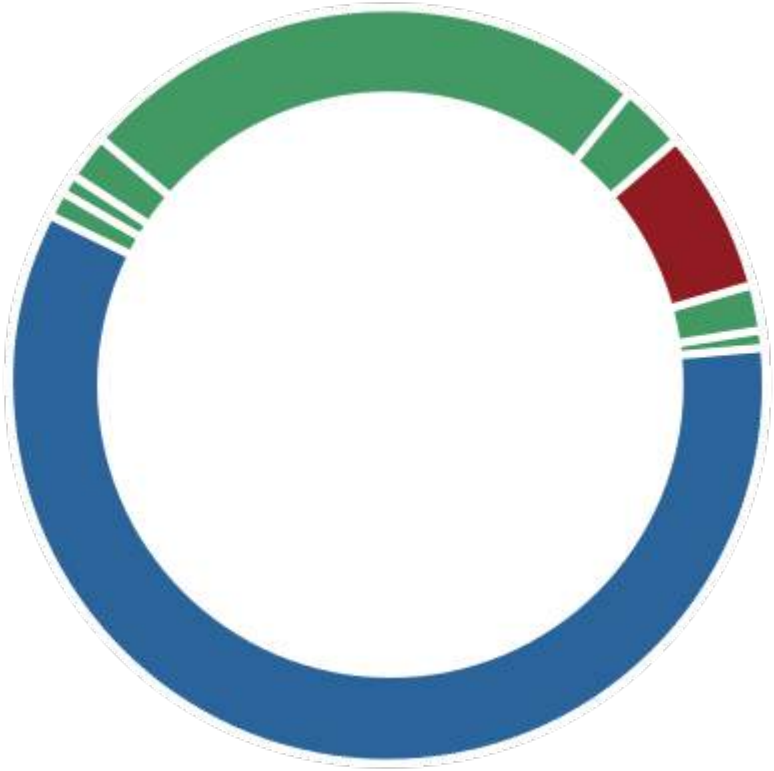
The Dashboarding Problem



Analytics team at Wikimedia Foundation

Goal:

Easily Visualize Editor Engagement for all Wikimedia projects

Side Goal:

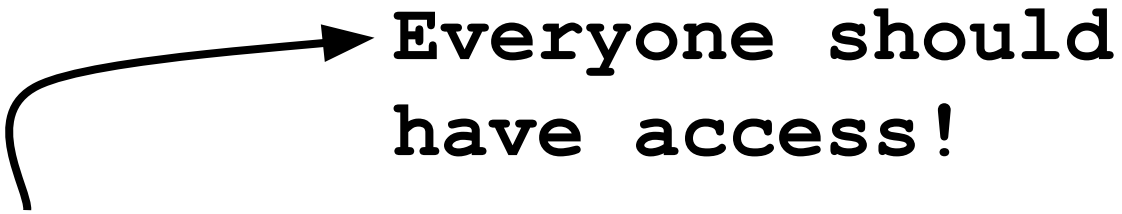
Explore Dashboarding Technologies

Date	project	pages created
20140101	arwiki	50
20140102	arwiki	40
…		
20140101	dewiki	100
20140102	dewiki	200

Problem #1

Problem #1

Access to Data



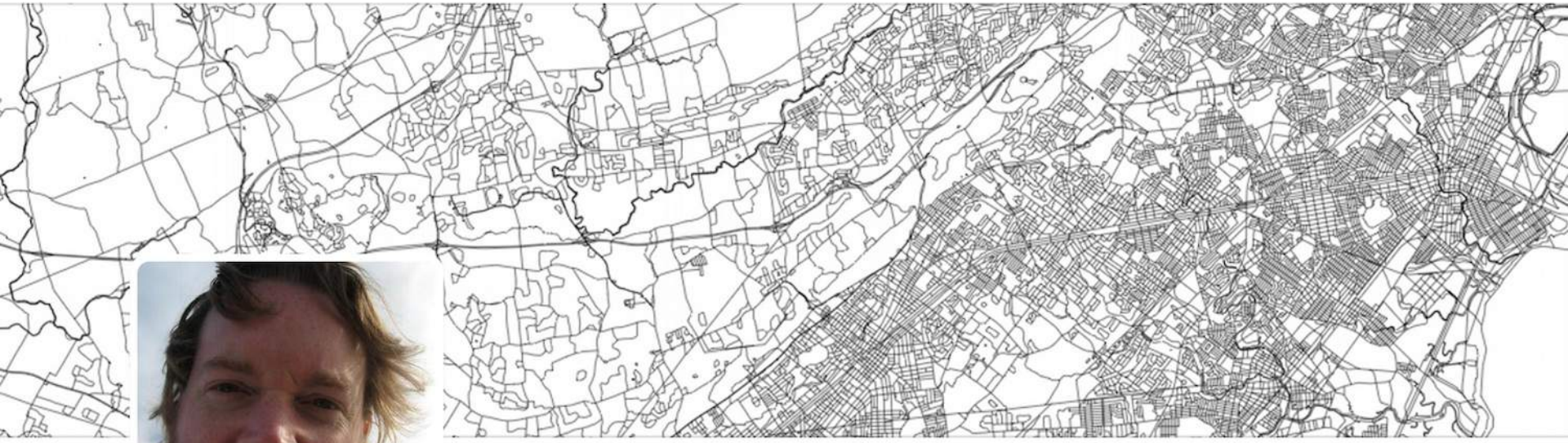
Easily Visualize Editor Engagement for **all** Wikimedia projects

Problem #1

Public Access to Data

Problem #2

Visualization



TWEETS 4,840 | PHOTOS/VIDEOS 95 | FOLLOWING 226 | FOLLOWERS 22.6K | More | Follow

Mike Bostock

@mbostock

Purveyor of fine misinformations. Graphics Editor @nytgraphics. Creator of #d3js.

San Francisco, CA

bost.ocks.org

Joined December 2006

Tweets | Tweets & replies

Pinned Tweet

Mike Bostock @mbostock · Jun 26

New post - Visualizing Algorithms - using vision to study processes & systems. bost.ocks.org/mike/algorithm… #eyeo2014

Visualization is an **easier** problem than **data access**

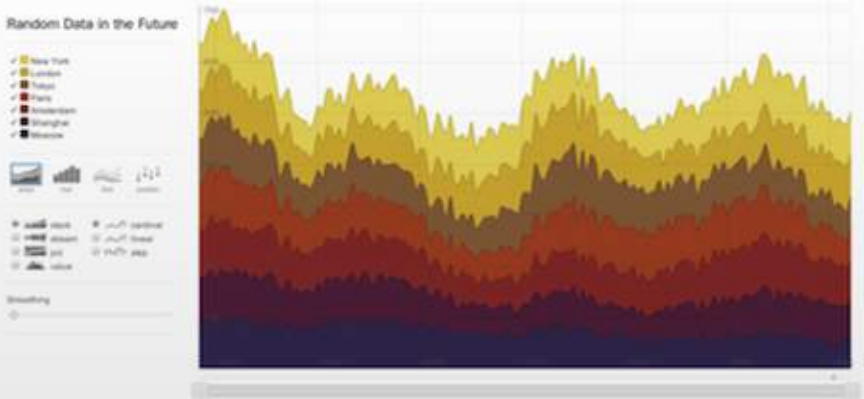
This JavaScript...

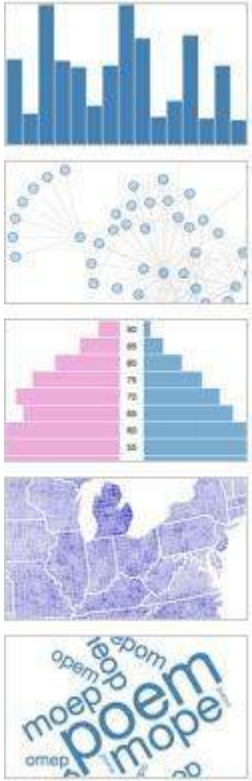
```
new Dygraph(div, "ny-vs-sf.txt", {
  legend: 'always',
  title: 'NYC vs. SF',
  showRoller: true,
  rollPeriod: 14,
  customBars: true,
  ylabel: 'Temperature (F)',
});
```

...makes this chart!

Rickshaw is a JavaScript toolkit for creating interactive time series graphs.

Rickshaw on GitHub





vega

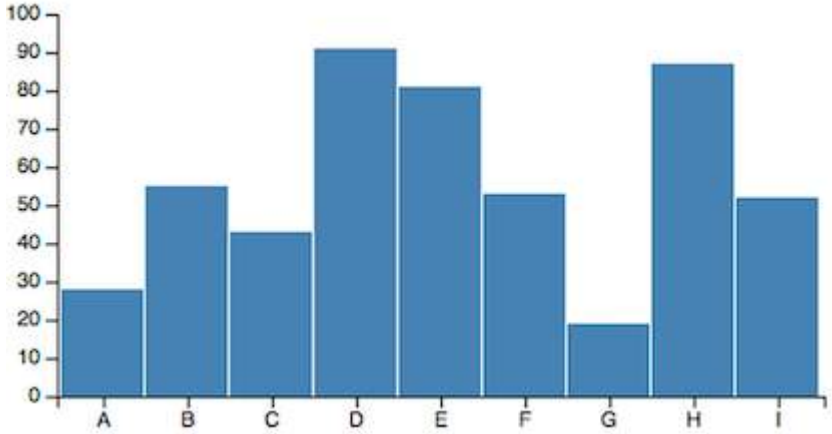
vega.min.js (125
Source (GitHub)

Fork me on GitHub

Vega is a visualization grammar, a declarative format for creating, saving and sharing visualization designs.

With Vega you can describe data visualizations in a JSON format, and generate interactive views using either HTML5 Canvas or SVG.

Read the tutorial, browse the documentation, join the discussion, and explore visualizations using the web-based Vega Editor.




```
{
  "width": 400,
  "height": 200,
  "padding": {"top": 10, "left": 30, "bottom": 20, "right": 10},
  "data": [
    {
      "name": "table",
      "values": [
        {"x":"A", "y":28}, {"x":"B", "y":55}, {"x":"C", "y":43},
        {"x":"D", "y":91}, {"x":"E", "y":81}, {"x":"F", "y":53},
        {"x":"G", "y":19}, {"x":"H", "y":87}, {"x":"I", "y":52}
      ]
    }
  ],
  "scales": [
    {"name":"x", "type":"ordinal", "range":"width", "domain":{"data":
    {"name":"y", "range":"height", "nice":true, "domain":{"data":"tab
  ],
  "axes": [
    {"type":"x", "scale":"x"},
    {"type":"y", "scale":"y"}
  ],
  "marks": [
    {
      "type": "rect",
      "from": {"data":"table"},
      "properties": {
        "enter": {
          "x": {"scale":"x", "field":"data.x"},
          "width": {"scale":"x", "band":true, "offset":-1},
          "y": {"scale":"y", "field":"data.y"},
          "y2": {"scale":"y", "value":0}
        },
        "update": { "fill": {"value":"steelblue"} },
        "hover": { "fill": {"value":"red"} }
      }
    }
  ]
}
```

Problem #3

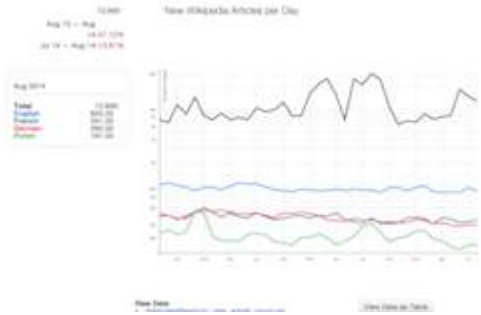
Problem #3

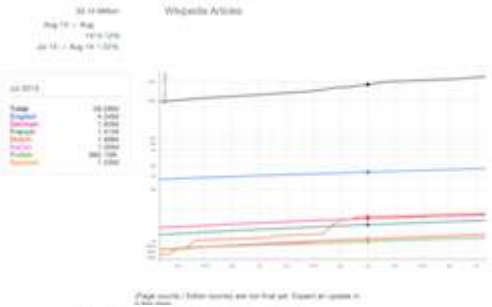
Information Architecture

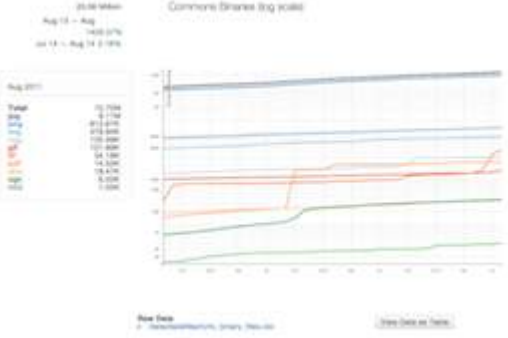
How do you explore data for more than 800 projects?

Forever
Scrolling
Creative Digital Experts

http://reportcard.wmflabs.org/#secondary-graphs-tab


New Wikipedia Articles per Day


Wikipedia Articles


Commons Binaries (log scale)

#1 Access to Data

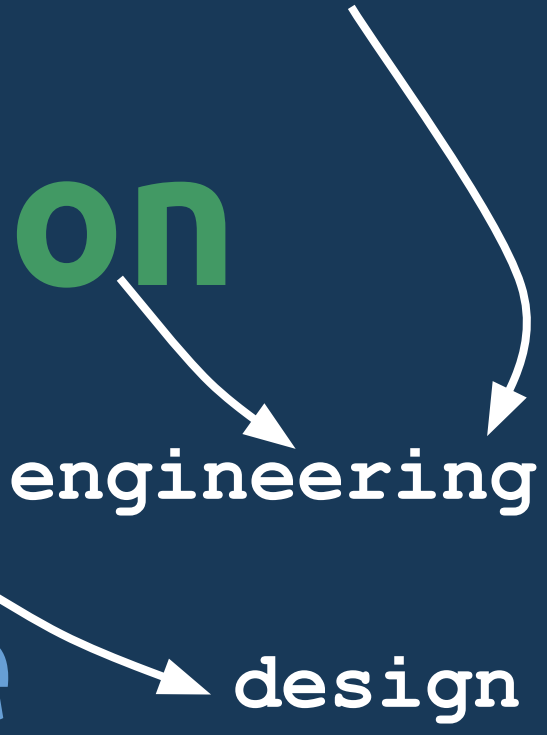
#2 Visualization

#3 Information Architecture

#1 Access to Data

#2 Visualization

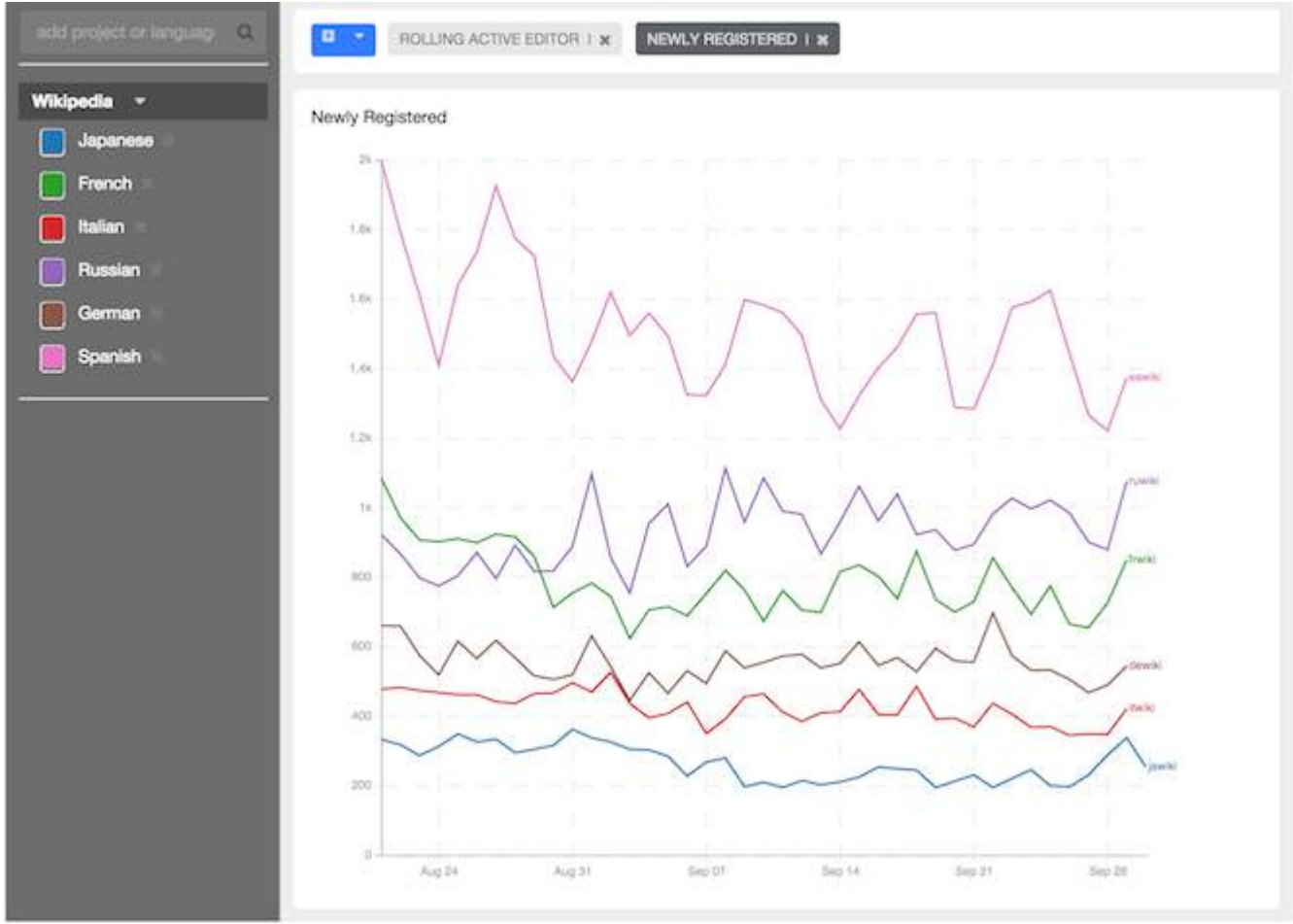
#3 Information Architecture



#3 Information Architecture

Design knows best

https://metrics.wmflabs.org/static/public/dash/



#1 Access to Data

Tech Decision: Make all data available via HTTP

https://metrics.wmflabs.org/static/public/datafiles/

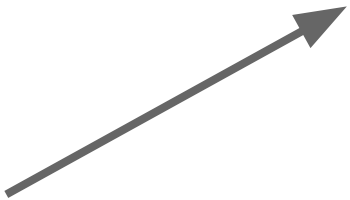
Index of /static/public/datafiles

Name	Last modified	Size	Description
Parent Directory		-	
NamespaceEdits/	04-Sep-2014 19:24	-	
NewlyRegistered/	21-Aug-2014 18:04	-	
PagesCreated/	04-Sep-2014 19:29	-	
RollingActiveEditor/	21-Aug-2014 18:07	-	
RollingNewActiveEditor/	12-Sep-2014 17:08	-	
RollingSurvivingNewActiveEditor/	12-Sep-2014 17:05	-	
available-metrics.json	16-Sep-2014 13:58	1.2K	
available-projects.json	12-Sep-2014 20:38	229K	
defaultDashboard.json	12-Sep-2014 20:38	192	

Apache/2.2.22 (Ubuntu) Server at metrics.wmflabs.org Port 80

Data is available in JSON format

```
{
    "result": {
        "Sum": {
            "newly_registered": {
                "2014-08-26 00:00:00": 364.0,
                "2014-08-24 00:00:00": 331.0,
                "2014-08-21 00:00:00": 382.0,
                "2014-08-30 00:00:00": 383.0,
                "2014-09-25 00:00:00": 372.0,
                "2014-08-22 00:00:00": 367.0,
                "2014-09-24 00:00:00": 410.0,
                "2014-09-06 00:00:00": 370.0,
                "2014-09-09 00:00:00": 370.0,
                "2014-09-27 00:00:00": 321.0,
                "2014-09-05 00:00:00": 317.0
            }
        }
    },
    "parameters": {
        "Cohort": "arwiki",
        "Created On": "2014-09-13 00:00:00",
        "Metric_end_date": "2014-09-13 00:00:00",
        "Metric_start_date": "2014-09-12 00:00:00",
        "Cohort Size": 0,
        "Metric": "NewlyRegistered"
    }
}
```

Includes metadata

Data is available to everyone.

Why not?...

Server-less

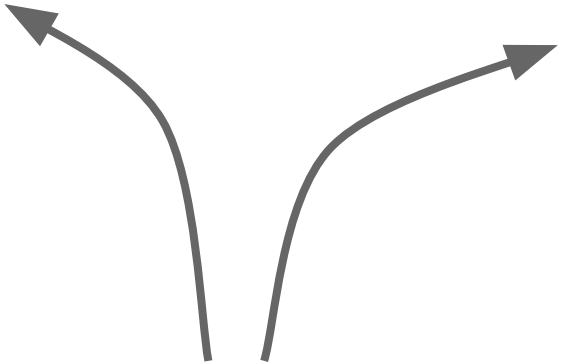
Why not?...

Server-less

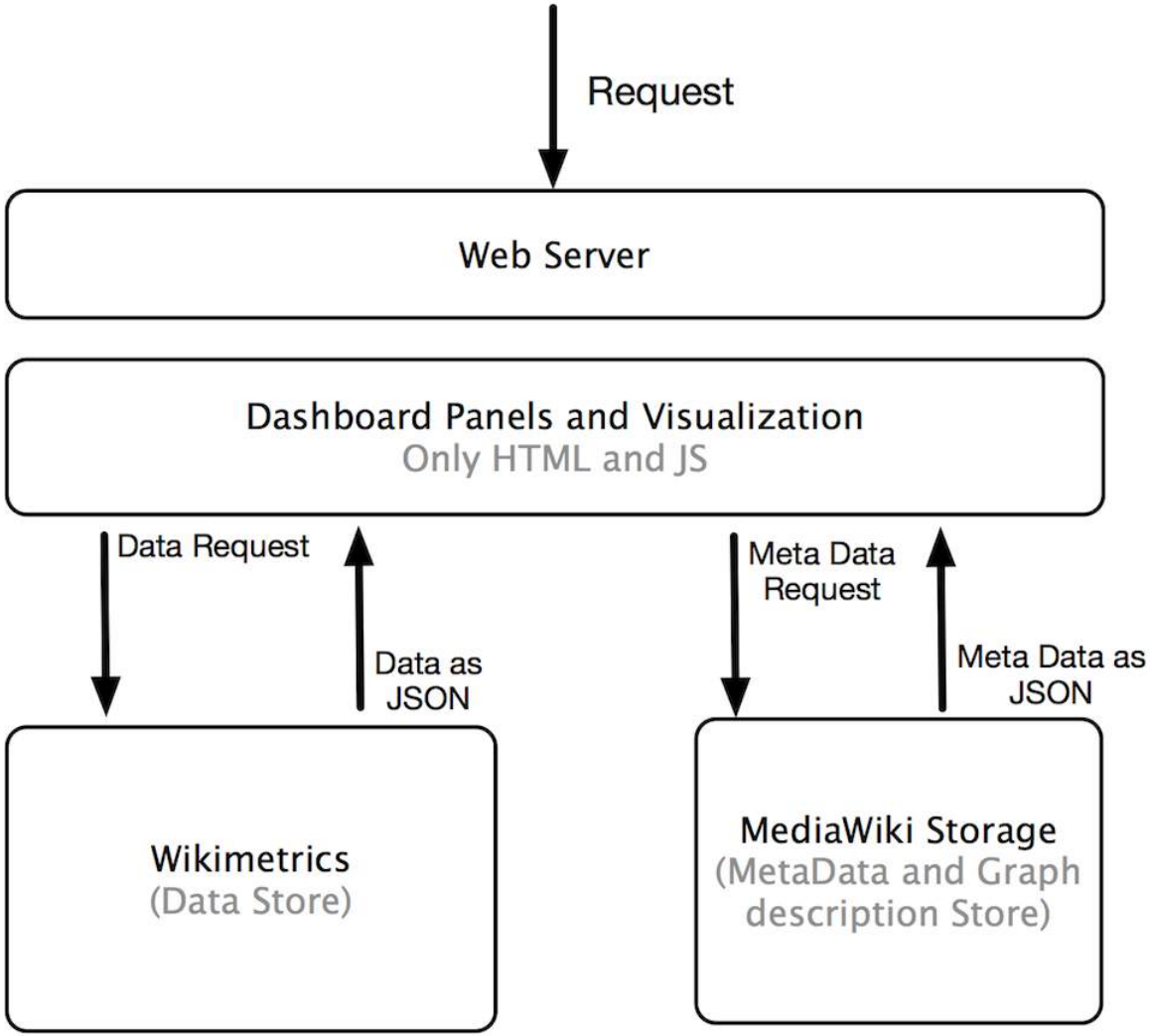
What is the middle tier for anyways?

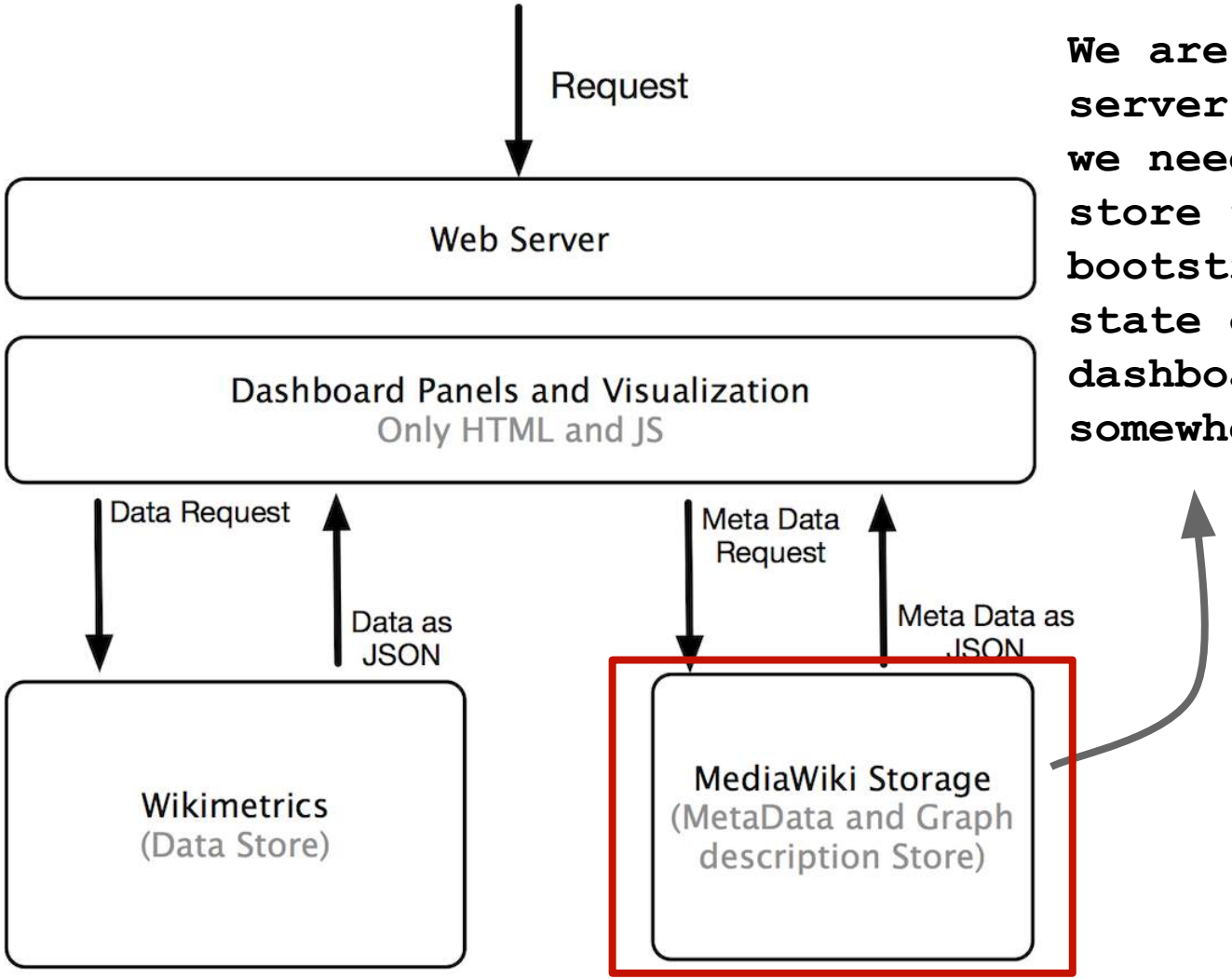
No authentication needed

Decouples visualization piece from data gathering piece



Server-less

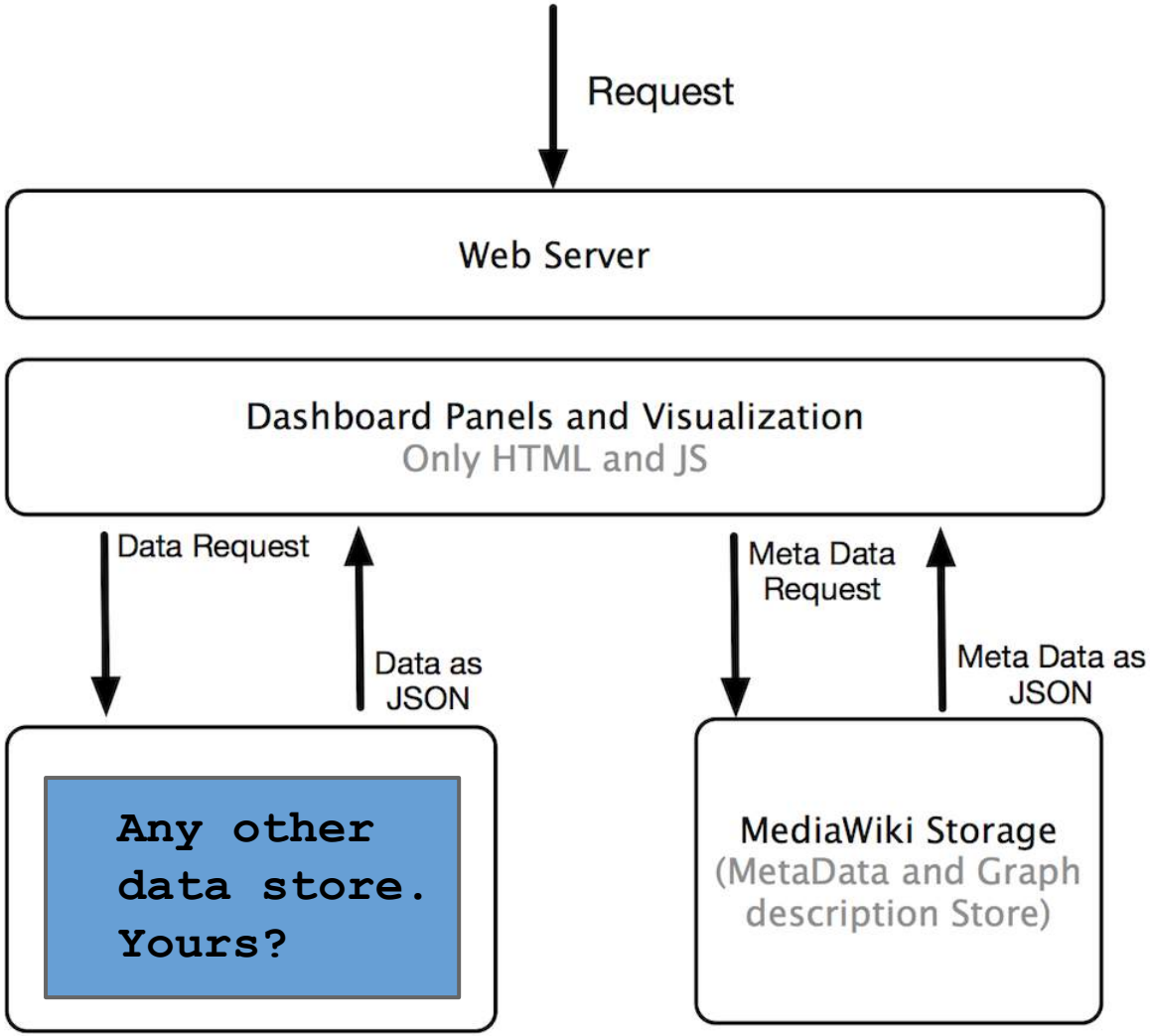

Request
Web Server
Dashboard Panels and Visualization
Only HTML and JS
Data Request
Data as JSON
Meta Data Request
Meta Data as JSON
Wikimetrics
(Data Store)
MediaWiki Storage
(MetaData and Graph description Store)



We are server-less, we need to store the bootstrap state of the dashboard somewhere

https://metrics.wmflabs.org/static/public/dash/


```
{
    "defaultProjects": [
        "enwiki", "eswiki", "dewiki", "frwiki", "ruwiki",
"jawiki", "itwiki"
    ],
    "defaultMetrics": [
        "RollingActiveEditor", "NewlyRegistered"
    ]
}
```


Request
Web Server
Dashboard Panels and Visualization
Only HTML and JS
Data Request
Data as JSON
Meta Data Request
Meta Data as JSON
Any other data store. Yours?
MediaWiki Storage
(MetaData and Graph description Store)

Technology Stack

Must Haves:

Package Manager
Dependency Loading
Data Binding Library
DOM manipulation and ajax
Testing

Technology Stack

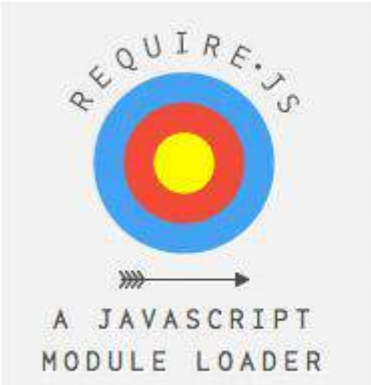
Micro libraries: “each library does 1 thing well”

Bower

A package manager for the web

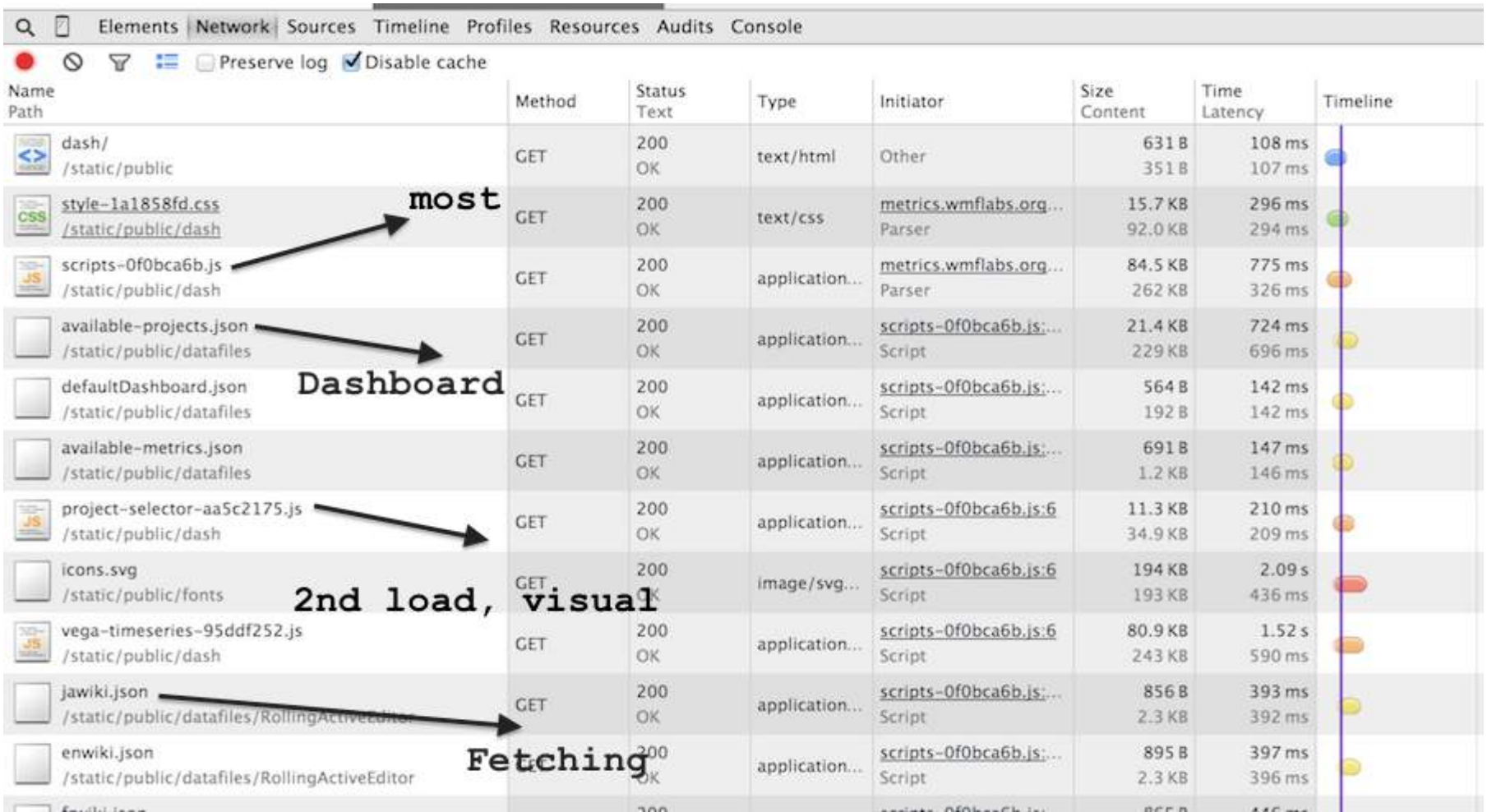
By Twitter

```
> Bower install
```


REQUIRE.JS
A JAVASCRIPT
MODULE LOADER

Jasmine

https://metrics.wmflabs.org/static/public/dash/

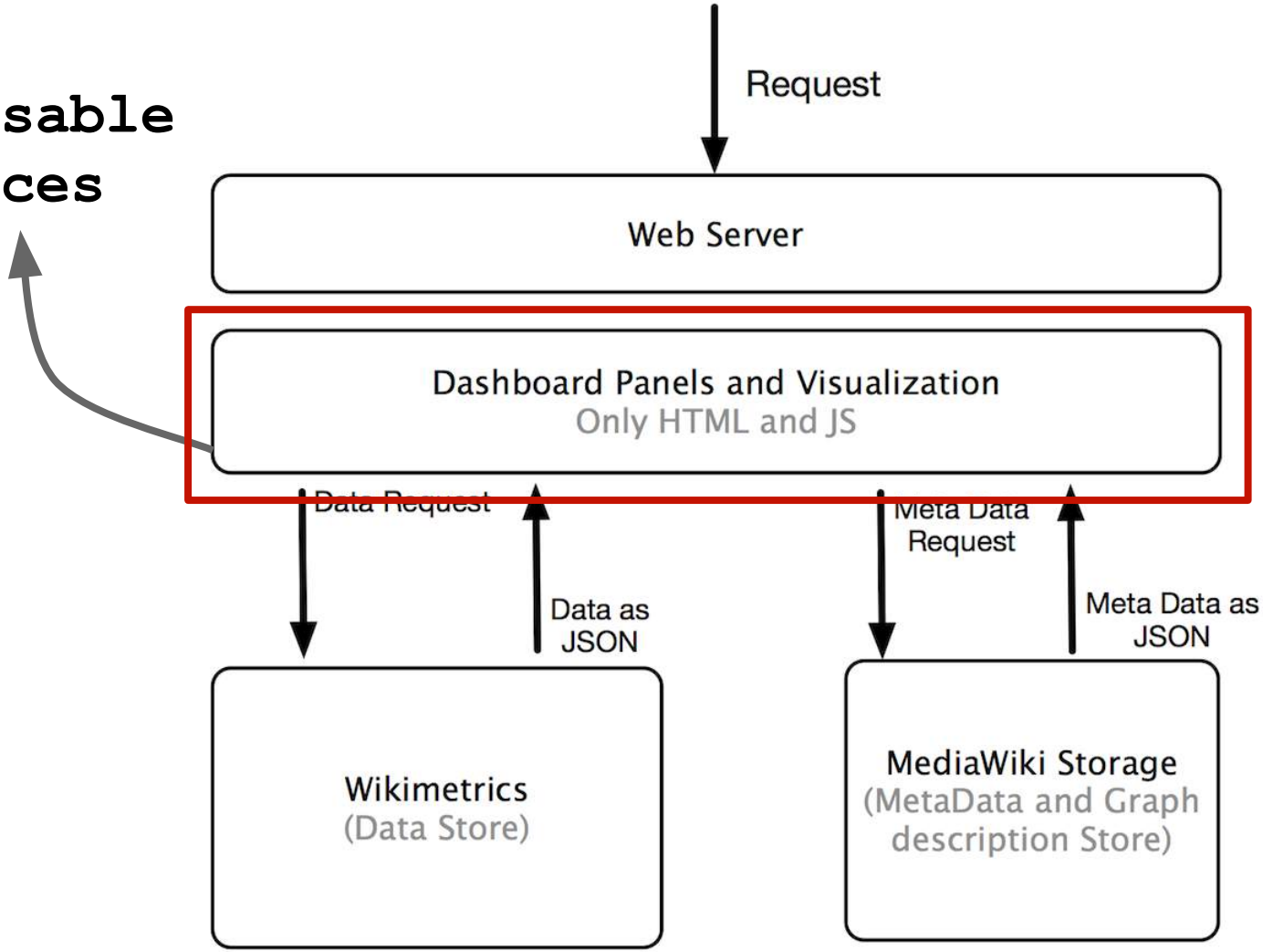


Knockout.

The Of Knockout Components

Knockout is not Polymer

Reusable pieces



ch
Projects
outreach 1 language
Languages
Achinese 1 project
Alemannisch 4 projects
Chamorro 3 projects
Chavacano de Zamboanga 1 project
Chechen 1 project
Cherokee 2 projects
Cheyenne 1 project
Chinese 7 projects
Rolling Active Editor
Newly Registered
ng Active Editor
enwiki
dewiki
itwiki
Aug 24
Aug 31
Sep 07
Sep 14
Sep 21
Sep 28

Show me the code….

```
<section class="ui six column celled page grid">
    <div class="row">
        <section class="column projects-container">
            <project-selector params="projectOptions:projectOptions,
                                      languageOptions:languageOptions,
                                      defaultProjects: defaultProjects,
                                      reverseLookup: reverseLookup,
                                      selectedProjects: selectedProjects">


            </project-selector>
        </section>
        <section class="five wide column main-container">
            <div class="ui segment">
                <metric-selector params="selectedMetric: selectedMetric,
                                         metrics: metrics,
                                         defaultSelection: defaultMetrics">
                </metric-selector>
                <time-selector></time-selector>
            </div>
            <div class="ui segment graph">
                <wikimetrics-visualizer params="projects: selectedProjects,
                                                metric: selectedMetric">
                </wikimetrics-visualizer>
            </div>
        </section>
    </div>
</section>
```

> *On the AngularJS team, we rely on testing and we always seek better tools to make our life easier. That's why we created Karma - a test runner that fits all our needs.*

View project on GitHub

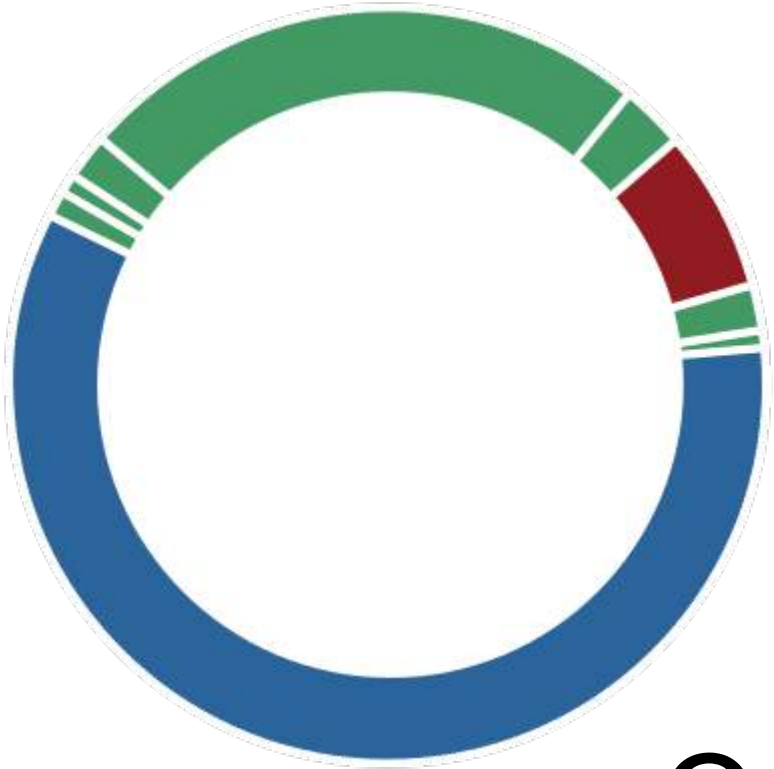
npm install karma

Dashiki: a thin wrapper on top of OS technology + information architecture. We can use **dashiki** to build **other** dashboards.

https://github.com/wikimedia/analytics-dashiki

Lessons:

Server-less = Reduced complexity
Easy(est) to deploy
CORS
No error log for now
Bootstrap default state
Keep things simple. No authentication.



Questions?