

"Big data" skills for online data

@suzymoat

@thedatascilab

Suzy Moat

Data Science Lab
Warwick Business School

THE ALAN
TURING
INSTITUTE

A map of the world built only from
GPS locations of Flickr photos

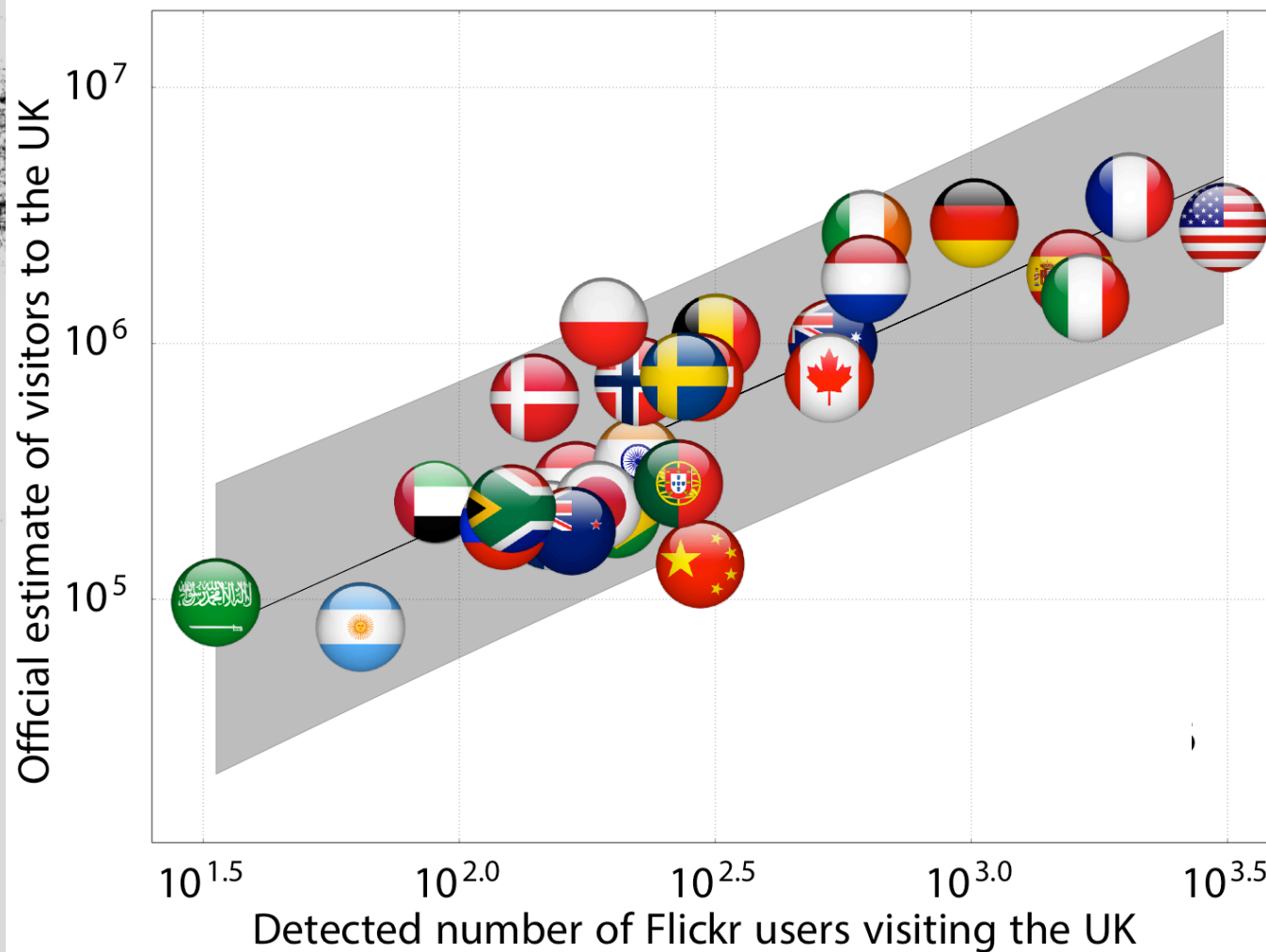


The big data explosion



International Passenger Survey

Flickr: global travel tracking



Alanyali, Preis & Moat (2016)



<http://dx.doi.org/10.1371/journal.pone.0150466>

Alanyali, Preis & Moat (2016)



<http://dx.doi.org/10.1371/journal.pone.0150466>

Seresinhe, Preis & Moat (2016)



<https://dx.doi.org/10.1098/rsos.160146>

Alanyali, Preis & Moat (2016)



<http://dx.doi.org/10.1371/journal.pone.0150466>

Seresinhe, Preis & Moat (2016)



<https://dx.doi.org/10.1098/rsos.160146>



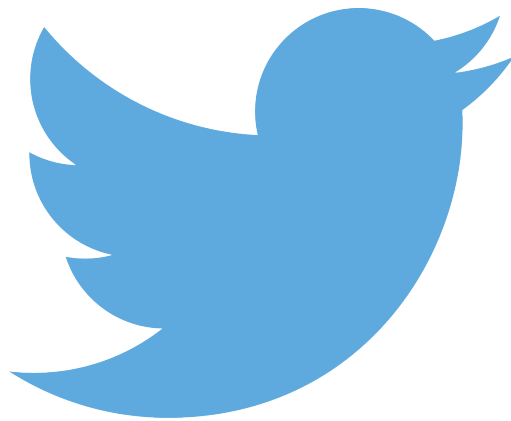
Preis, Moat, Bishop, Treleaven & Stanley (2013)

<http://www.nature.com/articles/srep03141>

flickr

Google

WIKIPEDIA



flickr

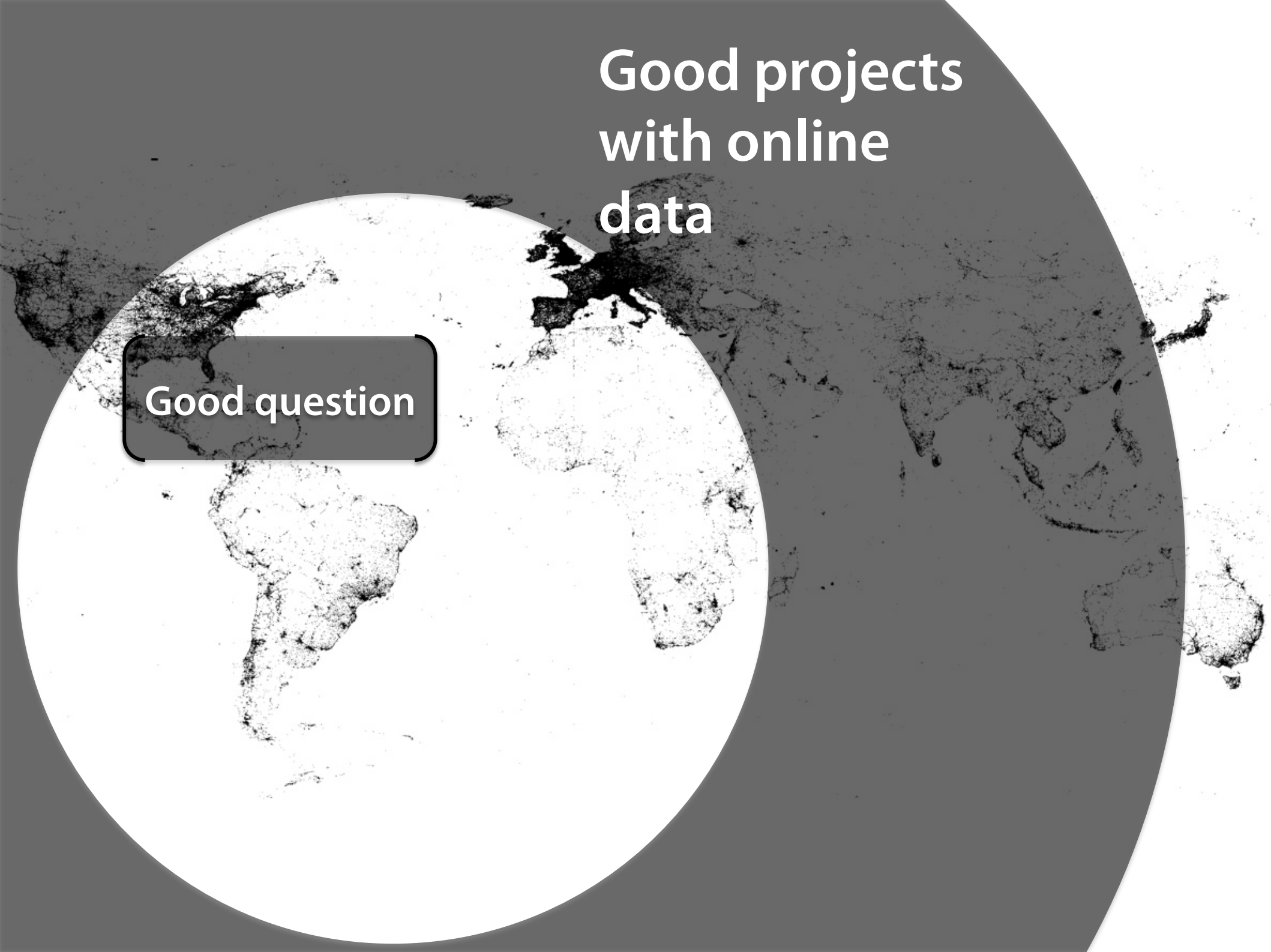
Other case studies:
[http://www.wbs.ac.uk/
about/person/suzy-moat/](http://www.wbs.ac.uk/about/person/suzy-moat/)

Good projects with online data



Good projects with online data

Good question



Good projects with online data

The background of the slide is a dark gray world map. A large circular cutout is positioned on the left side, revealing a white background within it. This white area contains a faint, dark map of the Americas. Two dark gray rounded rectangular boxes are overlaid on this white area. The first box, containing the text 'Good question', is located in the upper left of the white circle. The second box, containing the text 'Statistical methodology', is located in the lower right of the white circle.

Good question

Statistical
methodology

Good projects with online data



Good question

Statistical
methodology

Access to data

Good projects with online data



Good question

Statistical
methodology

Access to data

What's the problem?

A screenshot of a web browser window displaying the 'Wikipedia article traffic statistics' website. The browser's address bar shows 'http://stats.grok.se/'. The page title is 'Wikipedia article traffic statistics'. Below the title, a paragraph asks questions about Wikipedia's readership, such as 'What do Wikipedia's readers care about? Is Britney Spears more popular than Brittany?'. It then provides a search interface with a dropdown for 'English', a text input with '201401', and 'Go' and 'Top' buttons. A 'News:' section lists three updates from February and March 2008. At the bottom, a link 'About these stats' is followed by a disclaimer: 'The raw data is available here. This is very much a beta service and may disappear or change at any time.'

Wikipedia article traffic statistics

What do Wikipedia's readers care about? Is [Britney Spears](#) more popular than [Brittany](#)? Is [Asia Carrera](#) more popular than [Asia](#)? How many people looked at the article on [Santa Claus](#) in December? How many looked at the article on [Ron Paul](#)?

What can *you* find?

Enter a wikipedia article title and press Go

English 201401 Go Top

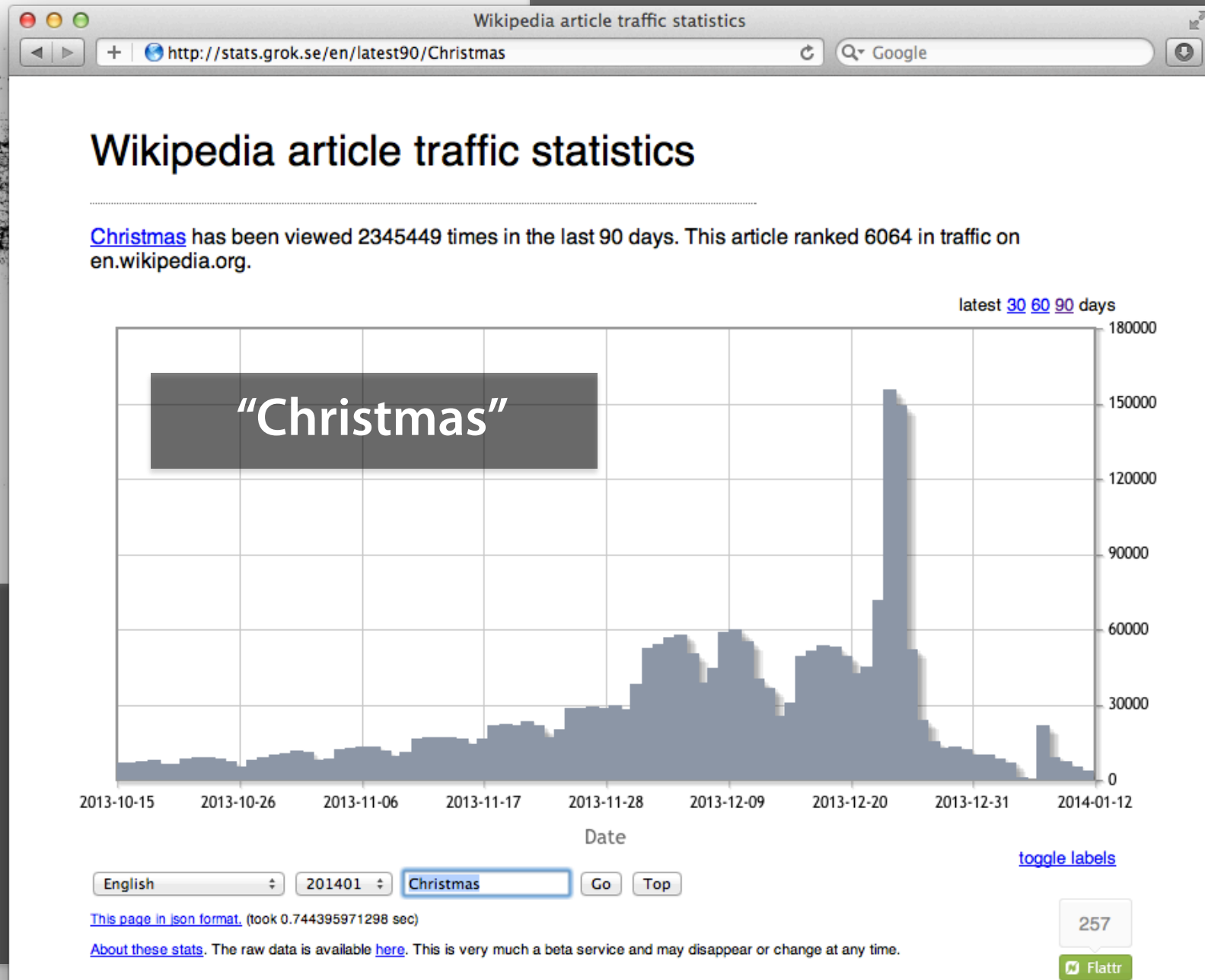
News:

- 2008-02-29: Non-English statistics added (available from February)
- 2008-03-05: Added [top list](#) for English Wikipedia. Other projects will be added soon.
- 2008-03-19: Top 1000 articles in February for all projects.

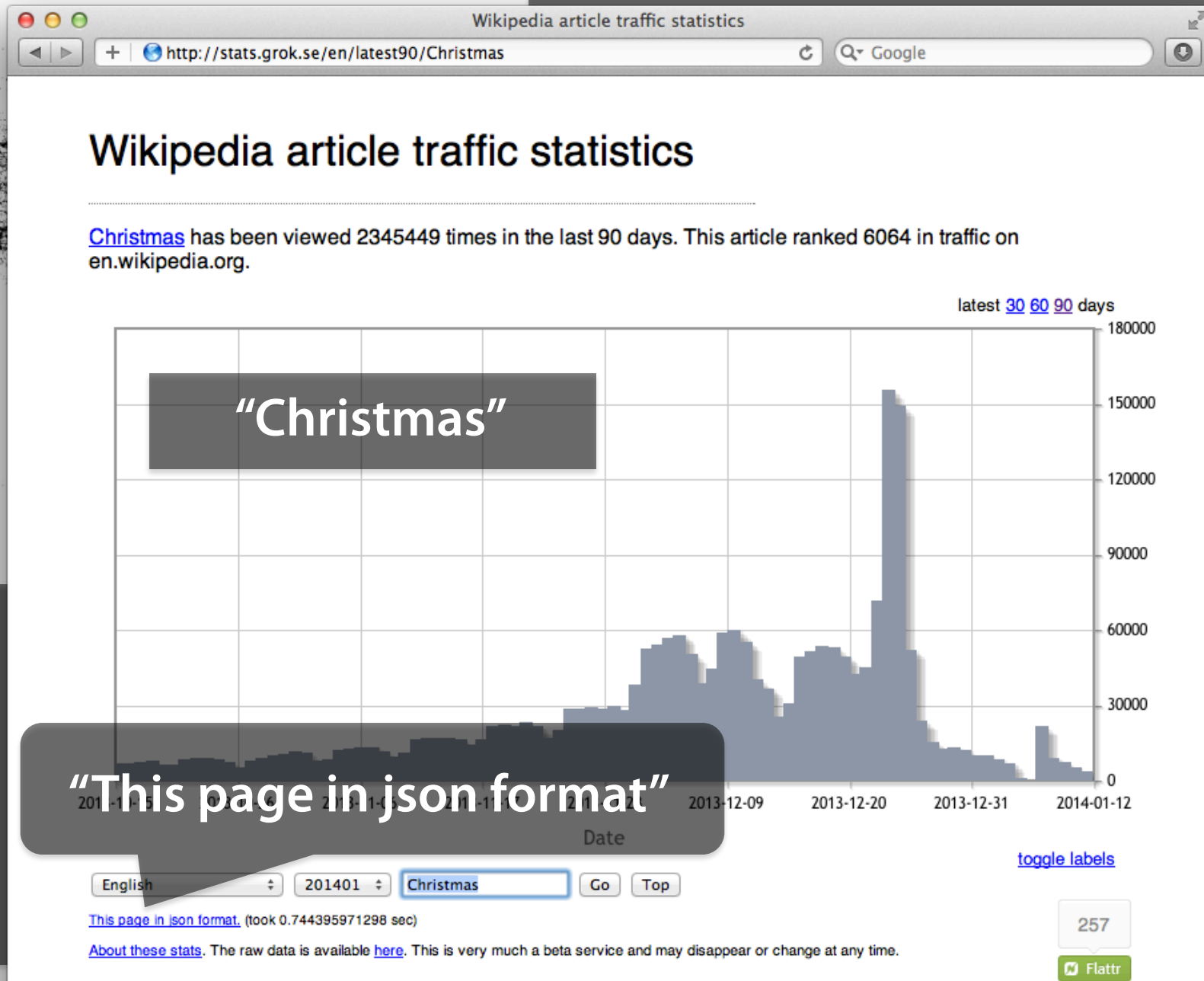
[About these stats](#). The raw data is available [here](#). This is very much a beta service and may disappear or change at any time.

<http://stats.grok.se/>

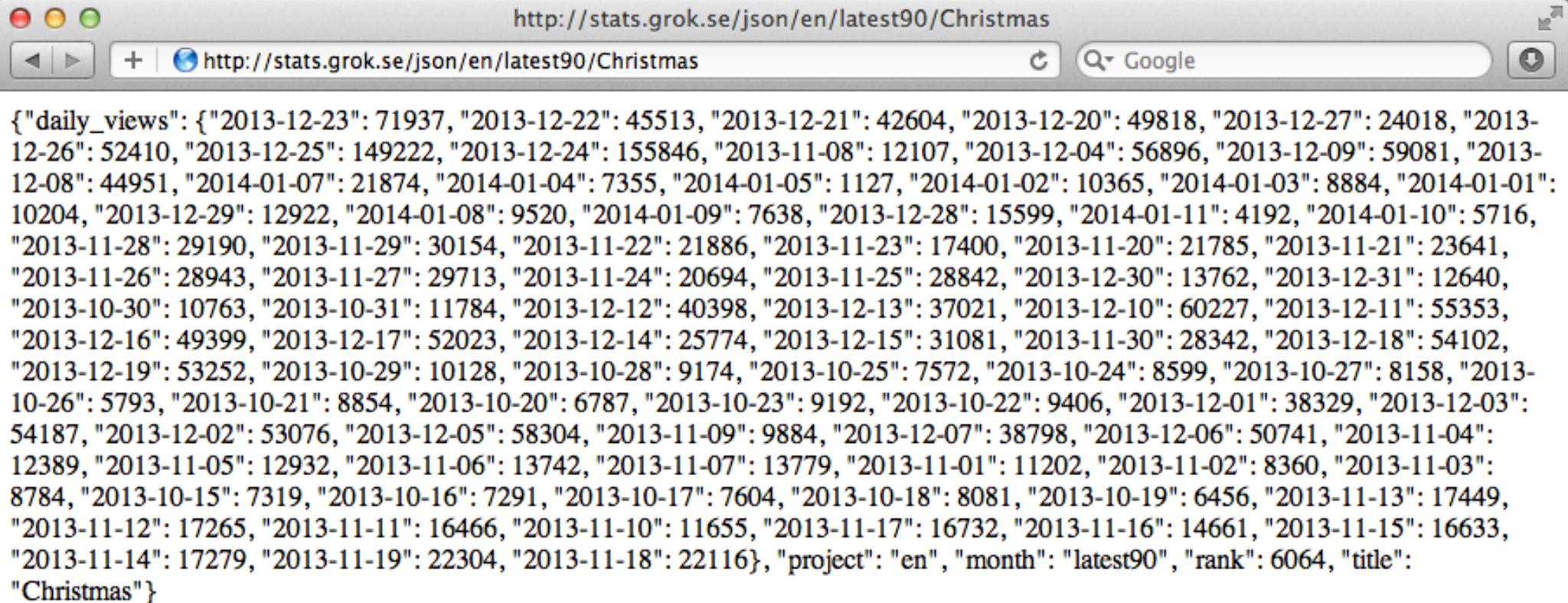
What's the problem?



What's the problem?



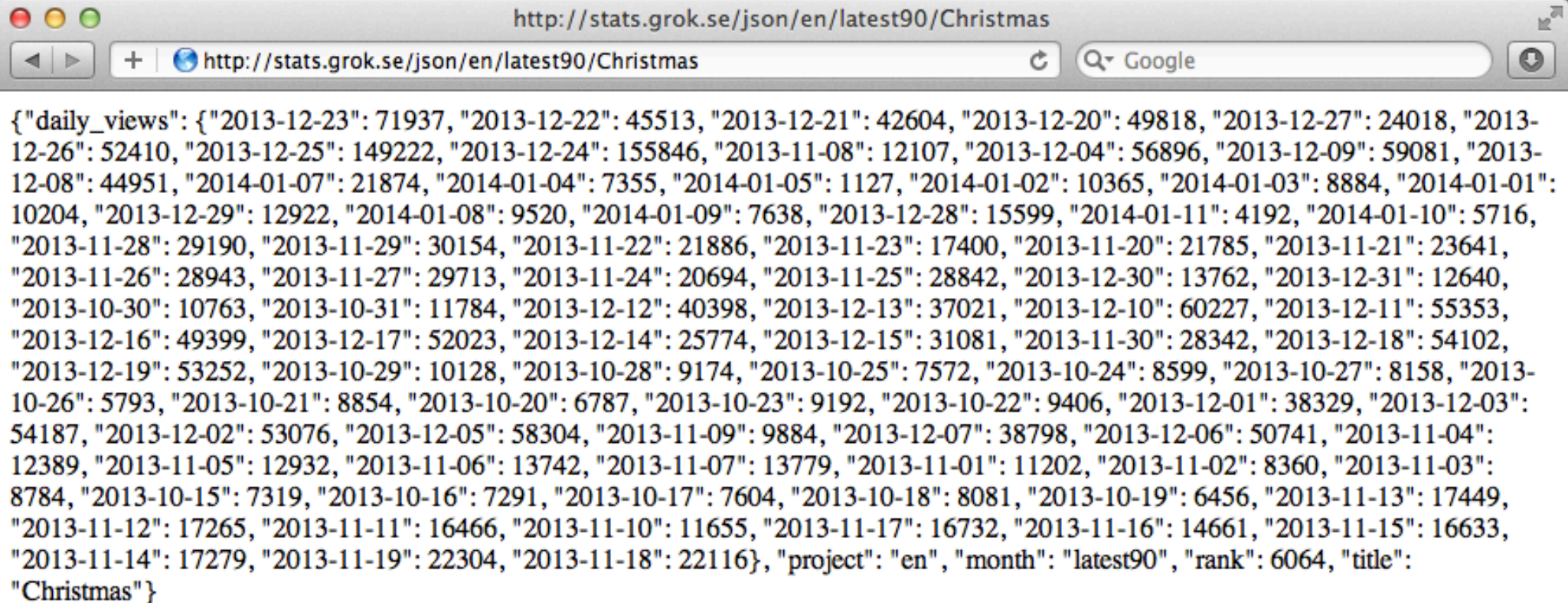
What's the problem?



A screenshot of a web browser window. The address bar shows the URL `http://stats.grok.se/json/en/latest90/Christmas`. The page content displays a large JSON object representing daily views for the month of December 2013. The JSON is formatted with single quotes for keys and values, which is not standard JSON syntax. The data includes a list of date-view pairs and a summary object at the end.

```
{ "daily_views": { "2013-12-23": 71937, "2013-12-22": 45513, "2013-12-21": 42604, "2013-12-20": 49818, "2013-12-27": 24018, "2013-12-26": 52410, "2013-12-25": 149222, "2013-12-24": 155846, "2013-11-08": 12107, "2013-12-04": 56896, "2013-12-09": 59081, "2013-12-08": 44951, "2014-01-07": 21874, "2014-01-04": 7355, "2014-01-05": 1127, "2014-01-02": 10365, "2014-01-03": 8884, "2014-01-01": 10204, "2013-12-29": 12922, "2014-01-08": 9520, "2014-01-09": 7638, "2013-12-28": 15599, "2014-01-11": 4192, "2014-01-10": 5716, "2013-11-28": 29190, "2013-11-29": 30154, "2013-11-22": 21886, "2013-11-23": 17400, "2013-11-20": 21785, "2013-11-21": 23641, "2013-11-26": 28943, "2013-11-27": 29713, "2013-11-24": 20694, "2013-11-25": 28842, "2013-12-30": 13762, "2013-12-31": 12640, "2013-10-30": 10763, "2013-10-31": 11784, "2013-12-12": 40398, "2013-12-13": 37021, "2013-12-10": 60227, "2013-12-11": 55353, "2013-12-16": 49399, "2013-12-17": 52023, "2013-12-14": 25774, "2013-12-15": 31081, "2013-11-30": 28342, "2013-12-18": 54102, "2013-12-19": 53252, "2013-10-29": 10128, "2013-10-28": 9174, "2013-10-25": 7572, "2013-10-24": 8599, "2013-10-27": 8158, "2013-10-26": 5793, "2013-10-21": 8854, "2013-10-20": 6787, "2013-10-23": 9192, "2013-10-22": 9406, "2013-12-01": 38329, "2013-12-03": 54187, "2013-12-02": 53076, "2013-12-05": 58304, "2013-11-09": 9884, "2013-12-07": 38798, "2013-12-06": 50741, "2013-11-04": 12389, "2013-11-05": 12932, "2013-11-06": 13742, "2013-11-07": 13779, "2013-11-01": 11202, "2013-11-02": 8360, "2013-11-03": 8784, "2013-10-15": 7319, "2013-10-16": 7291, "2013-10-17": 7604, "2013-10-18": 8081, "2013-10-19": 6456, "2013-11-13": 17449, "2013-11-12": 17265, "2013-11-11": 16466, "2013-11-10": 11655, "2013-11-17": 16732, "2013-11-16": 14661, "2013-11-15": 16633, "2013-11-14": 17279, "2013-11-19": 22304, "2013-11-18": 22116}, "project": "en", "month": "latest90", "rank": 6064, "title": "Christmas" }
```

What's the problem?

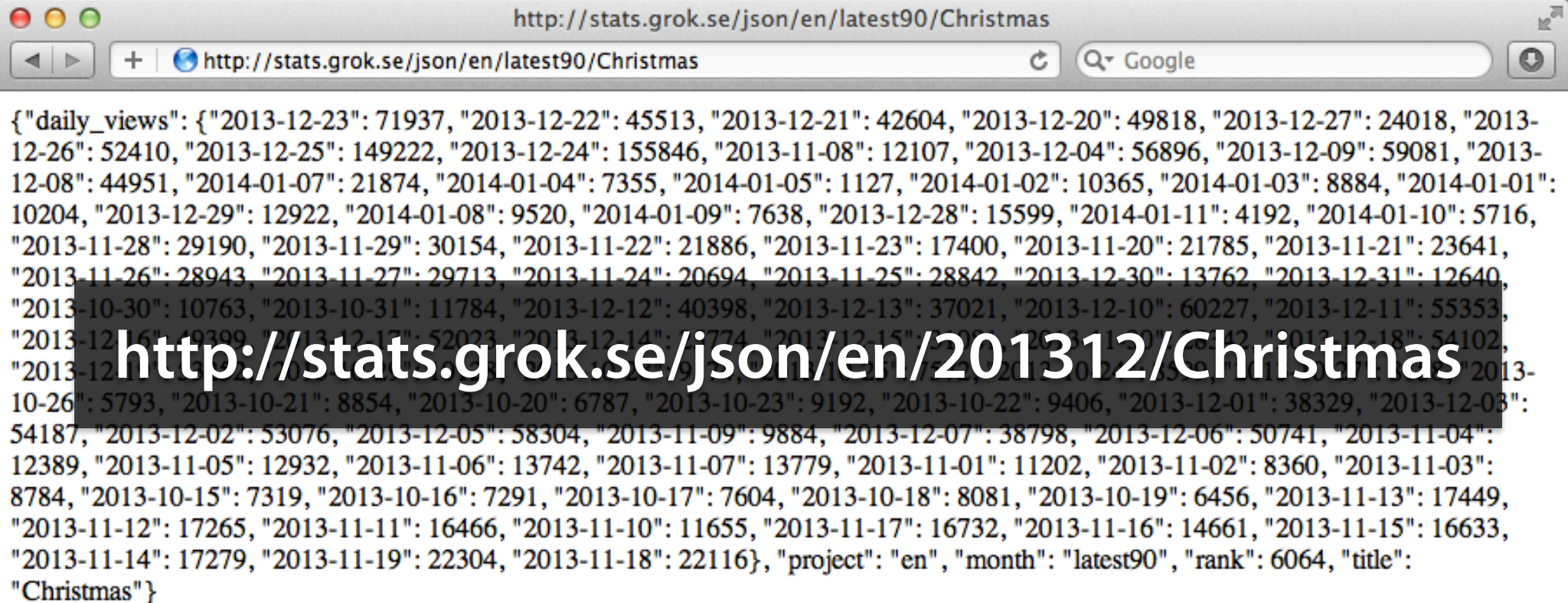


http://stats.grok.se/json/en/latest90/Christmas

```
{ "daily_views": { "2013-12-23": 71937, "2013-12-22": 45513, "2013-12-21": 42604, "2013-12-20": 49818, "2013-12-27": 24018, "2013-12-26": 52410, "2013-12-25": 149222, "2013-12-24": 155846, "2013-11-08": 12107, "2013-12-04": 56896, "2013-12-09": 59081, "2013-12-08": 44951, "2014-01-07": 21874, "2014-01-04": 7355, "2014-01-05": 1127, "2014-01-02": 10365, "2014-01-03": 8884, "2014-01-01": 10204, "2013-12-29": 12922, "2014-01-08": 9520, "2014-01-09": 7638, "2013-12-28": 15599, "2014-01-11": 4192, "2014-01-10": 5716, "2013-11-28": 29190, "2013-11-29": 30154, "2013-11-22": 21886, "2013-11-23": 17400, "2013-11-20": 21785, "2013-11-21": 23641, "2013-11-26": 28943, "2013-11-27": 29713, "2013-11-24": 20694, "2013-11-25": 28842, "2013-12-30": 13762, "2013-12-31": 12640, "2013-10-30": 10763, "2013-10-31": 11784, "2013-12-12": 40398, "2013-12-13": 37021, "2013-12-10": 60227, "2013-12-11": 55353, "2013-12-16": 49399, "2013-12-17": 52023, "2013-12-14": 25774, "2013-12-15": 31081, "2013-11-30": 28342, "2013-12-18": 54102, "2013-12-19": 53252, "2013-10-29": 10128, "2013-10-28": 9174, "2013-10-25": 7572, "2013-10-24": 8599, "2013-10-27": 8158, "2013-10-26": 5793, "2013-10-21": 8854, "2013-10-20": 6787, "2013-10-23": 9192, "2013-10-22": 9406, "2013-12-01": 38329, "2013-12-03": 54187, "2013-12-02": 53076, "2013-12-05": 58304, "2013-11-09": 9884, "2013-12-07": 38798, "2013-12-06": 50741, "2013-11-04": 12389, "2013-11-05": 12932, "2013-11-06": 13742, "2013-11-07": 13779, "2013-11-01": 11202, "2013-11-02": 8360, "2013-11-03": 8784, "2013-10-15": 7319, "2013-10-16": 7291, "2013-10-17": 7604, "2013-10-18": 8081, "2013-10-19": 6456, "2013-11-13": 17449, "2013-11-12": 17265, "2013-11-11": 16466, "2013-11-10": 11655, "2013-11-17": 16732, "2013-11-16": 14661, "2013-11-15": 16633, "2013-11-14": 17279, "2013-11-19": 22304, "2013-11-18": 22116}, "project": "en", "month": "latest90", "rank": 6064, "title": "Christmas" }
```

Not tabular

What's the problem?



The screenshot shows a web browser window with the address bar displaying `http://stats.grok.se/json/en/latest90/Christmas`. The page content is a large block of JSON text. A semi-transparent box is overlaid on the middle of the JSON, containing the URL `http://stats.grok.se/json/en/201312/Christmas` in white text.

```
{ "daily_views": { "2013-12-23": 71937, "2013-12-22": 45513, "2013-12-21": 42604, "2013-12-20": 49818, "2013-12-27": 24018, "2013-12-26": 52410, "2013-12-25": 149222, "2013-12-24": 155846, "2013-11-08": 12107, "2013-12-04": 56896, "2013-12-09": 59081, "2013-12-08": 44951, "2014-01-07": 21874, "2014-01-04": 7355, "2014-01-05": 1127, "2014-01-02": 10365, "2014-01-03": 8884, "2014-01-01": 10204, "2013-12-29": 12922, "2014-01-08": 9520, "2014-01-09": 7638, "2013-12-28": 15599, "2014-01-11": 4192, "2014-01-10": 5716, "2013-11-28": 29190, "2013-11-29": 30154, "2013-11-22": 21886, "2013-11-23": 17400, "2013-11-20": 21785, "2013-11-21": 23641, "2013-11-26": 28943, "2013-11-27": 29713, "2013-11-24": 20694, "2013-11-25": 28842, "2013-12-30": 13762, "2013-12-31": 12640, "2013-10-30": 10763, "2013-10-31": 11784, "2013-12-12": 40398, "2013-12-13": 37021, "2013-12-10": 60227, "2013-12-11": 55353, "2013-12-16": 49399, "2013-12-17": 52033, "2013-12-14": 12774, "2013-12-15": 12055, "2013-11-20": 21785, "2013-12-18": 54102, "2013-12-19": 54102, "2013-10-26": 5793, "2013-10-21": 8854, "2013-10-20": 6787, "2013-10-23": 9192, "2013-10-22": 9406, "2013-12-01": 38329, "2013-12-03": 54187, "2013-12-02": 53076, "2013-12-05": 58304, "2013-11-09": 9884, "2013-12-07": 38798, "2013-12-06": 50741, "2013-11-04": 12389, "2013-11-05": 12932, "2013-11-06": 13742, "2013-11-07": 13779, "2013-11-01": 11202, "2013-11-02": 8360, "2013-11-03": 8784, "2013-10-15": 7319, "2013-10-16": 7291, "2013-10-17": 7604, "2013-10-18": 8081, "2013-10-19": 6456, "2013-11-13": 17449, "2013-11-12": 17265, "2013-11-11": 16466, "2013-11-10": 11655, "2013-11-17": 16732, "2013-11-16": 14661, "2013-11-15": 16633, "2013-11-14": 17279, "2013-11-19": 22304, "2013-11-18": 22116}, "project": "en", "month": "latest90", "rank": 6064, "title": "Christmas" }
```

Not tabular

What's the problem?

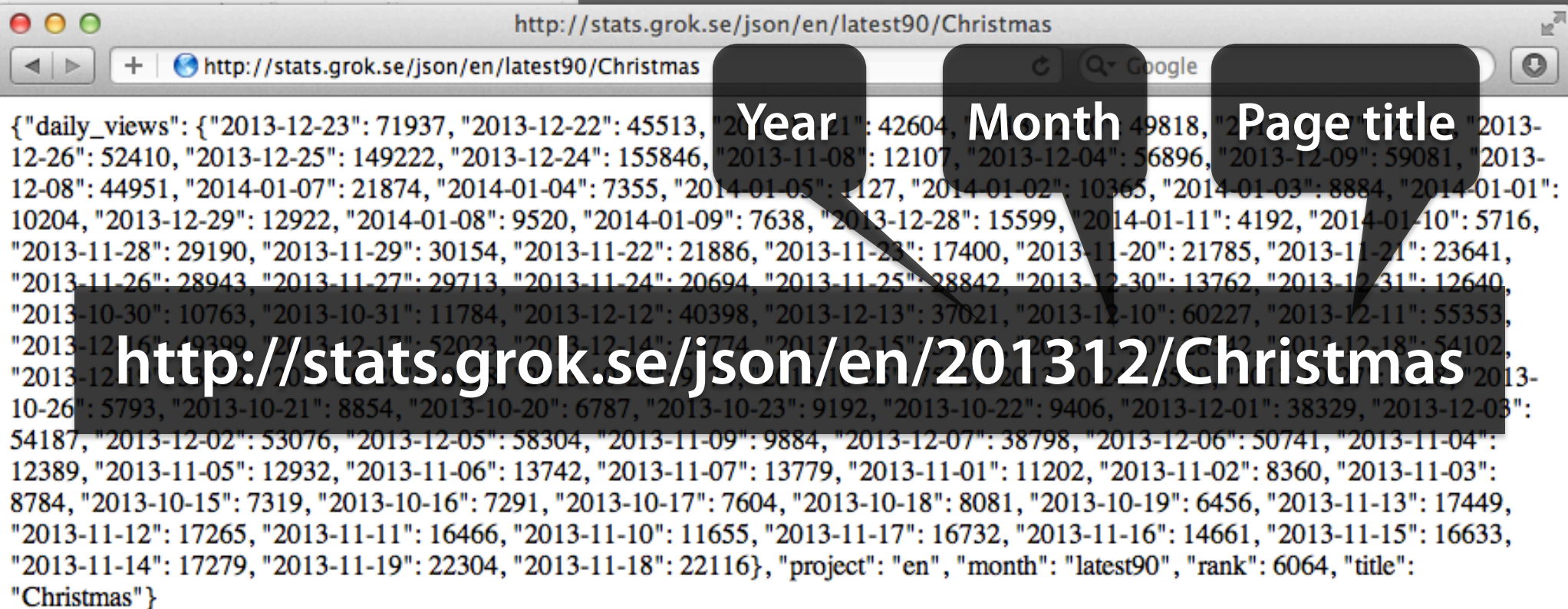
Year Month Page title

<http://stats.grok.se/json/en/201312/Christmas>

```
{
  "daily_views": {
    "2013-12-23": 71937,
    "2013-12-22": 45513,
    "2013-12-21": 42604,
    "2013-12-20": 49818,
    "2013-12-19": 52410,
    "2013-12-18": 54102,
    "2013-12-17": 52023,
    "2013-12-16": 49399,
    "2013-12-15": 5095,
    "2013-12-14": 774,
    "2013-12-13": 37021,
    "2013-12-12": 40398,
    "2013-12-11": 55353,
    "2013-12-10": 60227,
    "2013-12-09": 59081,
    "2013-12-08": 44951,
    "2013-12-07": 38798,
    "2013-12-06": 50741,
    "2013-12-05": 58304,
    "2013-12-04": 56896,
    "2013-12-03": 54187,
    "2013-12-02": 53076,
    "2013-12-01": 38329,
    "2013-11-30": 5793,
    "2013-11-29": 30154,
    "2013-11-28": 29190,
    "2013-11-27": 29713,
    "2013-11-26": 28943,
    "2013-11-25": 28842,
    "2013-11-24": 20694,
    "2013-11-23": 17400,
    "2013-11-22": 21886,
    "2013-11-21": 23641,
    "2013-11-20": 21785,
    "2013-11-19": 16633,
    "2013-11-18": 22116,
    "2013-11-17": 16732,
    "2013-11-16": 14661,
    "2013-11-15": 16633,
    "2013-11-14": 17279,
    "2013-11-13": 17449,
    "2013-11-12": 17265,
    "2013-11-11": 16466,
    "2013-11-10": 11655,
    "2013-11-09": 9884,
    "2013-11-08": 12107,
    "2013-11-07": 13779,
    "2013-11-06": 13742,
    "2013-11-05": 12932,
    "2013-11-04": 8784,
    "2013-11-03": 8360,
    "2013-11-02": 8360,
    "2013-11-01": 11202,
    "2013-10-31": 11784,
    "2013-10-30": 10763,
    "2013-10-29": 10763,
    "2013-10-28": 10763,
    "2013-10-27": 10763,
    "2013-10-26": 5793,
    "2013-10-25": 5793,
    "2013-10-24": 5793,
    "2013-10-23": 9192,
    "2013-10-22": 9406,
    "2013-10-21": 8854,
    "2013-10-20": 6787,
    "2013-10-19": 6456,
    "2013-10-18": 8081,
    "2013-10-17": 7604,
    "2013-10-16": 7291,
    "2013-10-15": 7319,
    "2013-10-14": 7319,
    "2013-10-13": 7319,
    "2013-10-12": 7319,
    "2013-10-11": 7319,
    "2013-10-10": 7319,
    "2013-10-09": 7319,
    "2013-10-08": 7319,
    "2013-10-07": 7319,
    "2013-10-06": 7319,
    "2013-10-05": 7319,
    "2013-10-04": 7319,
    "2013-10-03": 7319,
    "2013-10-02": 7319,
    "2013-10-01": 7319,
    "2013-09-30": 7319,
    "2013-09-29": 7319,
    "2013-09-28": 7319,
    "2013-09-27": 7319,
    "2013-09-26": 7319,
    "2013-09-25": 7319,
    "2013-09-24": 7319,
    "2013-09-23": 7319,
    "2013-09-22": 7319,
    "2013-09-21": 7319,
    "2013-09-20": 7319,
    "2013-09-19": 7319,
    "2013-09-18": 7319,
    "2013-09-17": 7319,
    "2013-09-16": 7319,
    "2013-09-15": 7319,
    "2013-09-14": 7319,
    "2013-09-13": 7319,
    "2013-09-12": 7319,
    "2013-09-11": 7319,
    "2013-09-10": 7319,
    "2013-09-09": 7319,
    "2013-09-08": 7319,
    "2013-09-07": 7319,
    "2013-09-06": 7319,
    "2013-09-05": 7319,
    "2013-09-04": 7319,
    "2013-09-03": 7319,
    "2013-09-02": 7319,
    "2013-09-01": 7319,
    "2013-08-31": 7319,
    "2013-08-30": 7319,
    "2013-08-29": 7319,
    "2013-08-28": 7319,
    "2013-08-27": 7319,
    "2013-08-26": 7319,
    "2013-08-25": 7319,
    "2013-08-24": 7319,
    "2013-08-23": 7319,
    "2013-08-22": 7319,
    "2013-08-21": 7319,
    "2013-08-20": 7319,
    "2013-08-19": 7319,
    "2013-08-18": 7319,
    "2013-08-17": 7319,
    "2013-08-16": 7319,
    "2013-08-15": 7319,
    "2013-08-14": 7319,
    "2013-08-13": 7319,
    "2013-08-12": 7319,
    "2013-08-11": 7319,
    "2013-08-10": 7319,
    "2013-08-09": 7319,
    "2013-08-08": 7319,
    "2013-08-07": 7319,
    "2013-08-06": 7319,
    "2013-08-05": 7319,
    "2013-08-04": 7319,
    "2013-08-03": 7319,
    "2013-08-02": 7319,
    "2013-08-01": 7319,
    "2013-07-31": 7319,
    "2013-07-30": 7319,
    "2013-07-29": 7319,
    "2013-07-28": 7319,
    "2013-07-27": 7319,
    "2013-07-26": 7319,
    "2013-07-25": 7319,
    "2013-07-24": 7319,
    "2013-07-23": 7319,
    "2013-07-22": 7319,
    "2013-07-21": 7319,
    "2013-07-20": 7319,
    "2013-07-19": 7319,
    "2013-07-18": 7319,
    "2013-07-17": 7319,
    "2013-07-16": 7319,
    "2013-07-15": 7319,
    "2013-07-14": 7319,
    "2013-07-13": 7319,
    "2013-07-12": 7319,
    "2013-07-11": 7319,
    "2013-07-10": 7319,
    "2013-07-09": 7319,
    "2013-07-08": 7319,
    "2013-07-07": 7319,
    "2013-07-06": 7319,
    "2013-07-05": 7319,
    "2013-07-04": 7319,
    "2013-07-03": 7319,
    "2013-07-02": 7319,
    "2013-07-01": 7319,
    "2013-06-30": 7319,
    "2013-06-29": 7319,
    "2013-06-28": 7319,
    "2013-06-27": 7319,
    "2013-06-26": 7319,
    "2013-06-25": 7319,
    "2013-06-24": 7319,
    "2013-06-23": 7319,
    "2013-06-22": 7319,
    "2013-06-21": 7319,
    "2013-06-20": 7319,
    "2013-06-19": 7319,
    "2013-06-18": 7319,
    "2013-06-17": 7319,
    "2013-06-16": 7319,
    "2013-06-15": 7319,
    "2013-06-14": 7319,
    "2013-06-13": 7319,
    "2013-06-12": 7319,
    "2013-06-11": 7319,
    "2013-06-10": 7319,
    "2013-06-09": 7319,
    "2013-06-08": 7319,
    "2013-06-07": 7319,
    "2013-06-06": 7319,
    "2013-06-05": 7319,
    "2013-06-04": 7319,
    "2013-06-03": 7319,
    "2013-06-02": 7319,
    "2013-06-01": 7319,
    "2013-05-31": 7319,
    "2013-05-30": 7319,
    "2013-05-29": 7319,
    "2013-05-28": 7319,
    "2013-05-27": 7319,
    "2013-05-26": 7319,
    "2013-05-25": 7319,
    "2013-05-24": 7319,
    "2013-05-23": 7319,
    "2013-05-22": 7319,
    "2013-05-21": 7319,
    "2013-05-20": 7319,
    "2013-05-19": 7319,
    "2013-05-18": 7319,
    "2013-05-17": 7319,
    "2013-05-16": 7319,
    "2013-05-15": 7319,
    "2013-05-14": 7319,
    "2013-05-13": 7319,
    "2013-05-12": 7319,
    "2013-05-11": 7319,
    "2013-05-10": 7319,
    "2013-05-09": 7319,
    "2013-05-08": 7319,
    "2013-05-07": 7319,
    "2013-05-06": 7319,
    "2013-05-05": 7319,
    "2013-05-04": 7319,
    "2013-05-03": 7319,
    "2013-05-02": 7319,
    "2013-05-01": 7319,
    "2013-04-30": 7319,
    "2013-04-29": 7319,
    "2013-04-28": 7319,
    "2013-04-27": 7319,
    "2013-04-26": 7319,
    "2013-04-25": 7319,
    "2013-04-24": 7319,
    "2013-04-23": 7319,
    "2013-04-22": 7319,
    "2013-04-21": 7319,
    "2013-04-20": 7319,
    "2013-04-19": 7319,
    "2013-04-18": 7319,
    "2013-04-17": 7319,
    "2013-04-16": 7319,
    "2013-04-15": 7319,
    "2013-04-14": 7319,
    "2013-04-13": 7319,
    "2013-04-12": 7319,
    "2013-04-11": 7319,
    "2013-04-10": 7319,
    "2013-04-09": 7319,
    "2013-04-08": 7319,
    "2013-04-07": 7319,
    "2013-04-06": 7319,
    "2013-04-05": 7319,
    "2013-04-04": 7319,
    "2013-04-03": 7319,
    "2013-04-02": 7319,
    "2013-04-01": 7319,
    "2013-03-31": 7319,
    "2013-03-30": 7319,
    "2013-03-29": 7319,
    "2013-03-28": 7319,
    "2013-03-27": 7319,
    "2013-03-26": 7319,
    "2013-03-25": 7319,
    "2013-03-24": 7319,
    "2013-03-23": 7319,
    "2013-03-22": 7319,
    "2013-03-21": 7319,
    "2013-03-20": 7319,
    "2013-03-19": 7319,
    "2013-03-18": 7319,
    "2013-03-17": 7319,
    "2013-03-16": 7319,
    "2013-03-15": 7319,
    "2013-03-14": 7319,
    "2013-03-13": 7319,
    "2013-03-12": 7319,
    "2013-03-11": 7319,
    "2013-03-10": 7319,
    "2013-03-09": 7319,
    "2013-03-08": 7319,
    "2013-03-07": 7319,
    "2013-03-06": 7319,
    "2013-03-05": 7319,
    "2013-03-04": 7319,
    "2013-03-03": 7319,
    "2013-03-02": 7319,
    "2013-03-01": 7319,
    "2013-02-28": 7319,
    "2013-02-27": 7319,
    "2013-02-26": 7319,
    "2013-02-25": 7319,
    "2013-02-24": 7319,
    "2013-02-23": 7319,
    "2013-02-22": 7319,
    "2013-02-21": 7319,
    "2013-02-20": 7319,
    "2013-02-19": 7319,
    "2013-02-18": 7319,
    "2013-02-17": 7319,
    "2013-02-16": 7319,
    "2013-02-15": 7319,
    "2013-02-14": 7319,
    "2013-02-13": 7319,
    "2013-02-12": 7319,
    "2013-02-11": 7319,
    "2013-02-10": 7319,
    "2013-02-09": 7319,
    "2013-02-08": 7319,
    "2013-02-07": 7319,
    "2013-02-06": 7319,
    "2013-02-05": 7319,
    "2013-02-04": 7319,
    "2013-02-03": 7319,
    "2013-02-02": 7319,
    "2013-02-01": 7319,
    "2013-01-31": 7319,
    "2013-01-30": 7319,
    "2013-01-29": 7319,
    "2013-01-28": 7319,
    "2013-01-27": 7319,
    "2013-01-26": 7319,
    "2013-01-25": 7319,
    "2013-01-24": 7319,
    "2013-01-23": 7319,
    "2013-01-22": 7319,
    "2013-01-21": 7319,
    "2013-01-20": 7319,
    "2013-01-19": 7319,
    "2013-01-18": 7319,
    "2013-01-17": 7319,
    "2013-01-16": 7319,
    "2013-01-15": 7319,
    "2013-01-14": 7319,
    "2013-01-13": 7319,
    "2013-01-12": 7319,
    "2013-01-11": 7319,
    "2013-01-10": 7319,
    "2013-01-09": 7319,
    "2013-01-08": 9520,
    "2013-01-07": 21874,
    "2013-01-06": 7319,
    "2013-01-05": 1127,
    "2013-01-04": 7355,
    "2013-01-03": 8884,
    "2013-01-02": 10365,
    "2013-01-01": 10204
  },
  "project": "en",
  "month": "latest90",
  "rank": 6064,
  "title": "Christmas"
}
```

Not tabular

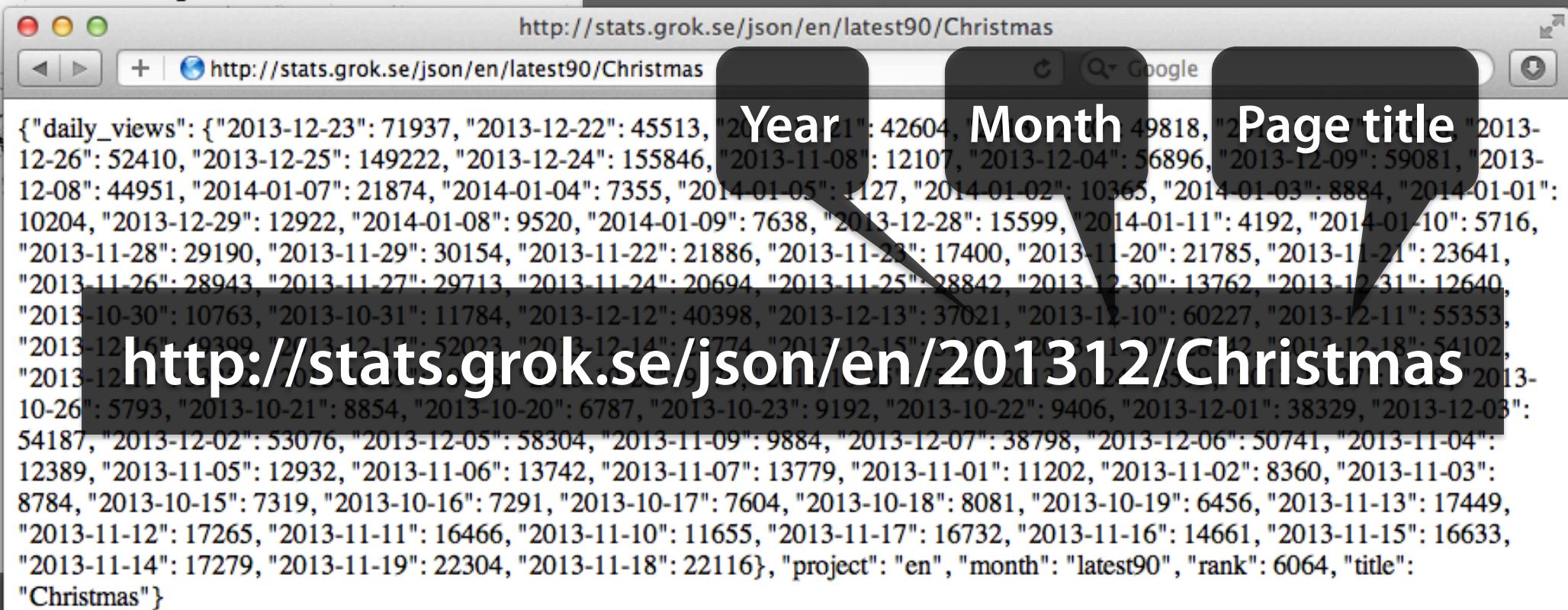
What's the problem?



Not tabular

Many pieces

What's the problem?



Not tabular

"APIs"

Many pieces

What's the problem?



A screenshot of a web browser window displaying a JSON API response. The browser's address bar shows the URL `http://stats.grok.se/json/en/latest90/Christmas`. The page content is a large, dense block of JSON text. Overlaid on this text is a semi-transparent dark grey rounded rectangle containing the text "Easy to overcome – if you can program".

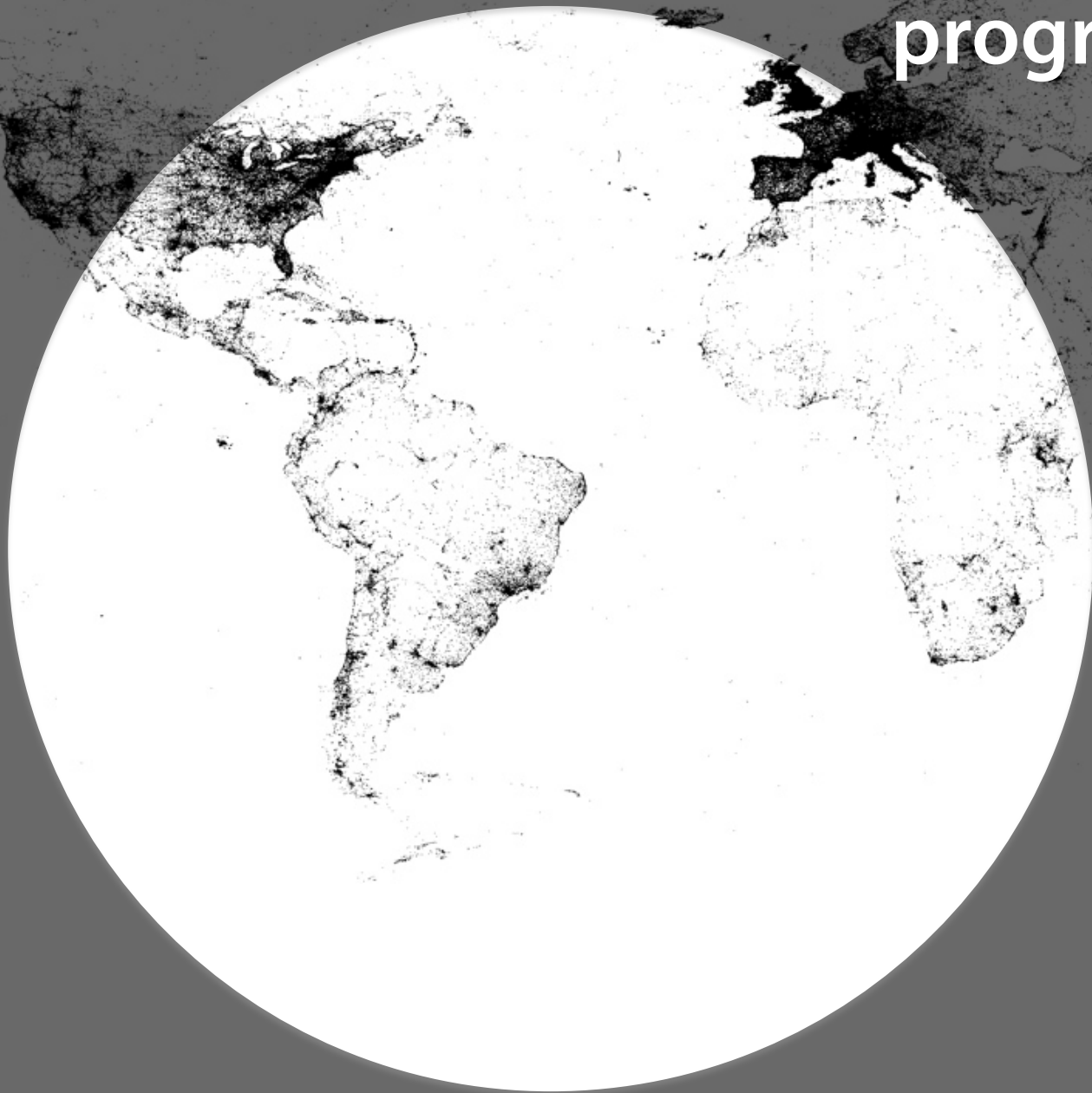
```
{ "daily_views": { "2013-12-23": 71937, "2013-12-22": 45513, "2013-12-21": 42604, "2013-12-20": 49818, "2013-12-27": 24018, "2013-12-26": 52410, "2013-12-25": 149222, "2013-12-24": 155846, "2013-11-08": 12107, "2013-12-04": 56896, "2013-12-09": 59081, "2013-12-08": 44951, "2014-01-07": 21874, "2014-01-04": 7355, "2014-01-05": 1127, "2014-01-02": 10365, "2014-01-03": 8884, "2014-01-01": 10204, "2013-12-29": 12922, "2014-01-08": 9520, "2014-01-09": 7638, "2013-12-28": 15599, "2014-01-11": 4192, "2014-01-10": 5716, "2013-11-28": 29190, "2013-11-29": 30154, "2013-11-22": 21886, "2013-11-23": 17400, "2013-11-20": 21785, "2013-11-21": 23641, "2013-11-26": 28943, "2013-11-27": 10773, "2013-11-24": 5064, "2013-12-02": 22842, "2013-12-31": 12640, "2013-10-30": 10763, "2013-10-31": 1134, "2013-12-03": 10303, "2013-12-10": 3021, "2013-12-11": 55353, "2013-12-16": 49399, "2013-12-17": 5023, "2013-12-14": 25774, "2013-12-15": 31081, "2013-11-30": 28342, "2013-12-18": 54102, "2013-12-19": 53252, "2013-10-29": 10128, "2013-10-28": 9174, "2013-10-25": 7572, "2013-10-24": 8599, "2013-10-27": 8158, "2013-10-26": 5793, "2013-10-27": 8854, "2013-10-20": 6787, "2013-10-23": 9192, "2013-10-22": 9406, "2013-12-01": 38329, "2013-12-03": 54187, "2013-12-02": 5176, "2013-11-25": 5804, "2013-11-09": 1384, "2013-11-07": 1870, "2013-11-06": 1261, "2013-11-04": 12389, "2013-11-05": 1232, "2013-11-01": 1342, "2013-11-07": 1779, "2013-11-01": 1102, "2013-11-02": 8460, "2013-11-03": 8784, "2013-10-15": 7319, "2013-10-16": 7291, "2013-10-17": 7604, "2013-10-18": 8081, "2013-10-19": 6456, "2013-11-13": 17449, "2013-11-12": 17265, "2013-11-11": 16466, "2013-11-10": 11655, "2013-11-17": 16732, "2013-11-16": 14661, "2013-11-15": 16633, "2013-11-14": 17279, "2013-11-19": 22304, "2013-11-18": 22116}, "project": "en", "month": "latest90", "rank": 6064, "title": "Christmas" }
```

Not tabular

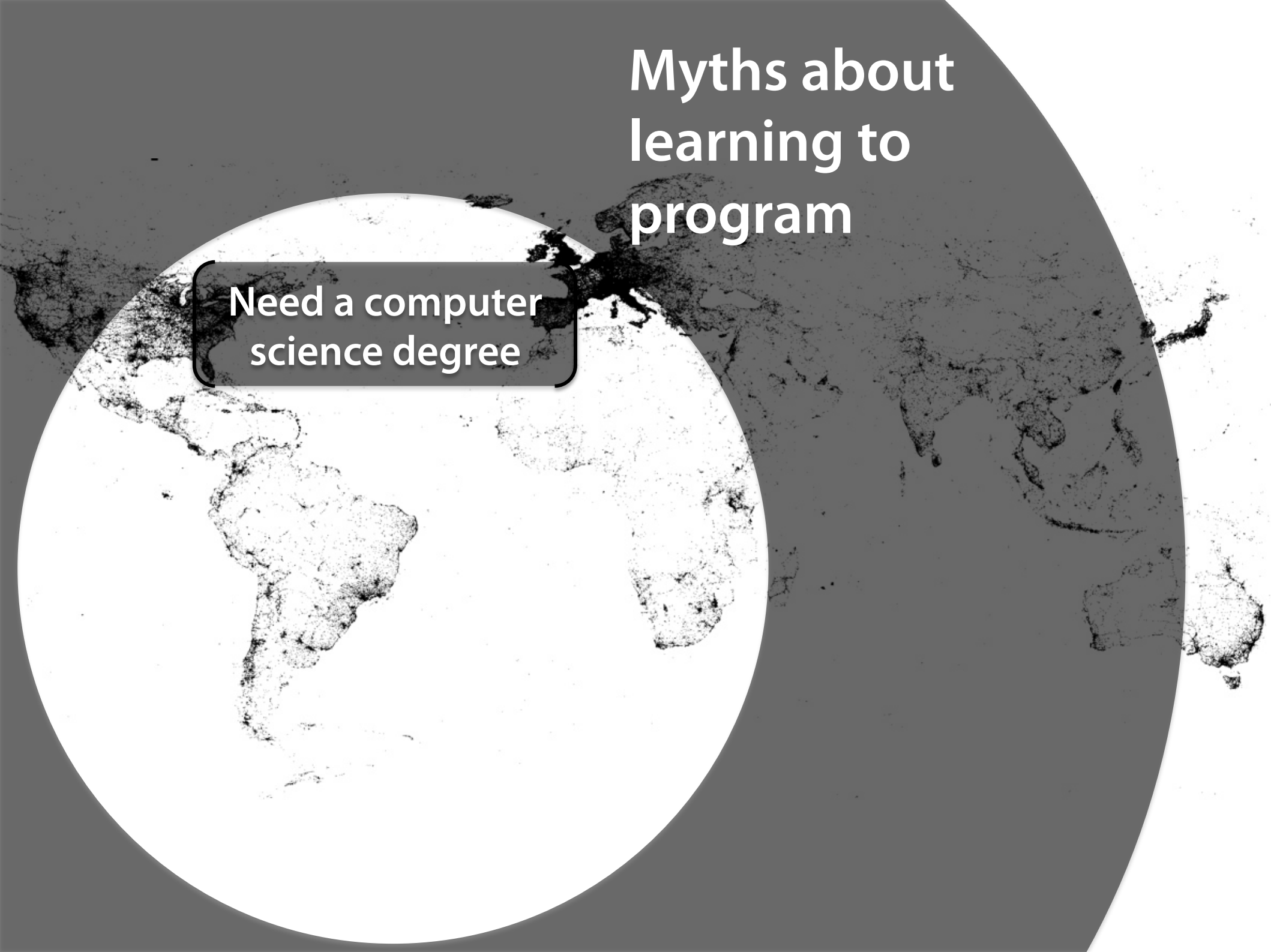
"APIs"

Many pieces

Myths about learning to program



Myths about learning to program



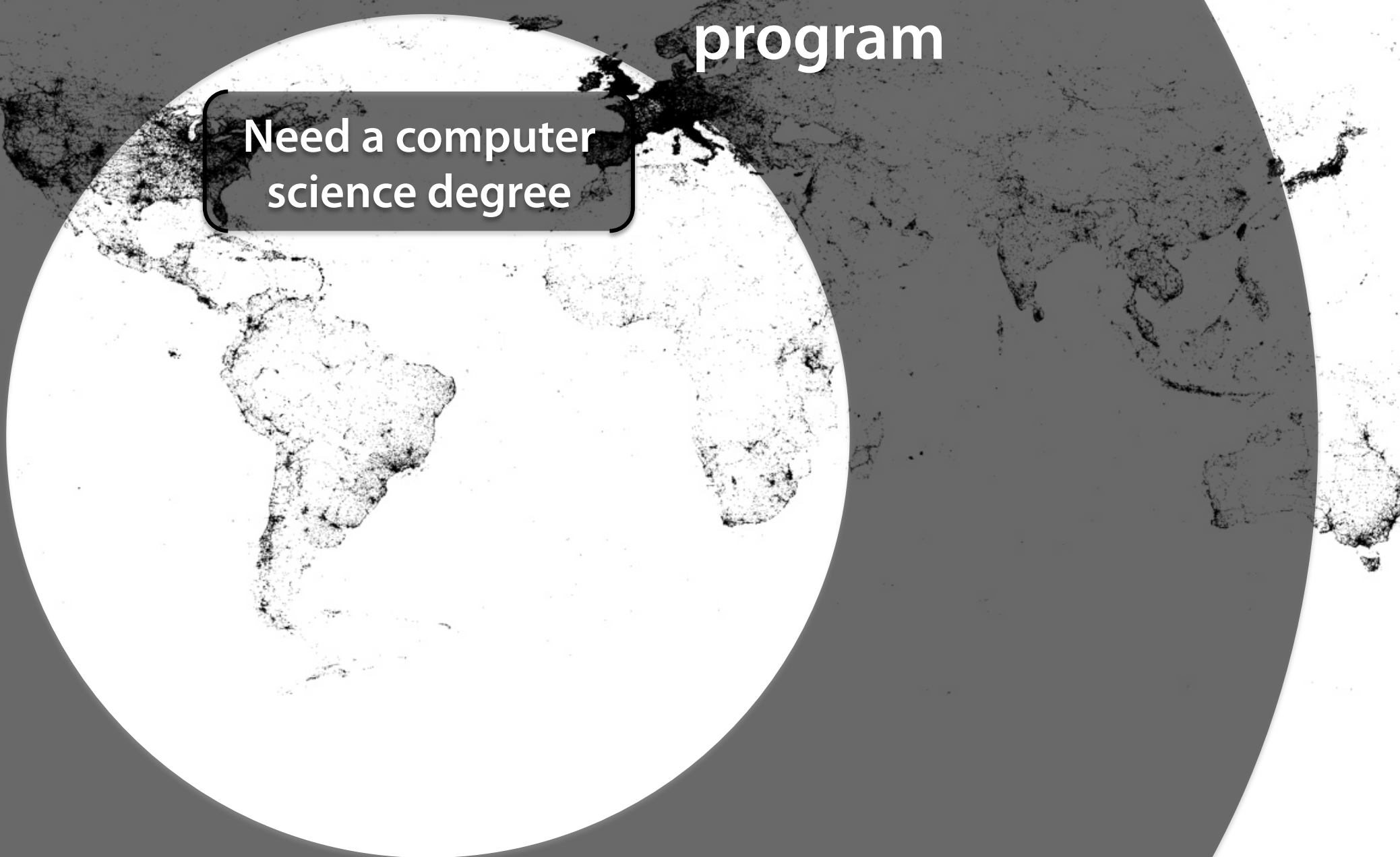
Need a computer
science degree



Photo: Dave Briccetti

Myths about learning to program

Need a computer
science degree



Myths about learning to program

Need a computer science degree

Takes too long

<https://www.codecademy.com/learn/python>

codecademy

Log in

Sign up



Python

Continue your learning by starting Python.

Start

Estimated time: 13 hours

Myths about learning to program

Need a computer science degree

Takes too long

Myths about learning to program

Need a computer science degree

Takes too long

I can't do it – I just get errors

Don't worry about errors!



Sam Grigg

@samgrigg



 Follow

How to write software:

- 1) Write a line of code
- 2) Google error message
- 3) Read stackoverflow
- 4) Goto: 1

Programming questions
and answers website

 Reply  Retweet  Favorited  More

1,316

RETWEETS

409

FAVORITES



7:48 PM - 25 Apr 2013

Myths about learning to program

Need a computer science degree

Takes too long

I can't do it – I just get errors

Myths about learning to program

Need a computer science degree

Takes too long

I can't do it – I just get errors

Can just use graphical interfaces instead

Free R packages: examples

```
install.packages(c("RCurl", "RJSONIO", "twitterR"))
```

Goes to get data from the
Internet for you

```
library(RCurl)
```

Free R packages: examples

```
install.packages(c("RCurl", "RJSONIO", "twitterR"))
```

Goes to get data from the Internet for you

```
library(RCurl)
```

```
library(RJSONIO)
```

Turns "JSON" into a table you can understand

Free R packages: examples

```
install.packages(c("RCurl", "RJSONIO", "twitterR"))
```

```
library(RCurl)
```

Goes to get data from the Internet for you

```
library(RJSONIO)
```

Turns "JSON" into a table you can understand

```
library(twitterR)
```

Goes to get data from Twitter for you

Free R packages: examples

```
install.packages(c("RCurl", "RJSONIO", "twitterR"))
```

```
library(RCurl)
```

Goes to get data from the Internet for you

```
library(RJSONIO)
```

Turns "JSON" into a table you can understand

```
library(twitterR)
```

Goes to get data from Twitter for you

... and lots more!

Myths about learning to program

Need a computer science degree

Takes too long

I can't do it – I just get errors

Can just use graphical interfaces instead

Myths about learning to program

Need a computer science degree

Takes too long

I can't do it – I just get errors

Can just use graphical interfaces instead

Only of use for new forms of data

What to learn: R and RStudio



<http://www.r-project.org/>
<http://www.rstudio.com/>

What to learn: Python

codecademy

Log in

Sign up



Python

Continue your learning by starting Python.

Start

<https://www.codecademy.com/learn/python>

<https://www.rstudio.com/online-learning/#R>



[rstudio::conf](#)

[Products](#)

[Resources](#)

[Pricing](#)

[About Us](#)

[Blogs](#)



R Programming

There are hundreds of websites that can help you learn R. Here's how you can use some of the best to become a productive R programmer.

Start by downloading [R](#) and [RStudio](#).

Learn the basics

Visit [DataCamp](#) or [Try R](#) to learn how to write basic R code. Both sites provide interactive lessons that will get you writing real code in minutes. They are a great place to make mistakes and test out new skills. You are told immediately when you go wrong and given a chance to fix your code.

Or try [Leada](#). Leada gets you programming in your own environment with videos and exercises. The first few courses are free and cover how to install R and RStudio.

Big Data: Measuring and Predicting Human Behaviour


Go to course – started 20 Apr

INTRODUCTION


Join us to explore how the vast amounts of data generated today can help us understand and even predict how humans behave.

WATCH THE TRAILER



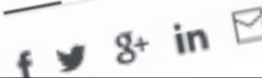
 FREE online course

 Duration: 9 weeks

 4 hours pw

 Certificates available

SHARE



Over 25,000 learners to date

Runs again in 2017

<https://www.futurelearn.com/courses/big-data>

Programming skills will unlock huge volumes
of interesting social data for you –
great return on investment



Data
Science
Lab