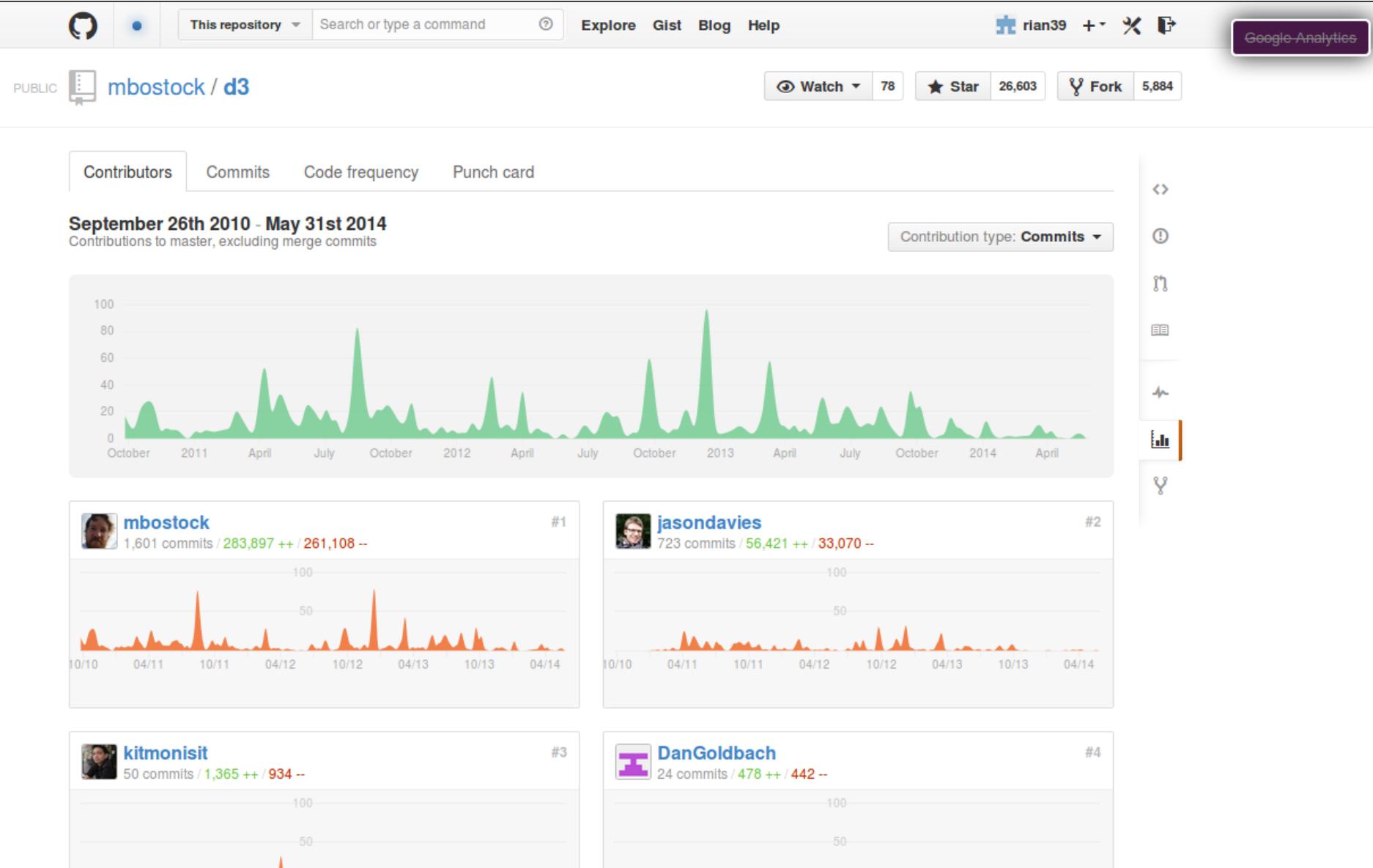
'Their story is our story'

# 13.2 million repositories: 'their story' as showcase

The screenshot shows the GitHub Explore page with the following sections:

- Explore GitHub** header with search bar and navigation links.
- All Showcases Trending Stars** navigation.
- Explore** section with the sub-instruction: "Browse interesting projects, solving all types of interesting problems."
- Web application frameworks** (blue box)
- Projects that power GitHub for Mac** (green box)
- Projects that power GitHub for Windows** (green box)
- NoSQL databases** (orange box)
- Design essentials** (red box)
- DevOps tools** (brown box)
- Package managers** (purple box) with a note: "Across programming languages and platforms, package managers make it easier to..." and stats: 635 14 6.
- Security** (teal box)
- See all >** link.
- Trending repositories this week** (flame icon) with a dropdown arrow.
- See all >** link.
- Repository cards:
  - tiimgreen/github-cheat-sheet** (3,699 stars, 227 forks)
  - cantino/huginn** (2,747 stars, 173 forks)
  - qrohlf/trianglify** (1,418 stars, 63 forks)
  - vhf/free-programming-books** (1,187 stars, 263 forks)
  - almandevis/vis** (900 stars, 51 forks)
  - Famous/famous** (843 stars, 113 forks)
- Starred by GitHub staff this week** (star icon) with a dropdown arrow.
- See all >** link.
- Repository cards:
  - cantino/huginn** (9 stars, 0 forks)
  - tiimgreen/github-cheat-sheet** (7 stars, 0 forks)
  - qrohlf/trianglify** (4 stars, 0 forks)
  - NYTimes/pourover** (3 stars, 0 forks)
  - tpope/heroku-binstubs** (3 stars, 0 forks)
  - Sirupsen/logrus** (2 stars, 0 forks)

# 'Their story': 1 repository at a time



# Publishing events: 210 million (1 March 2011 – today) : [www.githubarchive.org](http://www.githubarchive.org)



Open-source developers all over the world are working on millions of projects: writing code & documentation, fixing & submitting bugs, and so forth. GitHub Archive is a project to **record** the public GitHub timeline, **archive it**, and **make it easily accessible** for further analysis.

Looking for the [daily top new & watched repository](#) reports? Sign up here.



GitHub provides [18 event types](#), which range from new commits and fork events, to opening new tickets, commenting, and adding members to a project. The activity is aggregated in hourly archives, which you can access with any HTTP client:

| Query                                | Command  |
|--------------------------------------|--|
| Activity for April 11, 2012, 3PM UTC | <code>wget http://data.githubarchive.org/2012-04-11-15.json.gz</code>            |
| Activity for April 11, 2012          | <code>wget http://data.githubarchive.org/2012-04-11-{0..23}.json.gz</code>       |
| Activity for April 2012              | <code>wget http://data.githubarchive.org/2012-04-{01..31}-{0..23}.json.gz</code> |

*Note: timeline data is available starting February 12, 2011.*

Each archive contains a stream of JSON encoded GitHub events ([sample](#)), which you can process in any language. Ruby example:

```
1 require 'open-uri'
2 require 'zlib'
3 require 'yajl'
4
5 gz = open('https://data.githubarchive.org/2012-02-11-15.json.gz')
```

COMPOSE QUERY

Query History  
Job History

## Table Details: timeline

Schema Details Query Table

## Description

Describe this table...

## Table Info

|                |                               |
|----------------|-------------------------------|
| Table ID       | githubarchive:github.timeline |
| Table Size     | 112 GB                        |
| Number of Rows | 185,116,106                   |
| Creation Time  | 7:49am, 29 Apr 2012           |
| Last Modified  | 9:22am, 31 Mar 2014           |

## Devices re-format published events

## Preview

## Query Results 11:19am, 31 Mar 2014

| Row | created_at          | type              | actor          | repository_name        | url   |
|-----|---------------------|-------------------|----------------|------------------------|---|
| 1   | 2014-03-15 09:04:43 | PushEvent         | vvakame        | DefinitelyTyped        | <a href="https://github.com/borisyankov/DefinitelyTyped/compare/3d3832de3e...93a2313f62">https://github.com/borisyankov/DefinitelyTyped/compare/3d3832de3e...93a2313f62</a>                     |
| 2   | 2014-03-15 09:04:42 | PushEvent         | gizmomogwai    | plist                  | <a href="https://github.com/gizmomogwai/plist/compare/12eb82d283...83ad8b5f26">https://github.com/gizmomogwai/plist/compare/12eb82d283...83ad8b5f26</a>   |
| 3   | 2014-03-15 09:04:48 | IssueCommentEvent | ben-lin        | node.inflection        | <a href="https://github.com/dreamerslab/node.inflection/issues/16#issuecomment-37721104">https://github.com/dreamerslab/node.inflection/issues/16#issuecomment-37721104</a>                     |
| 4   | 2014-03-15 09:04:48 | WatchEvent        | cbmd           | mongofill              | <a href="https://github.com/koubas/mongofill">https://github.com/koubas/mongofill</a>   |
| 5   | 2014-03-15 09:04:48 | WatchEvent        | svett          | molokai                | <a href="https://github.com/tomasr/molokai">https://github.com/tomasr/molokai</a>   |
| 6   | 2014-03-15 09:04:47 | GollumEvent       | swordray       | wiki                   | <a href="https://github.com/ruby-china/wiki/wiki/RubyGems">https://github.com/ruby-china/wiki/wiki/RubyGems</a>   |
| 7   | 2014-03-15 09:04:46 | PushEvent         | elfet          | purephp                | <a href="https://github.com/elfet/purephp/compare/dc24b70f00...c94c524e76">https://github.com/elfet/purephp/compare/dc24b70f00...c94c524e76</a>   |
| 8   | 2014-03-15 09:04:46 | PullRequestEvent  | nikkypx        | rets_data              | <a href="https://github.com/arcticleo/rets_data/pull/2">https://github.com/arcticleo/rets_data/pull/2</a>   |
| 9   | 2014-03-15 09:04:46 | WatchEvent        | mjaneczek      | google-authenticator   | <a href="https://github.com/jaredonline/google-authenticator">https://github.com/jaredonline/google-authenticator</a>   |
| 10  | 2014-03-15 09:04:53 | PushEvent         | micmath        | Rye                    | <a href="https://github.com/micmath/Rye/compare/69c02b1aea...7885ea9e36">https://github.com/micmath/Rye/compare/69c02b1aea...7885ea9e36</a>   |
| 11  | 2014-03-15 09:04:53 | PushEvent         | fitret         | daigon                 | <a href="https://github.com/fitret/daigon/compare/8897fa9404...77e5f48276">https://github.com/fitret/daigon/compare/8897fa9404...77e5f48276</a>   |
| 12  | 2014-03-15 09:04:53 | WatchEvent        | joeyates       | spork                  | <a href="https://github.com/sporkrb/spork">https://github.com/sporkrb/spork</a>   |
| 13  | 2014-03-15 09:04:53 | CreateEvent       | grcnva         | mcs-chatboard          | <a href="https://github.com/grcnva/mcs-chatboard">https://github.com/grcnva/mcs-chatboard</a>   |
| 14  | 2014-03-15 09:04:51 | PushEvent         | albatrossen    | bungeebouncer          | <a href="https://github.com/albatrossen/bungeebouncer/compare/57a3771590...7794ffb531">https://github.com/albatrossen/bungeebouncer/compare/57a3771590...7794ffb531</a>                         |
| 15  | 2014-03-15 09:04:51 | PushEvent         | catalinstanciu | ChessEngine-Xboard-cpp | <a href="https://github.com/catalinstanciu/ChessEngine-Xboard-cpp/compare/9bedce1744...8b98d2ff45">https://github.com/catalinstanciu/ChessEngine-Xboard-cpp/compare/9bedce1744...8b98d2ff45</a> |

COMPOSE QUERY

## New Query

?

X

Query History

Job History

githubarchive

▶ github

▶ github\_explore

▶ github\_proper

▶ stack\_overflow

```

1 SELECT created_at, type, actor, repository_name, url FROM [githubarchive:github.timeline]
2 WHERE regexp_match(repository_name, 'cloud')
3 LIMIT 50

```

**RUN QUERY**

Save Query

Save View

Enable Options

Query complete (3.0s elapsed, 23.2 GB processed)



|                   |                                |                        |   |
|-------------------|--------------------------------|------------------------|---|
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| IssueCommentEvent | buildhive                      | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516027">https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516027</a>         |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| IssueCommentEvent | buildhive                      | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516091">https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516091</a>         |
| IssueCommentEvent | annashipman                    | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516194">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516194</a>               |
| PushEvent         | srunner                        | tilecloud              | <a href="https://github.com/srunner/tilecloud/compare/185a371ce5...7d69504ad1">https://github.com/srunner/tilecloud/compare/185a371ce5...7d69504ad1</a>                                 |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PushEvent         | maksimov                       | dasein-cloud-core      | <a href="https://github.com/maksimov/dasein-cloud-core/compare/90827bb84c...ca08115701">https://github.com/maksimov/dasein-cloud-core/compare/90827bb84c...ca08115701</a>               |
| PushEvent         | cyrichardson                   | docs-cloud-files       | <a href="https://github.com/rackerlabs/docs-cloud-files/compare/75bc24a55e...6f1a9daa3c">https://github.com/rackerlabs/docs-cloud-files/compare/75bc24a55e...6f1a9daa3c</a>             |
| PushEvent         | laszewsk                       | cloudmesh              | <a href="https://github.com/cloudmesh/cloudmesh/compare/1b327284fb...abcdabd325">https://github.com/cloudmesh/cloudmesh/compare/1b327284fb...abcdabd325</a>                             |
| IssueCommentEvent | cloudbees-pull-request-builder | jclouds-labs           | <a href="https://github.com/jclouds/jclouds-labs/pull/62#issuecomment-43516613">https://github.com/jclouds/jclouds-labs/pull/62#issuecomment-43516613</a>                               |
| IssueCommentEvent | mattbostock                    | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516643">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516643</a>               |
| CreateEvent       | jabley                         | vcloud-walker          | <a href="https://github.com/alphagov/vcloud-walker/compare/migrate-fog-to-core">https://github.com/alphagov/vcloud-walker/compare/migrate-fog-to-core</a>                               |
| IssueCommentEvent | mikepea                        | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516718">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516718</a>               |
| PushEvent         | dankilman                      | cloudify-manager       | <a href="https://github.com/cloudify-cosmo/cloudify-manager/compare/813efdc5e...c909eb45ee">https://github.com/cloudify-cosmo/cloudify-manager/compare/813efdc5e...c909eb45ee</a>       |
| PullRequestEvent  | maksimov                       | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/pull/192">https://github.com/greese/dasein-cloud-core/pull/192</a>   |
| IssuesEvent       | duglin                         | cloud_controller_ng    | <a href="https://github.com/cloudfoundry/cloud_controller_ng/issues/246">https://github.com/cloudfoundry/cloud_controller_ng/issues/246</a>   |
| IssueCommentEvent | cf-gitbot                      | cloud_controller_ng    | <a href="https://github.com/cloudfoundry/cloud_controller_ng/issues/246#issuecomment-43516802">https://github.com/cloudfoundry/cloud_controller_ng/issues/246#issuecomment-43516802</a> |
| PushEvent         | andylyall                      | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/compare/65a45c29c0...208773ac2e">https://github.com/greese/dasein-cloud-core/compare/65a45c29c0...208773ac2e</a>                   |
| PullRequestEvent  | andylyall                      | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/pull/192">https://github.com/greese/dasein-cloud-core/pull/192</a>   |
| PullRequestEvent  | jabley                         | vcloud-walker          | <a href="https://github.com/alphagov/vcloud-walker/pull/76">https://github.com/alphagov/vcloud-walker/pull/76</a>   |
| PushEvent         | BrunoMatry                     | cloudbook              | <a href="https://github.com/BrunoMatry/cloudbook/compare/58d7be347a...03d3225351">https://github.com/BrunoMatry/cloudbook/compare/58d7be347a...03d3225351</a>                           |

# **The datastream and its platform stories**

# 'Countless stories to tell'

Search or type a command

Explore Gist Blog Help

rian39 + X ↗ Google Analytics

## The GitHub Data Challenge

May 1, 2012 briandoll Watercooler

The GitHub public timeline is [now easy to query and analyze](#). With hundreds of thousands of events in the timeline every day, there are countless stories to tell.



Featured

All Posts

New Features

Engineering

Enterprise

Meetups

New Hires

**Watercooler**

Subscribe

<https://camo.githubusercontent.com/5e6eb0b00d714eb5b8ec84254205c61c2a97c68d/687474703a2f2f6f63746f6465782e6769746875622e636f6d2f696d616765732f77616c646f6361742e6a7067>

# 'What stories will be told this year? Its up to you'

The screenshot shows a GitHub repository page for 'The GitHub Data Challenge II'. The page has a header with navigation links for Explore, Gist, Blog, and Help. A search bar is at the top right. On the left, there's a sidebar with a 'Featured' section containing links to All Posts, New Features, Engineering, Enterprise, Meetups, New Hires, and Watercooler (which is highlighted). The main content area features a post by briandoll from April 3, 2013, about GitHub activity. It includes a sidebar with a 'Subscribe' button.

**The GitHub Data Challenge II**

April 3, 2013   briandoll   Watercooler

There are millions of projects on GitHub. Every day, people from around the world are working to make these projects better. Opening issues, pushing code, submitting Pull Requests, discussing project details — GitHub activity is a papertrail of progress. Have you ever wondered what all that data looks like? There are millions of stories to tell; you just have to look.

Last year we held our [first data challenge](#). We saw [incredible visualizations](#), [interesting timelines](#) and [compelling analysis](#).

What stories will be told this year? It's up to you!

**To Enter**

Send a link to a GitHub repository or gist with your graph(s) along with a description to [data@github.com](mailto:data@github.com) before **midnight, May 8th, 2013 PST**.

**Data access**

The GitHub public timeline is a featured public dataset available on [Google BigQuery](#). The "timeline" table has over a year's worth of public activity and is approaching 100M rows. You can find even more data available in JSON format on [The GitHub Archive project](#).

You are free to use any tools you like. If you choose to use BigQuery, running queries against the GitHub dataset is free for the first 100GB of query processing. Pricing information for additional query processing is [available here](#). After [signing up for BigQuery](#), add the project name "githubarchive".

# Stories: individuals and their similarities

BuySellAds  
Google Analytics

Dear recruiters:

While you read this, make sure that you remember that GitHub is not your CV, and that these stats only provide a *biased and one-sided view*. This is just a toy. Don't take it too seriously!

OK, I promise!

Powered by Fusion

THE  
OPEN SOURCE  
REPORT CARD

Enter a GitHub username to see a dynamically generated progress report for their open source contributions

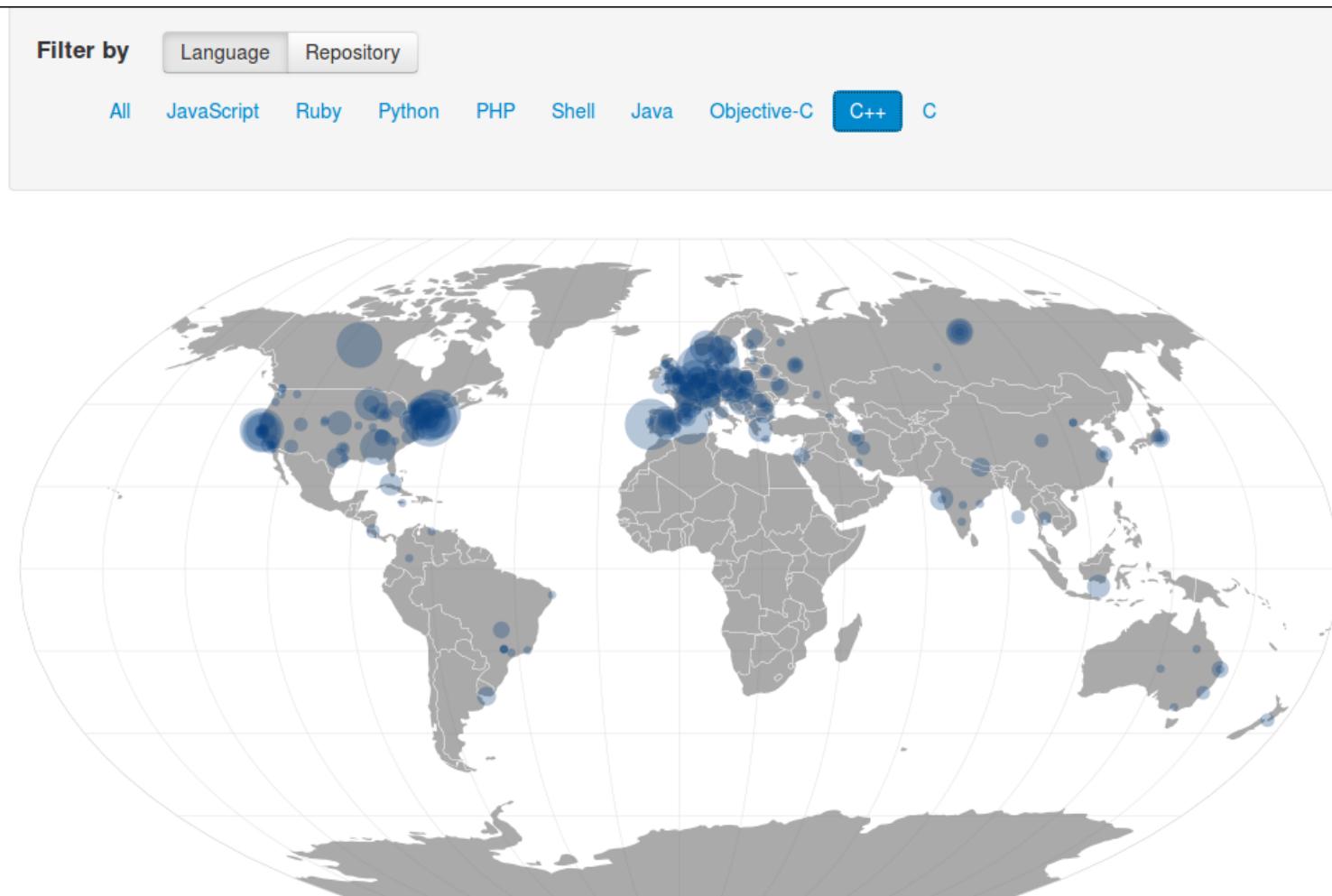
Enter a GitHub username

HOW IT WORKS

Every day, many thousands of open source contributions are made on GitHub by developers around the world. This data is publicly available through the API and—even more conveniently—on the GitHub Archive. This is generally a pretty fun dataset to play with but it is particularly exciting for hackers because we get to play with data that describes *our own behavior!* Last year, shortly after the full event stream was publicly released, the first annual GitHub data challenge produced some sick data visualizations and it's clear that people at GitHub have been thinking about how to Use The Data For Good™.

The one graph that is especially awesome in all sorts of surprising ways is the contributions heat map on every user's profile page. What sets this apart from the other visualizations that already exist on the site? It makes a general statement about one specific

# Repositories and their places



GitHub's [public timeline](#) summarizes the open source work of millions of developers the world over.

This data set contains contributions to the top 200 GitHub repositories during the first four months of



# Exploring Expressions of Emotions in GitHub Commit Messages



Last week [GitHub announced](#) that the data collected over at the [GitHub Archive](#) was made available as a public dataset on [Google's BigQuery web service](#). Pretty exciting news for data analysts, considering that the timeline dataset currently contains more than 7 million records and is growing quickly.

To see what developers come up with analyzing the GitHub timeline and language correlation datasets, they started [a data challenge](#) which is open for submissions until May 21st.

Being interested in both natural and programming languages, I decided to expand on this analysis of [profanity in git commit messages per programming language](#) with a more general exploration of expressions of emotions as well as looking at messages indicating issues and those containing swear words.

## Method of Collecting and Analyzing Data

Reading millions of commit messages and categorizing them is certainly not a task I'd be able to accomplish in less than 3 weeks and honestly wouldn't want to. Instead, I assembled lists of words expressing emotions using the resources listed at the [bottom of this article](#) and wrote corresponding database queries to count occurrences of these words grouped by programming languages.

Then for each language I calculated the percentage of messages that meet the criteria, i. e. contain certain expressions, in relation to the total number of commits for that language. Only non-empty commit messages of projects with an assigned language were included.

Since the number of commits varies considerably across different languages only those with at least 40,000

## Geek Shirts

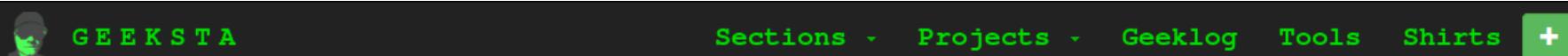
Get the [Bash fork bomb shirt](#) for only \$14 for a Hanes tagless tee. Limited until May 20th, 2014.



Create your personalized [born at unix time t-shirt](#).

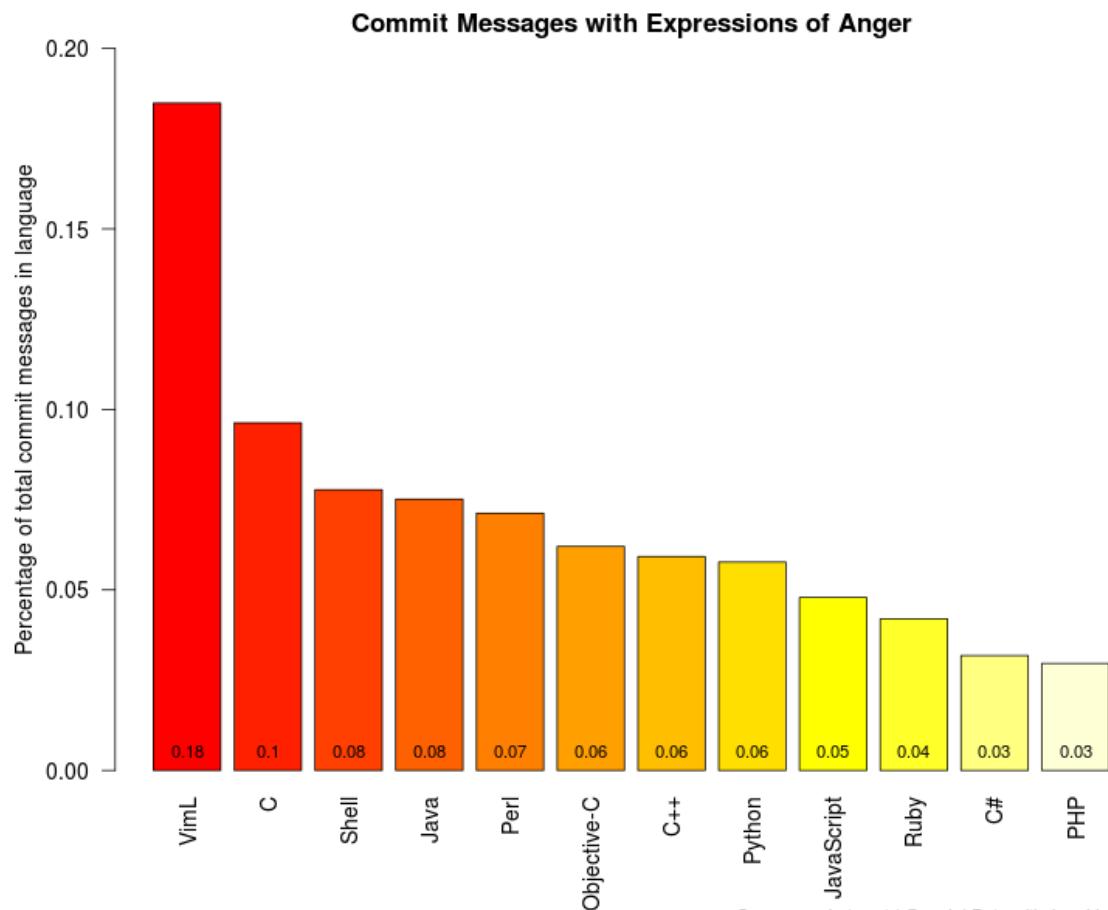


# Affects into repositories



Regular Expression

```
(?i)\b(a+rgh|angry|annoyed|annoying|appalled|bitter|cranky|hate|hating|mad)\b'
```



# Live reports of event counts

Octoboard

Global activity   Repository   Public   Fork   Pull Request   Issue   GH PR   GitHub Search   Google Analytics

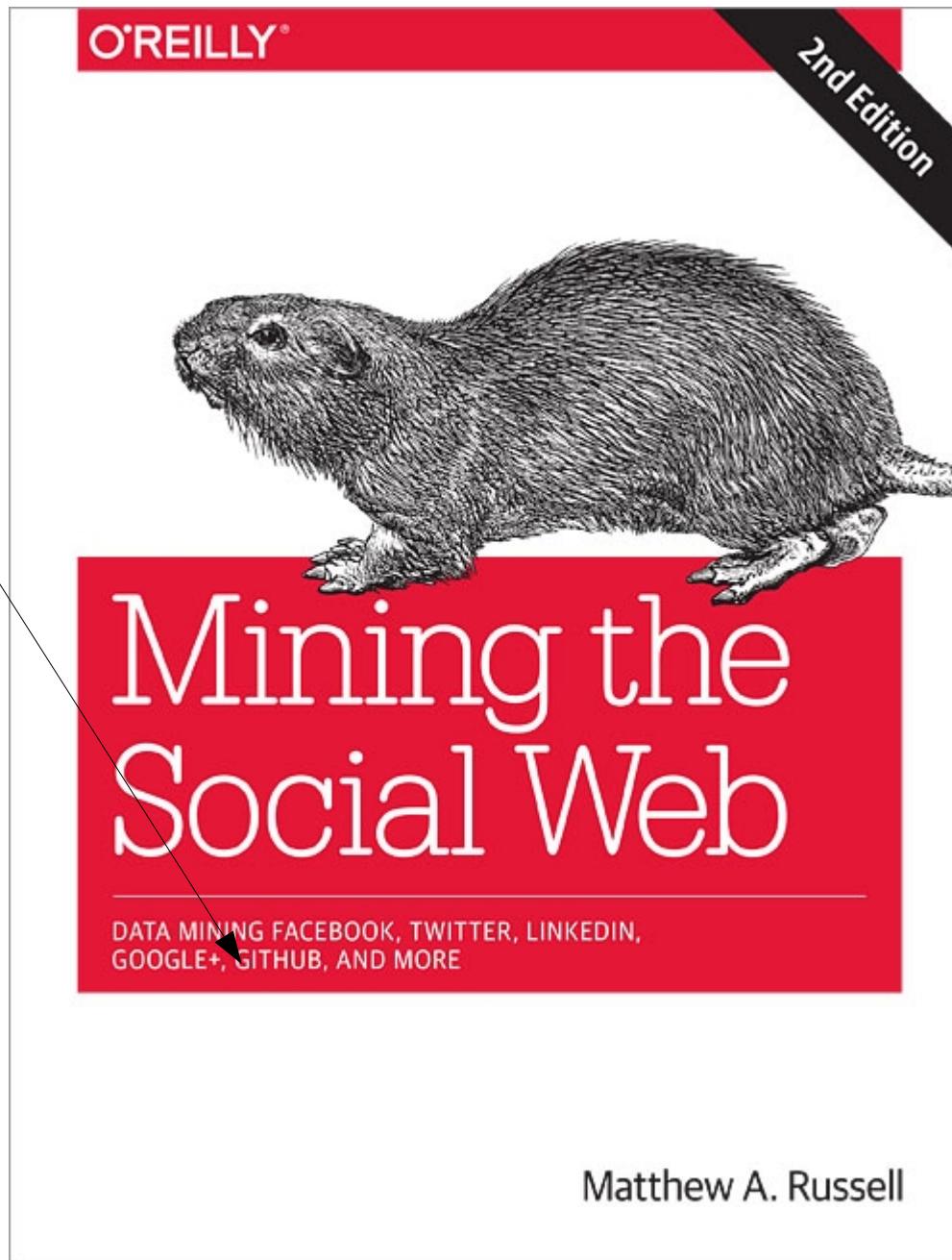
## GitHub activity dashboard.

Octoboard is based on GitHub Archive : each day, it scans new GitHub events archives and computes a few stats, with a 15 days history. You can see some general data on this page, or use menu for more information about language and history.

Since march 11th, 2012 :

|   |  |   |  |
|---|--|---|--|
| <b>Global Activity</b><br>158,547,611<br>+ 198,185 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 88% above average. | <b>Created Repositories</b><br>7,758,991<br>+ 9,699 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 81% above average. | <b>Open Sourced Repositories</b><br>149,497<br>+ 187 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 66% above average. | <b>Forked Repositories</b><br>5,931,778<br>+ 7,415 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 103% above average. |
| <b>Opened Pull Requests</b><br>501,755<br>+ 2,023 / day<br><br>Yesterday<br><br>It was a <b>bad</b> day, they were down 26% below average.      | <b>Opened Issues</b><br>5,474,375<br>+ 6,843 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 89% above average.        | <b>New GitHub Pages</b><br>311,154<br>+ 389 / day<br><br>Yesterday<br><br>It was a <b>very good</b> day, they were up 89% above average.          | <b>Created Gists</b><br>2,374,309<br>+ 3,005 / day<br><br>Yesterday<br><br>It was a <b>good</b> day, they were up 24% above average.             |

**Github data as  
examplar for  
coding work**



**Unformatted imitation in the data stream**

# *names in datastreams*

'Consider for a moment the more general conditions of naming. First, a name is offered, given, imposed by someone or by some set of someones, and it is attributed to someone else. It requires an intersubjective context, but also a mode of address, for the name emerges as the addressing of a coinage to another, and in that address, a rendering of that coinage proper' (Butler, 1997: 29)

## Names in STS

'it is important to explore the variety in current naming traditions in terms of what they do, the impact of the speech acts involved, integrating analyses of such technologies into studies of the constitution of subjects, the politics of belonging, exclusion, and control' (Palsson, 2014: 628).

```
repo  
maxkirchoff/google-music-dupe-killer  
MaratKarimov/TwitterClient  
seejay/xgoogle  
canercandan/mini-google-in-perl  
chengdujin/Mining-the-Social-Web  
twinslash/omniauth-vkontakte  
Deminem/google-glass-mirror-ruby-sinatra-scaffold  
tfountain/oauth-2---facebook---zend-framework-...  
alesaudate/camel-twitter-websockets  
soequelle/facebook-iphone-sdk  
shadowmaru/monitor-twitter  
Hiroki-Takayama/python-flask-social-oauth-face...  
davcro/facebook_client  
pratyush-nigam/Learning-RoR  
jabranr/instagram-api-php
```

dpackcartonmachine/3layer-5layer-7layer-corrug...  
dpackcartonmachine/corrugated-box-machine--817258  
dpackcartonmachine/corrugated-box-machine--854402  
dpackcartonmachine/corrugated-box-machine--709152  
dpackcartonmachine/5-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/3-layers-corrugated-box-mac...  
dpackcartonmachine/3-layers-corrugated-box-mac...  
dpackcartonmachine/5-layers-corrugated-box-mac...  
dpackcartonmachine/corrugated-box-machine--529003  
dpackcartonmachine/corrugated-box-machine--058569  
dpackcartonmachine/corrugated-box-machine--564559  
dpackcartonmachine/semi-automatic-four-color-p...  
dpackcartonmachine/3layer-5layer-7layer-Corrug...  
dpackcartonmachine/3layer-corrugated-box-machi...  
dpackcartonmachine/3layer5layer7layer-corrugat...  
dpackcartonmachine/corrugated-box-machine--408286  
dpackcartonmachine/3layer-5layer-7layer-Corrug...  
dpackcartonmachine/corrugated-box-machine--952283  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac... 4

# Copied names as affective imitation: forks, mirrors

FORKCOUNT: **4,468128** (approximately 4.5 million repos are forks)

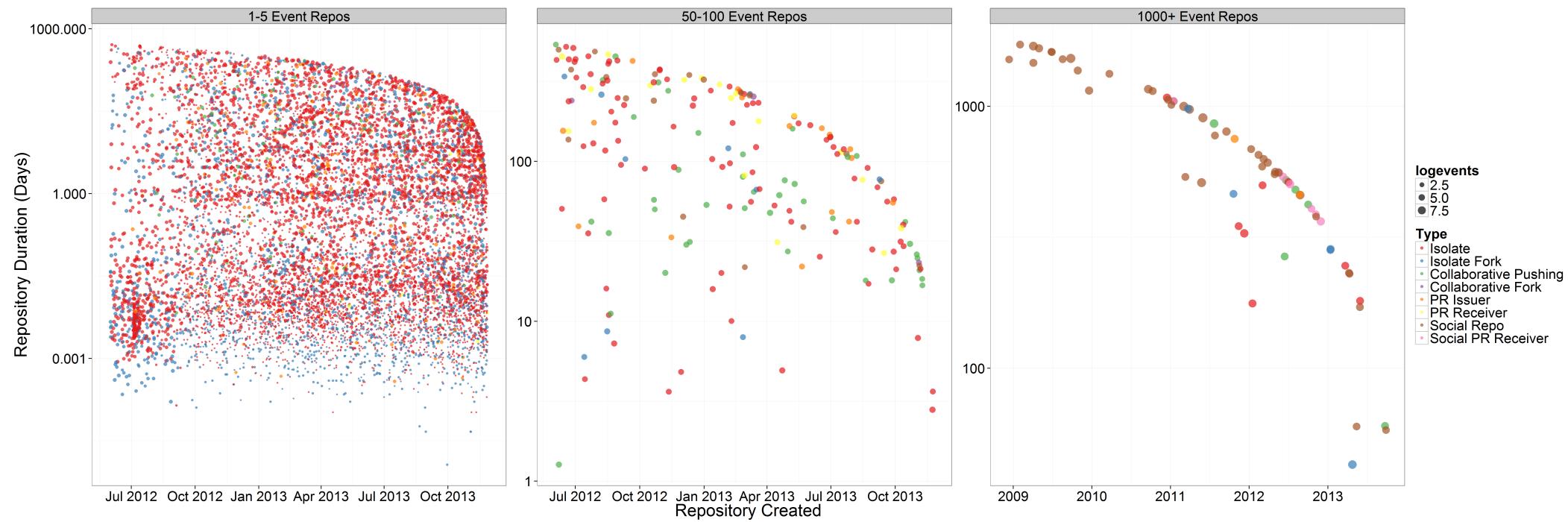
```
SELECT count(*) as fork_count FROM [metacommunities:github_proper.repo_list]  
where fork=1
```

UNIQUE NAMES: **5,323,230** (approximately 5.3 million repos have unique names, so 7 million names are imitations)

```
SELECT count(distinct(repository_name)) as name_count FROM  
[githubarchive:github.timeline] LIMIT 1000;
```

'incorporation into collectivities involves **affective imitation**' (Gatens & Lloyd, 1999)

## 13.2 million repositories, but mostly imitative incorporations



# 'passionate imitation'

## Names as bundles

repository\_url  
https://github.com/mantrala/twitter\_clone  
https://github.com/estebistec/django-twitter-bootstrap  
https://github.com/notonalot/LoL-twitter-stats  
https://github.com/sherzberg/twitter-stream-learning  
https://github.com/RouxRC/TwitterApiDocumentationRSS  
https://github.com/stephenturner/twitterarchive  
https://github.com/imausrokihi/twitter-activator  
https://github.com/sferik/sign-in-with-twitter  
https://github.com/konstandino-a/twitter-analyser  
https://github.com/thecodepath/ios\_twitter  
https://github.com/jacurtis/twitter\_clone  
https://github.com/nst/STTwitter  
https://github.com/Nodejs-Colorado/node-twitter-sentiment-promises  
https://github.com/mjhea0/node-twitter-sentiment-generators  
https://github.com/spiritdead/Twitterizer  
https://github.com/mjhea0/node-twitter-sentiment  
https://github.com/tonytonyjan/ruby-conf-tw-2014-twitter-wall  
https://github.com/mjhea0/node-twitter-sentiment-promises  
https://github.com/yusuke/twitter4j  
https://github.com/margaritoCajero/twitter  
https://github.com/kathrynsiegel/twitterbot  
https://github.com/cwhittemore1/TwitterAPI  
https://github.com/marek5050/TwitterScrap  
https://github.com/AvianFlu/ntwitter  
https://github.com/jamesfebin/meteor-twitter-api  
https://github.com/mrbellek/twitterbot  
https://github.com/tchu88/diff-twitter

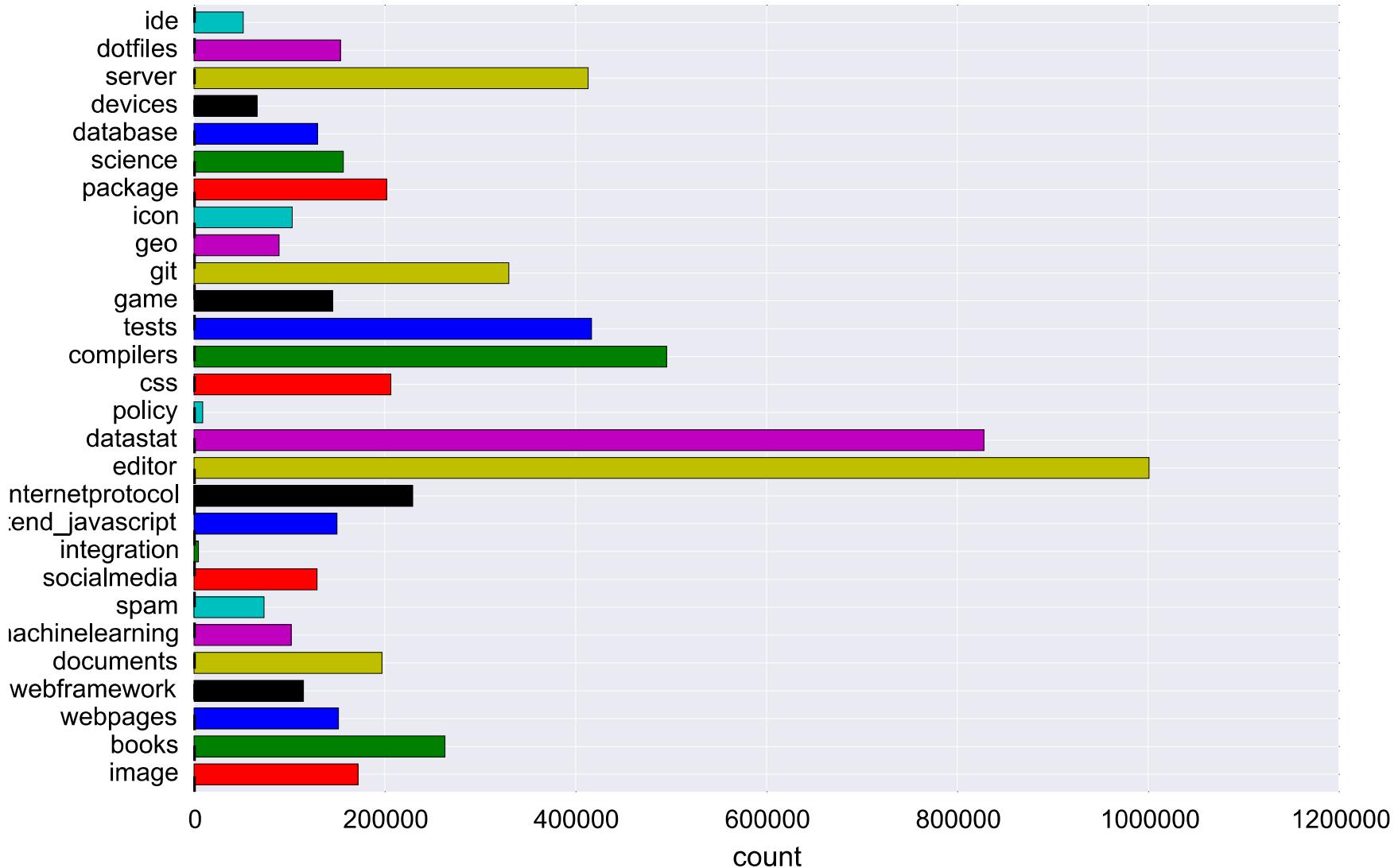
## Most common repository names

|                 |       |
|-----------------|-------|
| test            | 85404 |
| try_git         | 67393 |
| dotfiles        | 39928 |
| first_app       | 34772 |
| sample_app      | 33848 |
| demo_app        | 30076 |
| bootstrap       | 25814 |
| spoon-knife     | 24603 |
| hello-world     | 19989 |
| bitstarter      | 13076 |
| blog            | 11392 |
| helloworld      | 10362 |
| android         | 9722  |
| homebrew        | 8305  |
| rails           | 8281  |
| jquery          | 7540  |
| angular.js      | 6778  |
| myrepo          | 6555  |
| node            | 6499  |
| testrepo        | 6320  |
| html5-boilerpla | 6188  |
| scripts         | 6109  |
| test1           | 6054  |
| demo            | 5890  |
| projects        | 5718  |
| git             | 5706  |
| github          | 5681  |
| phonegap-star   | 5667  |
| testing         | 5494  |
| sandbox         | 5474  |

the importance of suggestion and imitation in social life, and the limits of forms of analysis which confine themselves to the study of representation or inscription. (Barry and Thrift, 2007: 511)

# Namespace as device-specific imitation

Repository topics on Github



# Querying device-specific affective imitation

IP[y]: Notebook

1a\_Generate\_repo\_topic\_data Last Checkpoint: May 30 10:28 (autosaved)

File Edit View Insert Cell Kernel Help



Cell Toolbar: None

```
vz|students circle network|stumbleupon>tagged|talkbiznow|taltopia|taringa!|teachstreet|termwiki|the sphere|travbuddy.com|travellers point|tribe.net|trombi.com|tuenti|tumblr|twitter|tylted|vkontakte|vampirefreaks.com|viadeo|virb|vox|wattpad|wayn|weeworld|we heart it|wellwer|weourfamily|wepolls.com|wer-kennt-wen|weread|wiser.org|wooxie|writeaprisoner.com|xanga|xing|xt3|yammer|yelp, inc.|zoo.gr|zooppa
```

In [91]: `socialmedia_df = construct_repo_df(socialmediaquery, description = False)`

```
starting bigquery ...
wikipedia|43.7things|academia.edu|about.me|advogato|anobii|asianavenue|asmallworld|athlinks|audimated.com|bebo|biip.no|blackplanet|blauk|blogster|bolt.com|busuu|buzznet|cafemom|care2|caringbridge|classmates.com|cloob|couchsurfing|cozcot|cross.tv|crunchyroll|cyworld|dailybooth|dailystrength|delicious|deviantart|diaspora*|didlr|disaboom|dol2day|dontstayin|draugiem.lv|douban|dxy.cn|elftown|elixio|english.?baby!|epernicus|eons.com|etoro|experience.?project|exploroo|facebook|faceparty|faces.com|fetlife|filmaffinity|filmow|fledgeling|flixster|flickr|focus.com|formspring|fotki|fotolog|foursquare|friendica|friends.?reunited|friendster|frühstückstreff|fuelmyblog|fullcircle|gaia.?online|gamerdna|gapyear.com|gather.com|gays.com|geni.com|getglue|gogoyoko|goodreads|goodwizz|google.?govloop|grono.net|habbo|hi5|hospitality.?club|hotlist|hr.com|hub.?culture|hyves|ibibo|identi.ca|indaba.?music|instagram|irc.?galleria|italki.com|itsmy|iwiw|jaiku|jiepang|kaixin001|kiwibox|lafango|laibhaar|last.fm|librarything|lifeknot|linkedin|linkxpats|listograph|ylivejournal|livemocha|makeoutclub|meetin|meetup.?website|meettheboss|millatfacebook|mixi|mocospace|mog|mouthshut.com|mubi|myheritage|mylife|my.?opera|myspace|nasza.?klasa.pl|netlog|nexopia|ngo.?post|ning|odnoklassniki|open.?diary|orkut|outeverywhere|patientslike|partyflock|pingsta|pinterest|plaxo|playfire|playlist.com|plurk|poolwo|qapacity|quechup|qzone|raptr|ravelry|renren|reverbnation.com|ryze|sciencestage|sgroups|sharethemusic|shelfari|sina.?weibo|skoob|skyrock|socialvibe|sonico.com|soundcloud|spaces|stage.?32|stickam|studivz|students.?circle.?network|stumbleupon>tagged|talkbiznow|taltopia|taringa!|teachstreet|termwiki|the.?sphere|travbuddy.com|travellerspoint|tribe.net|trombi.com|tuenti|tumblr|twitter|tylted|vkontakte|vampirefreaks.com|viadeo|virb|vox|wattpad|wayn|weeworld|we.?heart.?it|wellwer|weourfamily|wepolls.com|wer.?kennt.?wen|weread|wiser.org|wooxie|writeaprisoner.com|xanga|xing|xt3|yammer|yelp.?inc.|zoo.gr|zooppa
```

Waiting on bqjob\_r1bc3395151b160a\_00000146203dd18f\_21 ... (0s) Current status: DONE  
(156058, 3)

In [148]: `red.sadd('social_media', *socialmedia_df.full_name)`

Out[148]: 156050

Almost 157k repositories here. Is that mostly noise?

In [93]: `socialmedia_df.head(30)`

Out[93]:

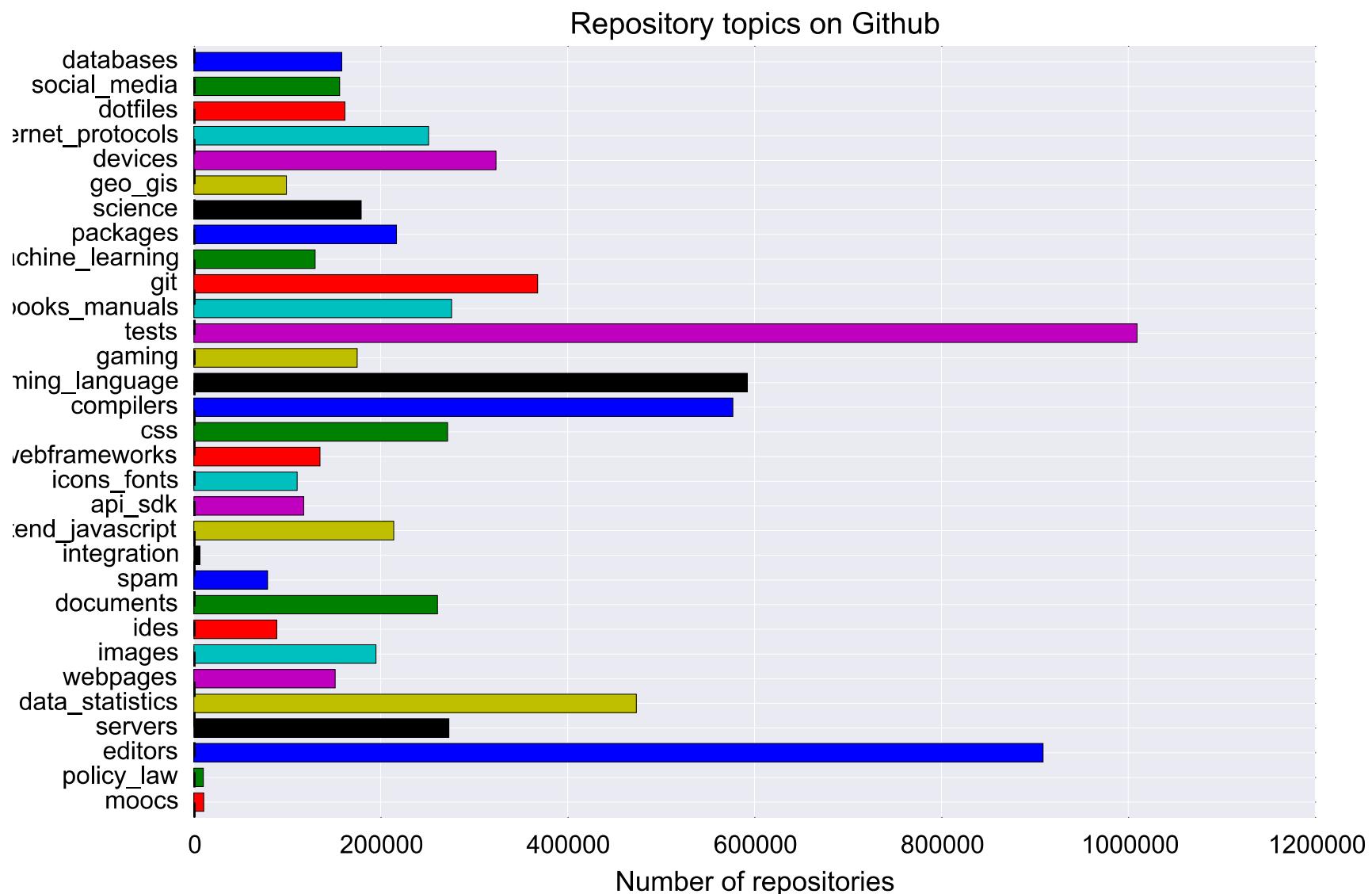
|   | name  | full_name | fork |
|---|-------|-----------|------|
| 0 | Blauk | Blauk     | 1    |

```
In [93]: socialmedia_df.head(30)
```

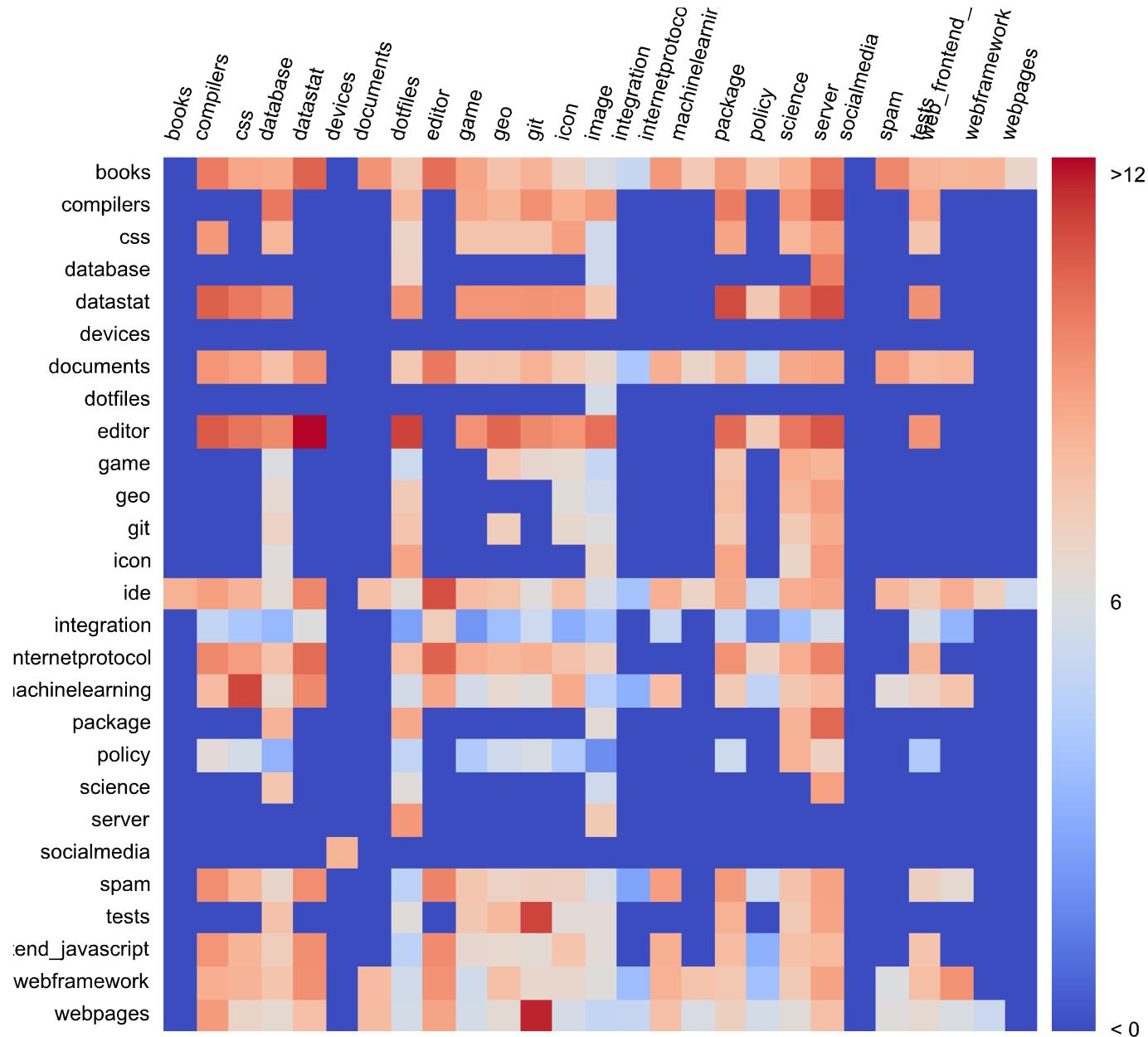
Out[93]:

|           | <b>name</b>                | <b>full_name</b>                      | <b>fork</b> |
|-----------|----------------------------|---------------------------------------|-------------|
| <b>0</b>  | PlurkCSS                   | howar31/PlurkCSS                      | 1           |
| <b>1</b>  | ZXing.Net.Mobile           | JerryLee7809/ZXing.Net.Mobile         | 1           |
| <b>2</b>  | twitteroauth               | mathijsvdhurk/twitteroauth            | 1           |
| <b>3</b>  | Recipes-for-Mining-Twitter | winnerdy/Recipes-for-Mining-Twitter   | 1           |
| <b>4</b>  | DMAActivityInstagram       | PrincessThan/DMAActivityInstagram     | 1           |
| <b>5</b>  | wikipedia_api              | gpherguson/wikipedia_api              | 0           |
| <b>6</b>  | facebook                   | harjeetsingh111/facebook              | 0           |
| <b>7</b>  | symphony-learning-download | teradotact/symphony-learning-download | 0           |
| <b>8</b>  | facebook-android-sdk       | slott/facebook-android-sdk            | 1           |
| <b>9</b>  | facebook-android-sdk       | derfshaya/facebook-android-sdk        | 1           |
| <b>10</b> | SwanFlickrUploader         | entercritical/SwanFlickrUploader      | 0           |
| <b>11</b> | iOSTraining                | tokyoaret/iOSTraining                 | 1           |
| <b>12</b> | angular-google-maps        | adato/angular-google-maps             | 1           |
| <b>13</b> | facebook-php-sdk           | cabaltho/facebook-php-sdk             | 1           |

# Unstable trajectories in the namespace



# Variations and inventions through imitation



# Recursively device-specific social research and its affective imitations

Screenshot of a GitHub search results page for the query "redis".

The search bar contains "redis".

The sidebar shows statistics: 7,292 repositories, 1,511,276 code files, 28,757 issues, and 56 users.

The main heading is "We've found 7,292 repository results".

The top result is "antirez/redis".

Description: Redis is an in-memory database that persists on disk. The data model is key-value, but many different kind of values are supported: Strings, Lists, Sets, Sorted Sets, Hashes.

Last updated on May 26.

Sort: Best match

## **Conclusion: datastream stories**

'their story is our story': repeating the platform's formatting

'countless stories to tell': mirroring the platform

'what stories will be told this year?': datastream as platform

# Conclusions

**Latour**: 'Let the agents produce a dynamics and collect the traces that their actions leave as they unfold so as to produce a rich data set'  
(Latour et al., 2012):

**Verran**: work on numbers (especially one-many/part-whole tensions)

**Marres/Weltevreden**: work on data formats to heighten ambivalent tensions between device-specific and social process

**Gatens/Barry/Thrift/Tarde**: Affective/passionate imitation and incorporation in collectives generates invention

**Riles**: momentary apprehensions of depth

# Acknowledgements

Research funded by ESRC (UK Economic and Social Research Council) under 'Google Data Analytics' programme, 2013-2014

Richard Mills and Stuart Sharples – research associates

Matthew Fuller and Andrew Goffey – co-researchers

# **Device-specific research and data stream events\***

Adrian Mackenzie  
Sociology, Lancaster University, UK

12 June 2014

\* “SELECT count(type) as event\_count FROM [githubarchive:github.timeline]”

**207 million events**

In any practical going-on with numbers, what matters is that they can be *made* to work, and *making them work* is a politics. Yet is a politics that completely evades conventional foundationist analysis (Verran, 2001: 88)

'momentary apprehensions of depth' (Anne-Lise Riles, 2001)

## The datastream

<http://api.github.com/events>

## Simple event?

```
{  
1. "id": "2111998059",  
2. "type": "WatchEvent",  
3. "actor": {  
4.   "id": 1459103,  
5.   "login": "mmemetea",  
6.   "gravatar_id": "4532d1e4885f579ca7d9aa8748418817",  
7.   "url": "https://api.github.com/users/mmemetea",  
8.   "avatar_url": "https://avatars.githubusercontent.com/u/1459103?"  
9. },  
10. "repo": {  
11.   "id": 14802742,  
12.   "name": "OpenSensorsIO/azondi",  
13.   "url": "https://api.github.com/repos/OpenSensorsIO/azondi"  
14. },  
15. "payload": {  
16.   "action": "started"  
17. },  
18. "public": true,  
19. "created_at": "2014-05-23T08:40:56Z",  
20. "org": {  
21.   "id": 5497318,  
22.   "login": "OpenSensorsIO",  
23.   "gravatar_id": "1e0218942846ec8ef59f5d679dbca782",  
24.   "url": "https://api.github.com/orgs/OpenSensorsIO",  
25.   "avatar_url": "https://avatars.githubusercontent.com/u/5497318?"  
    },  
},
```

# Event depth?

```
1. {
2.   "id": "2111998057",
3.   "type": "IssueCommentEvent",
4.   "actor": {
5.     "id": 179265,
6.     "login": "FroMage",
7.     "gravatar_id": "b8932676ae05ecdd084d3839d1f49da8",
8.     "url": "https://api.github.com/users/FroMage",
9.     "avatar_url": "https://avatars.githubusercontent.com/u/179265?"
10.    },
11.   "repo": {
12.     "id": 1287859,
13.     "name": "ceylon/ceylon-compiler",
14.     "url": "https://api.github.com/repos/ceylon/ceylon-compiler"
15.   },
16.   "payload": {
17.     "action": "created",
18.     "issue": {
19.       "url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490",
20.       "labels_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/labels{/name}",
21.       "comments_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/comments",
22.       "events_url": "https://api.github.com/repos/ceylon/ceylon-compiler/issues/1490/events",
23.       "html_url": "https://github.com/ceylon/ceylon-compiler/issues/1490",
24.       "id": 23820110,
25.       "number": 1490,
26.       "title": "CLI plugins: support distribution",
```

```
26.  
27. "user": {  
28.   "login": "FroMage",  
29.   "id": 179265,  
30.   "avatar_url": "https://avatars.githubusercontent.com/u/179265?",  
31.   "gravatar_id": "b8932676ae05ecdd084d3839d1f49da8",  
32.   "url": "https://api.github.com/users/FroMage",  
33.   "html_url": "https://github.com/FroMage",  
34.   "followers_url": "https://api.github.com/users/FroMage/followers",  
35.   "following_url": "https://api.github.com/users/FroMage/following{/other_user}",  
36.   "gists_url": "https://api.github.com/users/FroMage/gists{/gist_id}",  
37.   "starred_url": "https://api.github.com/users/FroMage/starred{/owner}{/repo}",  
38.   "subscriptions_url": "https://api.github.com/users/FroMage/subscriptions",  
39.   "organizations_url": "https://api.github.com/users/FroMage/orgs",  
40.   "repos_url": "https://api.github.com/users/FroMage/repos",  
41.   "events_url": "https://api.github.com/users/FroMage/events{/privacy}",  
42.   "received_events_url": "https://api.github.com/users/FroMage/received_events",  
43.   "type": "User",  
44.   "site_admin": false  
45. },  
46. "labels": [  
47.   {  
48.     "url": "https://api.github.com/repos/ceylon/ceylon-compiler/labels/FEATURE",  
49.     "name": "FEATURE",  
50.     "color": "02e10c"  
51.   },
```



**Work to:**

render analytically productive the **formatted, dynamic** character of digital networked **data**; (Marres & Weltevrede, 2013, 327)

render traffic between the **object** and **process** of social research analytically productive. .... the **formats and life cycles of online data** may lend structure to the analytic objects and findings of social research. (Marres and Weltevrede, 2013, p. 233)

**In order to:**

develop ethnographic sensibility in data;

momentarily apprehend depth.

## **The device/platform and its events**

Build software better, together.

Google Analytics

ABOUT

FOUNDED 2008

HOME BASE San Francisco

HUBBERS 239

LOVEABLE MASCOT The Octocat

Looking for help?

Check out our documentation or contact support.

Get in touch

Our contact form goes to real people and we'd love to hear from you. Contact us.

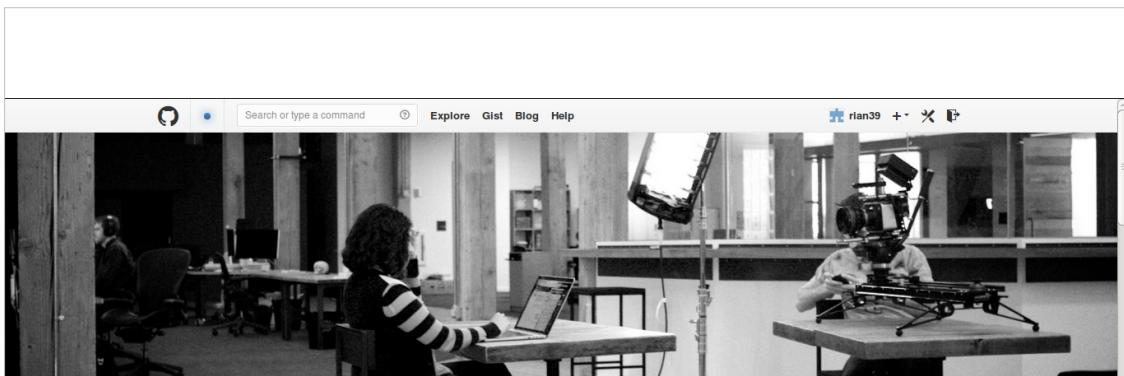
GitHub is the best place to share code with friends, co-workers, classmates, and complete strangers. Over five million people use GitHub to build amazing things together.

With the collaborative features of GitHub.com, our desktop and mobile apps, and GitHub Enterprise, it has never been easier for individuals and teams to write better code, faster.

Originally founded by Tom Preston-Werner, Chris Wanstrath, and PJ Hyett to simplify sharing code, GitHub has grown into the largest code host in the world.

© 2014 GitHub, Inc. Terms Privacy Security Contact

Status API Training Shop Blog About



About GitHub      Our Team      **Press**      Jobs

There are **6.1M** people collaborating right now across **13.2M** repositories on GitHub. Developers from all around the world are building amazing things together. Their story is our story.

[GitHub: Innovations That Mattered in 2013](#)  
GovLoop, December 12, 2013

[GitHub's Tom Preston-Werner: How We Went Mainstream](#)  
FastMrleMan, November 18, 2013

[press@github.com](mailto:press@github.com)

[GitHub Fact Sheet](#)

'Their story is our story'

## 13.2 million repositories: 'their story' as showcase

The screenshot shows the GitHub Explore page with a light gray background. At the top, there's a navigation bar with links for 'Explore', 'Grid', 'Blog', and 'Help'. A user icon 'mash9' is on the right. Below the navigation, a search bar says 'Search or type a command...'. The main area is titled 'Explore' with the subtitle 'Browse interesting projects, solving all types of interesting problems.' It features several cards:

- Web application frameworks**: Projects that power GitHub for Mac.
- Projects that power GitHub**: Projects that power GitHub.
- Redis databases**: Redis databases.
- Package managers**: Across programming languages and platforms, package managers like...
- Security**: Here are some of the great security projects that GitHub is using to...
- Design essentials**: Design essentials.
- Design tools**: Design tools.

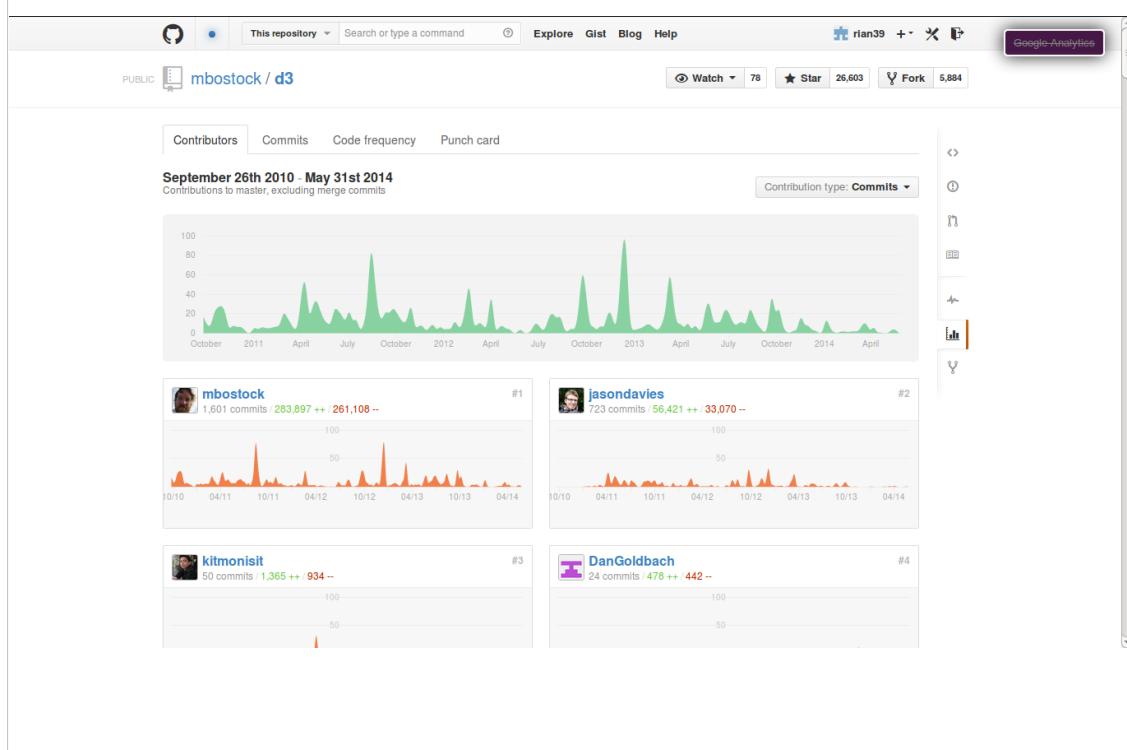
Below these cards is a 'See all' link. Further down, there are sections for 'Trending repositories this week' and 'Staffed by GitHub staff this week'. Each section lists five repositories with their names, descriptions, and statistics (stars, forks, issues, pull requests). For example, in the 'Trending repositories this week' section:

- timgreen/github-chat-sheet**: A lot of cool features for Git and GitHub. 4.5 stars, 227 forks.
- cantinohuginn**: Build agents that monitor and act on your behalf. Your agent... 4.5 stars, 1 fork.
- qphillipngifly**: One poly-style background generator with c3.js. 4.5 stars, 2 forks.
- vtfree-free-programming-books**: 4.5 stars, 2 forks.
- almandevia**: Dynamic, browser-based visualization library. 4.5 stars, 5 forks.

In the 'Staffed by GitHub staff this week' section:

- cantinohuginn**: Build agents that monitor and act on your behalf. Your agent... 4.5 stars.
- timgreen/github-chat-sheet**: A lot of cool features for Git and GitHub. 4.5 stars.
- qphillipngifly**: One poly-style background generator with c3.js. 4.5 stars.
- NYTimes/pouever**: 4.5 stars.
- tppop/heroku-binlogs**: Create binlogs to easily dispatch the heroku command for... 4.5 stars.

## 'Their story': 1 repository at a time



## Publishing events: 210 million (1 March 2011 – today) : [www.githubarchive.org](http://www.githubarchive.org)



Open-source developers all over the world are working on millions of projects: writing code & documentation, fixing & submitting bugs, and so forth. GitHub Archive is a project to **record** the public GitHub timeline, **archive it**, and **make it easily accessible** for further analysis.

Looking for the [daily top new & watched repository](#) reports? [Sign up here](#).



GitHub provides [18 event types](#), which range from new commits and fork events, to opening new tickets, commenting, and adding members to a project. The activity is aggregated in hourly archives, which you can access with any HTTP client:

| Query                                | Command  |
|--------------------------------------|--|
| Activity for April 11, 2012, 3PM UTC | <code>wget http://data.githubarchive.org/2012-04-11-15.json.gz</code>            |
| Activity for April 11, 2012          | <code>wget http://data.githubarchive.org/2012-04-11-(0..23).json.gz</code>       |
| Activity for April 2012              | <code>wget http://data.githubarchive.org/2012-04-(01..31)-(0..23).json.gz</code> |

*Note: timeline data is available starting February 12, 2011.*

Each archive contains a stream of JSON encoded GitHub events ([sample](#)), which you can process in any language. Ruby example:

```
1 require 'open-uri'
2 require 'zlib'
3 require 'yaml'
4
5 oj = open('https://data.githubarchive.org/2012-02-11-12.json.gz')
```

**Google bigquery**

**Table Details: timeline**

**Description**  
Describe this table...

**Table Info**

|                |                               |
|----------------|-------------------------------|
| Table ID       | githubarchive.github.timeline |
| Table Size     | 112 GB                        |
| Number of Rows | 185,116,106                   |
| Creation Time  | 7:49am, 29 Apr 2012           |
| Last Modified  | 9:22am, 31 Mar 2014           |

**Devices re-format published events**

**Query Results** 11:19am, 31 Mar 2014

| Row | created_at          | type              | actor       | repository_name        | url   |
|-----|---------------------|-------------------|-------------|------------------------|---|
| 1   | 2014-03-15 09:04:43 | PushEvent         | vvakame     | DefinitelyTyped        | <a href="https://github.com/borisyankov/DefinitelyTyped/compare/3d3832de3e...93a2313f62">https://github.com/borisyankov/DefinitelyTyped/compare/3d3832de3e...93a2313f62</a>               |
| 2   | 2014-03-15 09:04:42 | PushEvent         | gizmomogwai | plist                  | <a href="https://github.com/gizmomogwai/plist/compare/12eb82d283...83ad8b5f26">https://github.com/gizmomogwai/plist/compare/12eb82d283...83ad8b5f26</a>                                   |
| 3   | 2014-03-15 09:04:48 | IssueCommentEvent | ben-lin     | node.inflection        | <a href="https://github.com/dreamerstab/node.inflection/issues/16#issuecomment-37721104">https://github.com/dreamerstab/node.inflection/issues/16#issuecomment-37721104</a>               |
| 4   | 2014-03-15 09:04:48 | WatchEvent        | cbmd        | mongofill              | <a href="https://github.com/koubas/mongofill">https://github.com/koubas/mongofill</a>   |
| 5   | 2014-03-15 09:04:48 | WatchEvent        | svelt       | molokai                | <a href="https://github.com/tomasr/molokai">https://github.com/tomasr/molokai</a>   |
| 6   | 2014-03-15 09:04:47 | GollumEvent       | swordray    | wiki                   | <a href="https://github.com/ruby-china/wiki/wiki/RubyGems">https://github.com/ruby-china/wiki/wiki/RubyGems</a>   |
| 7   | 2014-03-15 09:04:46 | PushEvent         | elfet       | purephp                | <a href="https://github.com/elfet/purephp/compare/dc24b70f00...c94c524e76">https://github.com/elfet/purephp/compare/dc24b70f00...c94c524e76</a>   |
| 8   | 2014-03-15 09:04:46 | PullRequestEvent  | nikkypx     | rets_data              | <a href="https://github.com/arcticleo/rets_data/pull/2">https://github.com/arcticleo/rets_data/pull/2</a>   |
| 9   | 2014-03-15 09:04:46 | WatchEvent        | mjaneczek   | google-authenticator   | <a href="https://github.com/jaredonline/google-authenticator">https://github.com/jaredonline/google-authenticator</a>   |
| 10  | 2014-03-15 09:04:53 | PushEvent         | micmath     | Rye                    | <a href="https://github.com/micmath/Rye/compare/69c02b1aea...7885ea9e36">https://github.com/micmath/Rye/compare/69c02b1aea...7885ea9e36</a>   |
| 11  | 2014-03-15 09:04:53 | PushEvent         | fitret      | dagon                  | <a href="https://github.com/fitret/dagon/compare/8897fa9404...77e5f48276">https://github.com/fitret/dagon/compare/8897fa9404...77e5f48276</a>   |
| 12  | 2014-03-15 09:04:53 | WatchEvent        | joeyates    | spork                  | <a href="https://github.com/sporkrb/spork">https://github.com/sporkrb/spork</a>   |
| 13  | 2014-03-15 09:04:53 | CreateEvent       | grcnva      | mcs-chatboard          | <a href="https://github.com/grcnva/mcs-chatboard">https://github.com/grcnva/mcs-chatboard</a>   |
| 14  | 2014-03-15 09:04:51 | PushEvent         | albatrossen | bungeebouncer          | <a href="https://github.com/albatrossen/bungeebouncer/compare/57a3771590...7794ffb531">https://github.com/albatrossen/bungeebouncer/compare/57a3771590...7794ffb531</a>                   |
| 15  | 2014-03-15 09:04:51 | PushEvent         | cataliniciu | ChessEngine-Xboard-cpp | <a href="https://github.com/cataliniciu/ChessEngine-Xboard-cpp/compare/9bedce1744...8b98d2ff45">https://github.com/cataliniciu/ChessEngine-Xboard-cpp/compare/9bedce1744...8b98d2ff45</a> |

First < Prev Rows 1-15 of 50 Next > Last

## Google bigquery

COMPOSE QUERY

Query History  
Job History

githubarchive

- ▶ github
- ▶ github\_explore
- ▶ github\_proper
- ◀ click anywhere

New Query

```
1 SELECT created_at, type, actor, repository_name, url FROM [githubarchive:github.timeline]
2 WHERE regexp_match(repository_name, 'cloud')
3 LIMIT 50
```

RUN QUERY Save Query Save View Enable Options Query complete (3.0s elapsed, 23.2 GB processed)

|                   |                                |                        |   |
|-------------------|--------------------------------|------------------------|---|
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516027">https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516027</a>         |
| IssueCommentEvent | buildhive                      | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| IssueCommentEvent | buildhive                      | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516091">https://github.com/jclouds/jclouds-labs-openstack/pull/102#issuecomment-43516091</a>         |
| IssueCommentEvent | annashipman                    | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516194">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516194</a>               |
| PushEvent         | sbrunner                       | tilecloud              | <a href="https://github.com/sbrunner/tilecloud/compare/185a371ce5...7d69504ad1">https://github.com/sbrunner/tilecloud/compare/185a371ce5...7d69504ad1</a>                               |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PullRequestEvent  | zack-shoylev                   | jclouds-labs-openstack | <a href="https://github.com/jclouds/jclouds-labs-openstack/pull/102">https://github.com/jclouds/jclouds-labs-openstack/pull/102</a>   |
| PushEvent         | maksimov                       | dasein-cloud-core      | <a href="https://github.com/maksimov/dasein-cloud-core/compare/90827bb84c...ca08115701">https://github.com/maksimov/dasein-cloud-core/compare/90827bb84c...ca08115701</a>               |
| PushEvent         | cyrichardson                   | docs-cloud-files       | <a href="https://github.com/rackerlabs/docs-cloud-files/compare/75bc24a55e...6f1a9daa3c">https://github.com/rackerlabs/docs-cloud-files/compare/75bc24a55e...6f1a9daa3c</a>             |
| PushEvent         | laszewsk                       | cloudmesh              | <a href="https://github.com/cloudmesh/cloudmesh/compare/1b327284fb...abcdabd325">https://github.com/cloudmesh/cloudmesh/compare/1b327284fb...abcdabd325</a>                             |
| IssueCommentEvent | cloudbees-pull-request-builder | jclouds-labs           | <a href="https://github.com/jclouds-labs/pull/62#issuecomment-43516613">https://github.com/jclouds-labs/pull/62#issuecomment-43516613</a>   |
| IssueCommentEvent | mattbostock                    | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516643">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516643</a>               |
| CreateEvent       | jabley                         | vcloud-walker          | <a href="https://github.com/alphagov/vcloud-walker/compare/migrate-fog-to-core">https://github.com/alphagov/vcloud-walker/compare/migrate-fog-to-core</a>                               |
| IssueCommentEvent | mikepea                        | vcloud-edge_gateway    | <a href="https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516718">https://github.com/alphagov/vcloud-edge_gateway/pull/67#issuecomment-43516718</a>               |
| PushEvent         | dankilman                      | cloudify-manager       | <a href="https://github.com/cloudify-cosmo/cloudify-manager/compare/813efdfa5e...c909eb45ee">https://github.com/cloudify-cosmo/cloudify-manager/compare/813efdfa5e...c909eb45ee</a>     |
| PullRequestEvent  | maksimov                       | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/pull/192">https://github.com/greese/dasein-cloud-core/pull/192</a>   |
| IssuesEvent       | duglin                         | cloud_controller_ng    | <a href="https://github.com/cloudfoundry/cloud_controller_ng/issues/246">https://github.com/cloudfoundry/cloud_controller_ng/issues/246</a>   |
| IssueCommentEvent | cf-glibot                      | cloud_controller_ng    | <a href="https://github.com/cloudfoundry/cloud_controller_ng/issues/246#issuecomment-43516802">https://github.com/cloudfoundry/cloud_controller_ng/issues/246#issuecomment-43516802</a> |
| PushEvent         | andylyall                      | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/compare/65a45c29c0...208773ac2e">https://github.com/greese/dasein-cloud-core/compare/65a45c29c0...208773ac2e</a>                   |
| PullRequestEvent  | andylyall                      | dasein-cloud-core      | <a href="https://github.com/greese/dasein-cloud-core/pull/192">https://github.com/greese/dasein-cloud-core/pull/192</a>   |
| PullRequestEvent  | jabley                         | vcloud-walker          | <a href="https://github.com/alphagov/vcloud-walker/pull/76">https://github.com/alphagov/vcloud-walker/pull/76</a>   |
| PushEvent         | BrunoMatry                     | cloudbook              | <a href="https://github.com/BrunoMatry/cloudbook/compare/58d7be347a...03d3225351">https://github.com/BrunoMatry/cloudbook/compare/58d7be347a...03d3225351</a>                           |

## **The datastream and its platform stories**

## 'Countless stories to tell'

Search or type a command Explore Gist Blog Help rian39 + - X F Google Analytics

The GitHub Data Challenge

May 1, 2012 priandoll Watercooler

The GitHub public timeline is now easy to query and analyze. With hundreds of thousands of events in the timeline every day, there are countless stories to tell.



<https://camo.githubusercontent.com/5e6eb0b00d714eb5b8ec84254205c61c2a97c68d/687474703a2f2f63746f6465782e6769746875622e636f6d2f696d616765732f77616c646f6361742e6a7067>

Featured

- All Posts
- New Features
- Engineering
- Enterprise
- Meetups
- New Hires
- Watercooler

Subscribe

## 'What stories will be told this year? Its up to you'

The screenshot shows a GitHub blog post titled "The GitHub Data Challenge II". The post is dated April 3, 2013, by briandoll, and is categorized under "Watercooler". The content discusses the millions of projects on GitHub and the activity around them, including issues, code pushes, and pull requests. It mentions a previous data challenge and encourages users to submit their own graphs and descriptions to data@github.com by May 8th, 2013, PST. A sidebar on the right lists categories like Featured, All Posts, New Features, Engineering, Enterprise, Meetups, New Hires, and Watercooler (which is selected). A search bar and a "Subscribe" link are also visible.

April 3, 2013   briandoll   Watercooler

There are millions of projects on GitHub. Every day, people from around the world are working to make these projects better. Opening issues, pushing code, submitting Pull Requests, discussing project details — GitHub activity is a papertrail of progress. Have you ever wondered what all that data looks like? There are millions of stories to tell; you just have to look.

Last year we held our [first data challenge](#). We saw [incredible visualizations](#), [interesting timelines](#) and [compelling analysis](#).

What stories will be told this year? It's up to you!

To Enter

Send a link to a GitHub repository or gist with your graph(s) along with a description to [data@github.com](mailto:data@github.com) before **midnight, May 8th, 2013 PST**.

Subscribe

Data access

The GitHub public timeline is a featured public dataset available on [Google BigQuery](#). The "timeline" table has over a year's worth of public activity and is approaching 100M rows. You can find even more data available in JSON format on [The GitHub Archive project](#).

You are free to use any tools you like. If you choose to use BigQuery, running queries against the GitHub dataset is free for the first 100GB of query processing. Pricing information for additional query processing is [available here](#). After [signing up for BigQuery](#), add the project name "githubarchive".

## Stories: individuals and their similarities

The screenshot shows a web page titled "Stories: individuals and their similarities". At the top, there's a green header with a message to recruiters: "Dear recruiters: While you read this, make sure that you remember that GitHub is not your CV, and that these stats only provide a biased and one-sided view. This is just a toy. Don't take it too seriously!" Below this is a button labeled "OK, I promise!". In the top right corner, there's a small purple box with the text "Brought to you by Google Analytics".

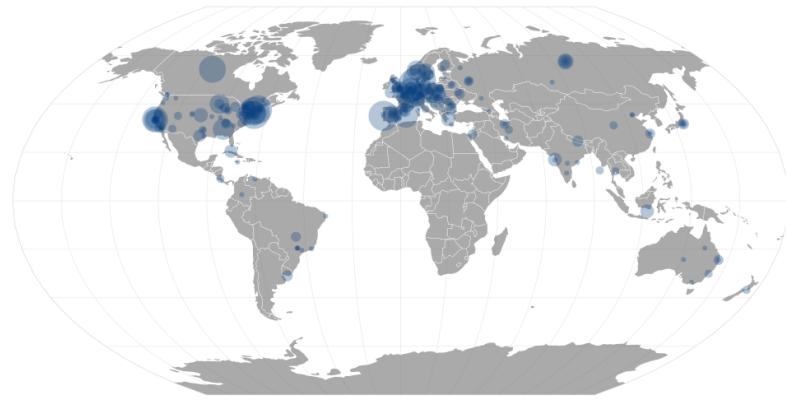
The main content area has a white background. It features the text "THE OPEN SOURCE REPORT CARD" at the top left. In the center, there's a form field with the placeholder "Enter a GitHub username". Below this, a section titled "HOW IT WORKS" contains text about open source contributions and data sources like the GitHub API and GitHub Archive.

On the right side of the page, there's a vertical sidebar with some icons and text, though the text is mostly illegible due to the screenshot quality.

## Repositories and their places

Filter by Language Repository

All JavaScript Ruby Python PHP Shell Java Objective-C C++ C



Use the scroll wheel to zoom in and out.

GitHub's [public timeline](#) summarizes the open source work of millions of developers the world over.

This data set contains contributions to the top 200 GitHub repositories during the first four months of

GEEKSTA

Sections · Projects · Geeklog · Tools · Site · [Disqus](#) · [Google Analytics](#) · [Gravatar](#) · [Geeklog Research Board](#)

## Exploring Expressions of Emotions in GitHub Commit Messages



Source: [Wikipedia](#)

Last week [GitHub announced](#) that the data collected over at the [GitHub Archive](#) was made available as a public dataset on [Google's BigQuery web service](#). Pretty exciting news for data analysts, considering that the timeline dataset currently contains more than 7 million records and is growing quickly.

To see what developers come up with analyzing the GitHub timeline and language correlation datasets, they started a [data challenge](#) which is open for submissions until May 21st.

Being interested in both natural and programming languages, I decided to expand on this [analysis of profanity in git commit messages per programming language](#) with a more general exploration of expressions of emotions as well as looking at messages indicating issues and those containing swear words.

### Method of Collecting and Analyzing Data

Reading millions of commit messages and categorizing them is certainly not a task I'd be able to accomplish in less than 3 weeks and honestly wouldn't want to. Instead, I assembled lists of words expressing emotions using the resources listed at the [bottom of this article](#) and wrote corresponding database queries to count occurrences of these words grouped by programming languages.

Then for each language I calculated the percentage of messages that meet the criteria, i. e. contain certain expressions, in relation to the total number of commits for that language. Only non-empty commit messages of projects with an assigned language were included.

Since the number of commits varies considerably across different languages only those with at least 40,000

**Geek Shirts**

Get the [Bash fork bomb shirt](#) for only \$14 for a Hanes tagless tee. Limited until May 20th, 2014.

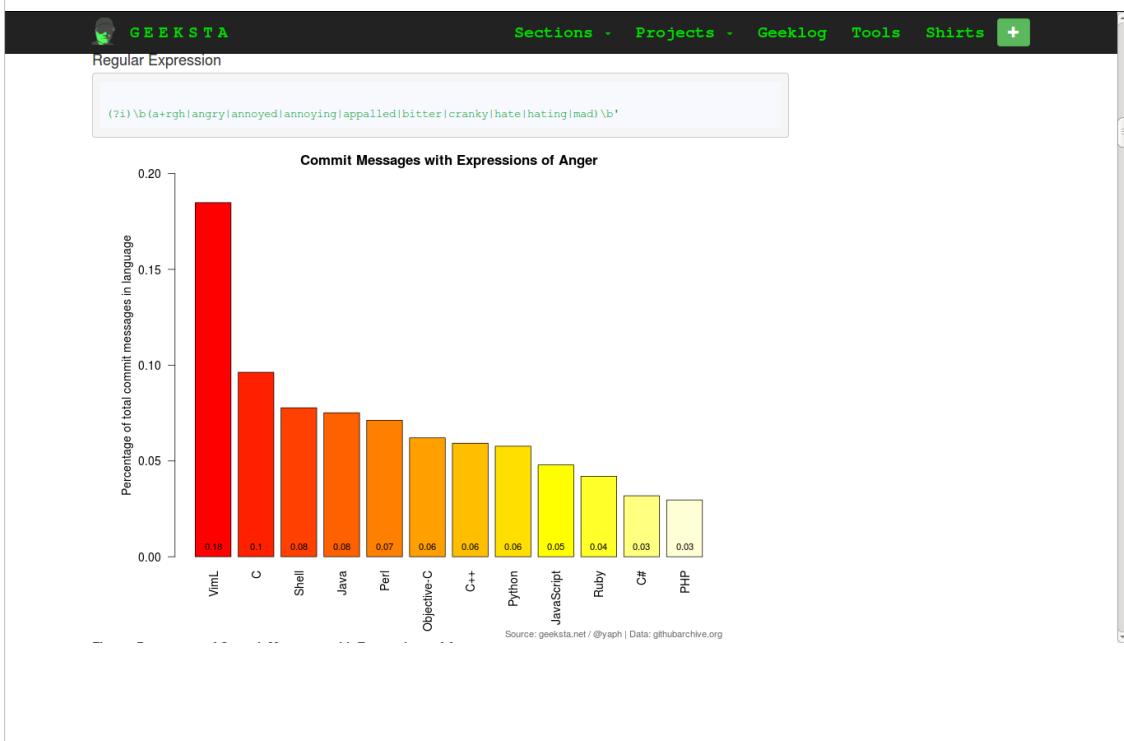
*Limited Offer*



Create your personalized [born at unix time t-shirt](#).



## Affects into repositories



## Live reports of event counts

Octoboard

Global activity Repository Public Fork Pull Request Issue GH Google Analytics

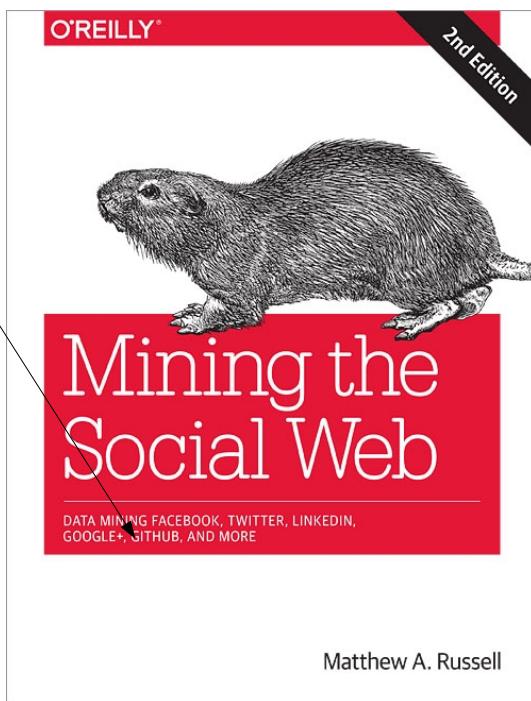
### GitHub activity dashboard.

Octoboard is based on GitHub Archive : each day, it scans new GitHub events archives and computes a few stats, with a 15 days history. You can see some general data on this page, or use menu for more information about language and history.

Since march 11th, 2012 :

|   |  |  |  |
|---|--|--|--|
| <b>Global Activity</b><br>158,547,611<br>+ 198,185 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 88% above average. | <b>Created Repositories</b><br>7,758,991<br>+ 9,699 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 81% above average. | <b>Open Sourced Repositories</b><br><b>149,497</b><br>+ 187 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 66% above average. | <b>Forked Repositories</b><br>5,931,778<br>+ 7,415 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 103% above average. |
| <b>Opened Pull Requests</b><br>501,755<br>- 2,023 / day<br>Yesterday<br>It was a <b>bad</b> day, they were down 26% below average.      | <b>Opened Issues</b><br>5,474,375<br>+ 6,843 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 89% above average.        | <b>New GitHub Pages</b><br>311,154<br>- 389 / day<br>Yesterday<br>It was a <b>very good</b> day, they were up 89% above average.                 | <b>Created Gists</b><br>2,374,309<br>+ 8,005 / day<br>Yesterday<br>It was a <b>good</b> day, they were up 24% above average.             |

**Github data as  
exemplar for  
coding work**



Matthew A. Russell

**Unformatted imitation in the data stream**

## *names in datastreams*

'Consider for a moment the more general conditions of naming. First, a name is offered, given, imposed by someone or by some set of someones, and it is attributed to someone else. It requires an intersubjective context, but also a mode of address, for the name emerges as the addressing of a coinage to another, and in that address, a rendering of that coinage proper' (Butler, 1997: 29)

## Names in STS

'it is important to explore the variety in current naming traditions in terms of what they do, the impact of the speech acts involved, integrating analyses of such technologies into studies of the constitution of subjects, the politics of belonging, exclusion, and control' (Palsson, 2014: 628).

```
repo
maxkirchoff/google-music-dupe-killer
MaratKarimov/TwitterClient
seejay/xgoogle
canercandan/mini-google-in-perl
chengdujin/Mining-the-Social-Web
twinslash/omniauth-vkontakte
Deminem/google-glass-mirror-ruby-sinatra-scaffold
tfountain/oauth-2---facebook---zend-framework-...
alesaudate/camel-twitter-websockets
soequelle/facebook-iphone-sdk
shadowmaru/monitor-twitter
Hiroki-Takayama/python-flask-social-oauth-face...
davcro/facebook_client
pratyush-nigam/Learning-RoR
jabranr/instagram-api-php
```

dpackcartonmachine/3layer-5layer-7layer-corrug...  
dpackcartonmachine/corrugated-box-machine--817258  
dpackcartonmachine/corrugated-box-machine--854402  
dpackcartonmachine/corrugated-box-machine--709152  
dpackcartonmachine/5-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/3-layers-corrugated-box-mac...  
dpackcartonmachine/3-layers-corrugated-box-mac...  
dpackcartonmachine/5-layers-corrugated-box-mac...  
dpackcartonmachine/corrugated-box-machine--529003  
dpackcartonmachine/corrugated-box-machine--058569  
dpackcartonmachine/corrugated-box-machine--564559  
dpackcartonmachine/semi-automatic-four-color-p...  
dpackcartonmachine/3layer-5layer-7layer-Corrug...  
dpackcartonmachine/3layer-corrugated-box-machi...  
dpackcartonmachine/3layer5layer7layer-corrugat...  
dpackcartonmachine/corrugated-box-machine--408286  
dpackcartonmachine/3layer-5layer-7layer-Corrug...  
dpackcartonmachine/corrugated-box-machine--952283  
dpackcartonmachine/7-layers-corrugated-box-mac...  
dpackcartonmachine/7-layers-corrugated-box-mac... 4

# Copied names as affective imitation: forks, mirrors

FORKCOUNT: **4,468128** (approximately 4.5 million repos are forks)

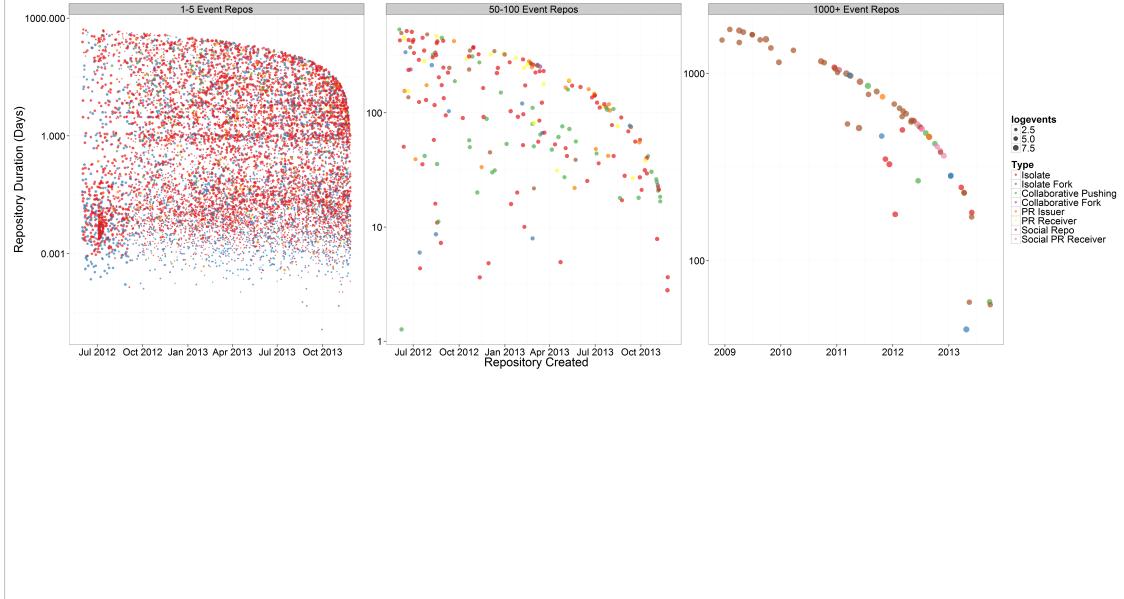
```
SELECT count(*) as fork_count FROM [metacommunities:github_proper.repo_list]  
where fork=1
```

UNIQUE NAMES: **5,323,230** (approximately 5.3 million repos have unique names, so 7 million names are imitations)

```
SELECT count(distinct(repository_name)) as name_count FROM  
[githubarchive:github.timeline] LIMIT 1000;
```

'incorporation into collectivities involves **affective imitation**' (Gatens & Lloyd, 1999)

## 13.2 million repositories, but mostly **imitative incorporations**



## 'passionate imitation'

### Names as bundles

```
repository_url
https://github.com/mantrala/twitter_clone
https://github.com/estebiste/django-twitter-bootstrap
https://github.com/notonalot/Lol-twitter-stats
https://github.com/sherberg/twitter-stream-learning
https://github.com/RouxRC/TwitterApiDocumentationRSS
https://github.com/stephenturner/twitterarchive
https://github.com/mausrokii/twitter-activator
https://github.com/sferik/sign-in-with-twitter
https://github.com/konstandino-a/twitter-analyser
https://github.com/thecodepath/ios_twitter
https://github.com/jacutis/twitter_clone
https://github.com/nst/STTwitter
https://github.com/Nodejs-Colorado/node-twitter-sentiment-promises
https://github.com/mjhead/node-twitter-sentiment-generators
https://github.com/spiritdead/Twitterizer
https://github.com/mjhead/node-twitter-sentiment
https://github.com/tonytonyan/ruby-conf-tw-2014-twitter-wall
https://github.com/mjhead/node-twitter-sentiment-promises
https://github.com/yusuke/twitter4j
https://github.com/margaritoCajero/twitter
https://github.com/kathrynsiegel/twitterbot
https://github.com/cwhitemore1/TwitterAPI
https://github.com/marek5050/TwitterScrap
https://github.com/AvianFlu/ntwitter
https://github.com/jamesfebin/meteor-twitter-api
https://github.com/mrbellek/twitterbot
https://github.com/tchu88/diff-twitter
```

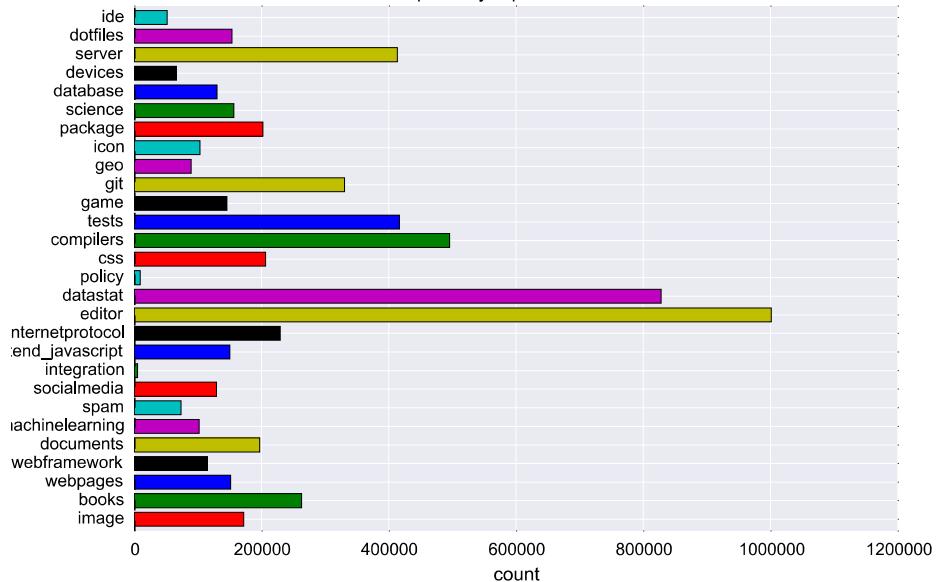
### Most common repository names

|                   |       |
|-------------------|-------|
| test              | 85404 |
| try_git           | 67393 |
| dotfiles          | 39928 |
| first_app         | 34772 |
| sample_app        | 33848 |
| demo_app          | 30076 |
| bootstrap         | 25814 |
| spoon-knife       | 24603 |
| hello-world       | 19989 |
| bitstarter        | 13076 |
| blog              | 11392 |
| helleworld        | 10362 |
| android           | 9722  |
| homebrew          | 8305  |
| rails             | 8281  |
| jquery            | 7540  |
| angular.js        | 6778  |
| myrepo            | 6555  |
| node              | 6499  |
| testrepo          | 6320  |
| html5-boilerplate | 6188  |
| scripts           | 6109  |
| test1             | 6054  |
| demo              | 5890  |
| projects          | 5718  |
| git               | 5706  |
| github            | 5681  |
| phonegap-star     | 5667  |
| testing           | 5494  |
| sandbox           | 5474  |

the importance of suggestion and imitation in social life, and the limits of forms of analysis which confine themselves to the study of representation or inscription. (Barry and Thrift, 2007: 511)

### Namespace as device-specific imitation

Repository topics on Github



## Querying device-specific affective imitation

IP[y]: Notebook 1a\_Generate\_repo\_topic\_data Last Checkpoint: May 30 10:28 (autosaved)

In [91]: `vz|students|circle|network|stumbleupon|tagged|talkbiznow|taltopia|tarining|teachstreet|termwiki|the_sphere|travbuddy.com|travellers point|tribe.net|trombi.com|tuenti|tumblr|twitter|tylted|vkontakte|vampirefreaks.com|viadeo|virb|vox|wattpad|wayn|weeworld|we heart it|wellwer|weourfamily|wepolls.com|wer|kennt|wen|weread|wiser.org|wooxie|writeaprisoner.com|xanga|xing|xt3|yammer|yelp, inc.|zoo.gr|zooppa`

In [91]: `socialmedia_df = construct_repo_df(socialmediaquery, description = False)`  
starting bigquery ...  
wikipedia[43,7things|academia.edu|about.me|advogato|anobii|asianavenue|asmallworld|athlinks|audimated.com|bebo|biip.no|blackplanet|blauk|blogster|bolt.com|busuu|buznet|cafemon|care2|caringbridge|classmates.com|cloob|couchsurfing|cozcot|cross.tv|crunchyroll|cyworld|dailybooth|dailystrength|delicious|deviantart|diaspora\*|didlr|disaboom|dol2day|dontstayindraigiem.lv|douban|dxy.cn|elftown|elgio|english.?baby|epernicus|eons.com|etoro|experience.?project|expoloroo|facebook|faceparty|faces.com|felife|filmaffinity|filmow|fledgeling|flixster|flickr|focus.com|formspring|fotki|fotoilog|foursquare|friendica|friends.?reunited|friendster|frühstückstrefffue|mbylog|fullcircle|gaiaplanoine|gamerdnal|gapear.com|gather.com|gays.com|getglue|gopoyoko|goodreads|goodwizz|google.?gvpl|oop|gron|net|habbo|habbo|hostel|hostel|hotlist|hotlist|hr.com|hub.?culture|hyve|ibib|identi.ca|indaba|7music|instagram|irc|galleria|jiltalki.com|itsmy|iwiw|jaiku|jlepang|kaixin001|kimibox|lafango|laibhaarilast fm|librarything|life knot|linkedin|linkexpats|listograph|ylivejournal|livemocha|makeoutclub|meetin|meetup|?website|meettheboss|millatfacebook|mixi|mocospace|mognmouthshut.com|mubi|myheritage|myle|myle|mymy|mymyspace|nasza.?klasa.pl|netlog|nexopia|ngo.?post|ning|odnoklassniki|open|dilary|orkut|outeverywhere|patientsli|keme|partyflock|pingsta|pinterest|plaxo|playfire|playlist|com|plurk|poolwe|qapacity|quechup|qzone|rapti|ravelry|renren|reverbnation|.com|ryze|sciencestage|sqrouples|sharethemusic|shelfari|sina|weibo|skoob|skyrock|socialvibe|sonico.com|soundcloud|spaces|stage\_732|stickam|studivz|students.?circle|?network|stumbleupon|tagged|talkbiznow|taltopia|tarining|teachstreet|termwiki|the\_sphere|travbuddy.com|travellers point|tribe.net|trombi.com|tuenti|tumblr|twitter|tylted|vkontakte|vampirefreaks.com|viadeo|virb|vox|wattpad|wayn|weeworld|we heart it|wellwer|weourfamily|wepolls.com|wer|kennt|wen|weread|wiser.org|wooxie|writeaprisoner.com|xanga|xing|xt3|yammer|yelp, inc.|zoo.gr|zooppa  
Waiting on bqjob\_r1bca3395151b160a\_00000146203dd18f\_21 ... (0s) Current status: DONE  
(156058, 3)

In [148]: `red.sadd('social_media', *socialmedia_df.full_name)`  
Out[148]: 156058

Almost 157k repositories here. Is that mostly noise?

In [93]: `socialmedia_df.head(30)`  
Out[93]:

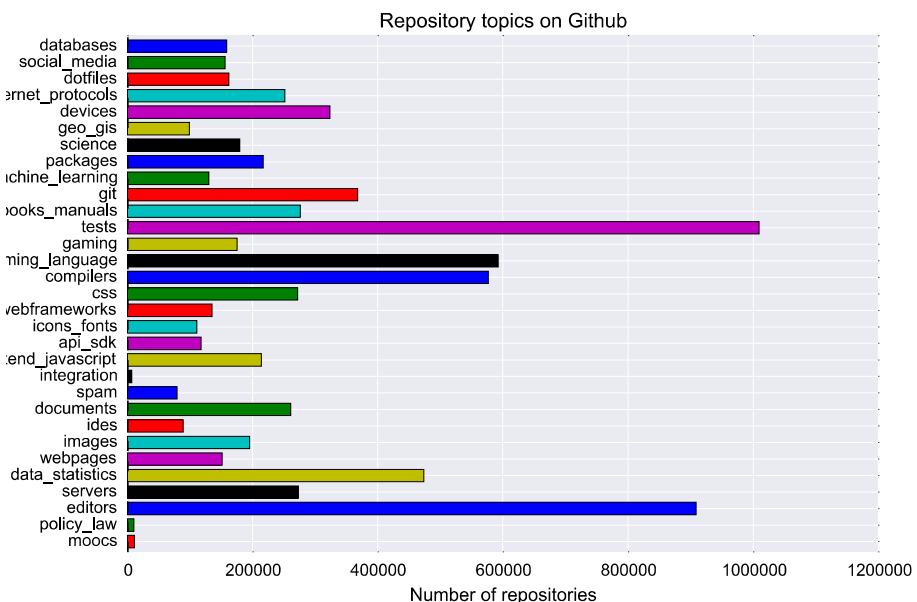
|    | name            | full_name       | fork            |
|----|-----------------|-----------------|-----------------|
| 1  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 2  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 3  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 4  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 5  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 6  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 7  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 8  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 9  | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 10 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 11 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 12 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 13 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 14 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 15 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 16 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 17 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 18 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 19 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 20 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 21 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 22 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 23 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 24 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 25 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 26 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 27 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 28 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 29 | ... ... ... ... | ... ... ... ... | ... ... ... ... |
| 30 | ... ... ... ... | ... ... ... ... | ... ... ... ... |

```
In [93]: socialmedia_df.head(30)
```

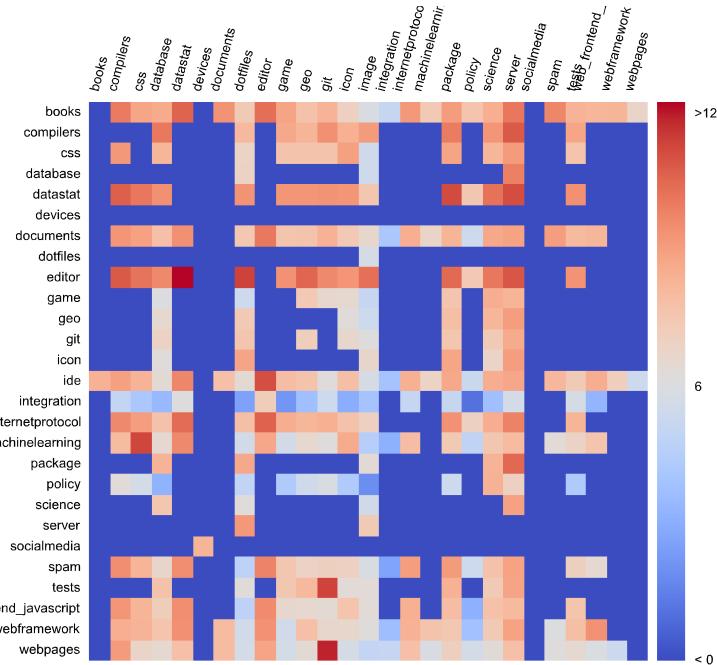
```
Out[93]:
```

|    | name                       | full_name                             | fork |
|----|----------------------------|---------------------------------------|------|
| 0  | PlurkCSS                   | howar31/PlurkCSS                      | 1    |
| 1  | ZXing.Net.Mobile           | JerryLee7809/ZXing.Net.Mobile         | 1    |
| 2  | twitteroauth               | mathijsvdhurk/twitteroauth            | 1    |
| 3  | Recipes-for-Mining-Twitter | winnerdy/Recipes-for-Mining-Twitter   | 1    |
| 4  | DMActivityInstagram        | PrincessThan/DMActivityInstagram      | 1    |
| 5  | wikipedia_api              | gpherguson/wikipedia_api              | 0    |
| 6  | facebook                   | harjeetsingh111/facebook              | 0    |
| 7  | symphony-learning-download | teradotact/symphony-learning-download | 0    |
| 8  | facebook-android-sdk       | slott/facebook-android-sdk            | 1    |
| 9  | facebook-android-sdk       | derfshaya/facebook-android-sdk        | 1    |
| 10 | SwanFlickrUploader         | entercritical/SwanFlickrUploader      | 0    |
| 11 | iOSTraining                | tokyoaret/iOSTraining                 | 1    |
| 12 | angular-google-maps        | adato/angular-google-maps             | 1    |
| 13 | facebook-php-sdk           | cabaltho/facebook-php-sdk             | 1    |

### Unstable trajectories in the namespace



### Variations and inventions through imitation



## Recursively device-specific social research and its affective imitations

The screenshot shows a search interface with a header bar featuring 'Explore', 'Gist', 'Blog', and 'Help' buttons, and a user profile 'rian39'. Below the header is a search bar with the query 'redis'. A message indicates 'We've found 7,292 repository results' and a 'Sort: Best match' dropdown. On the left, a sidebar displays statistics: 'Repositories 7,292', 'Code 1,511,276', 'Issues 28,757', and 'Users 56'. The main area highlights the top result, 'antirez/redis', which is described as an in-memory database supporting various data structures like strings, lists, sets, sorted sets, and hashes. It has 9,354 stars and 2,664 forks. The last update was on May 26.

## **Conclusion: datastream stories**

'their story is our story': repeating the platform's formatting

'countless stories to tell': mirroring the platform

'what stories will be told this year?': datastream as platform

# Conclusions

**Latour:** 'Let the agents produce a dynamics and collect the traces that their actions leave as they unfold so as to produce a rich data set' (Latour et al., 2012):

**Verran:** work on numbers (especially one-many/part-whole tensions)

**Marres/Weltevreden:** work on data formats to heighten ambivalent tensions between device-specific and social process

**Gatens/Barry/Thrift/Tarde:** Affective/passionate imitation and incorporation in collectives generates invention

**Riles:** momentary apprehensions of depth

# **Acknowledgements**

Research funded by ESRC (UK Economic and Social Research Council) under 'Google Data Analytics' programme, 2013-2014

Richard Mills and Stuart Sharples – research associates

Matthew Fuller and Andrew Goffey – co-researchers