Visualizing code bases

@codingandrey

About me

Andrey Adamovich

- Bio: Developer, coach, speaker, author
- Company: Aestas/IT (http://aestasit.com)

Contact details

- E-mail: andrey@aestasit.com
- Linkedin: http://www.linkedin.com/in/andreyadamovich
- Lanyrd: http://lanyrd.com/profile/andrey-adamovich
- GitHub: https://github.com/aadamovich
- SO: http://stackoverflow.com/users/162792/andrey-adamovich
- Twitter: @codingandrey, @aestasit

Let's start!

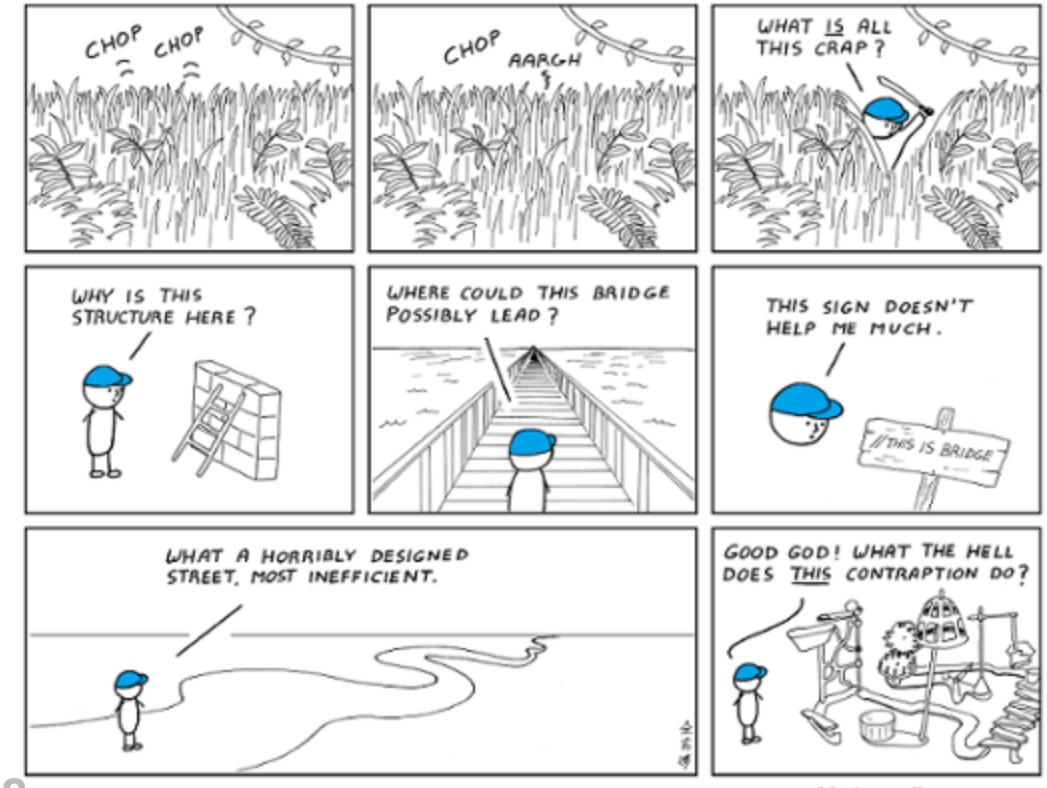
Background

- Find new clients
- Face big code bases
- Need quick analysis
- Need quick results

Code size

- Google: 2 billions LOC
- Facebook: 61 million LOC
- Me: from 20K to 2M LOC

I hate reading...



I hate reading
other people's code.

Reading code

```
01. I want this to work
02. This just works
03. Working is the definition of this block
04. Work! You, stupid!
05. The work should be here
```

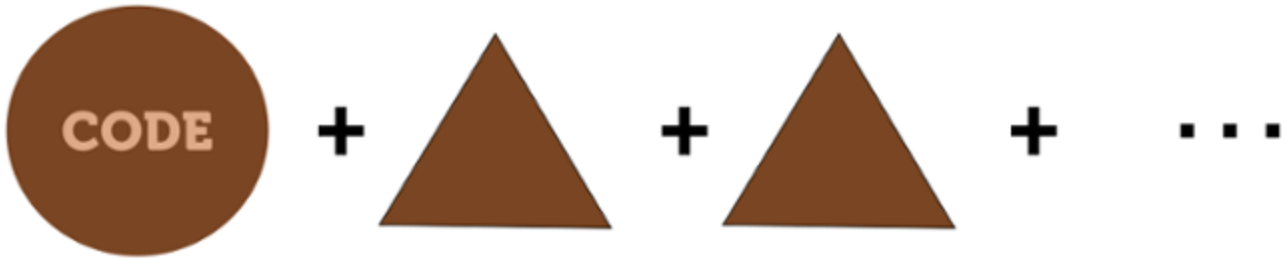
It's life, Jim, but not as we know it!

Code is data!


Well, big data

Snapshot

Temporal



HISTORY

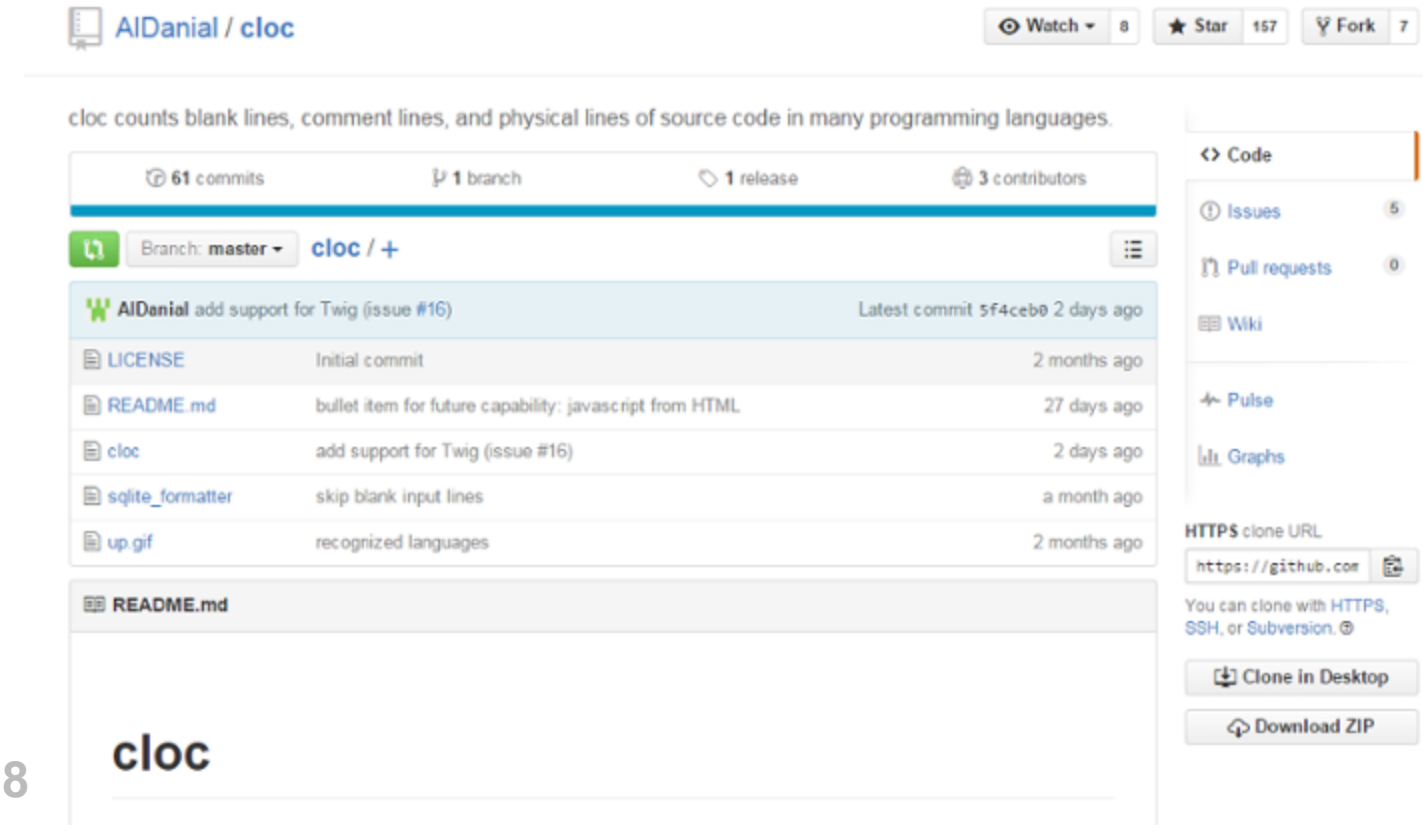
First steps

Count the lines!

Size does matter

- Gives you an estimate (on how much reading is needed).
- Most of the code bases are **polyglot**. Ratio between languages can tell a lot.
- Ratio between test code, comments, blank lines is also interesting.

cloc


AlDanial / cloc
Watch 8
Star 157
Fork 7
cloc counts blank lines, comment lines, and physical lines of source code in many programming languages.
61 commits
1 branch
1 release
3 contributors
Branch: master
cloc / +
AlDanial add support for Twig (issue #16)
Latest commit 5f4ceb0 2 days ago
LICENSE Initial commit 2 months ago
README.md bullet item for future capability: javascript from HTML 27 days ago
cloc add support for Twig (issue #16) 2 days ago
sqlite_formatter skip blank input lines a month ago
up.gif recognized languages 2 months ago
README.md
cloc
Code
Issues 5
Pull requests 0
Wiki
Pulse
Graphs
HTTPS clone URL
https://github.com
You can clone with HTTPS, SSH, or Subversion.
Clone in Desktop
Download ZIP

Usage

```
01. cloc --help
02. cloc --write-lang-def=lang.defs
```

Usage

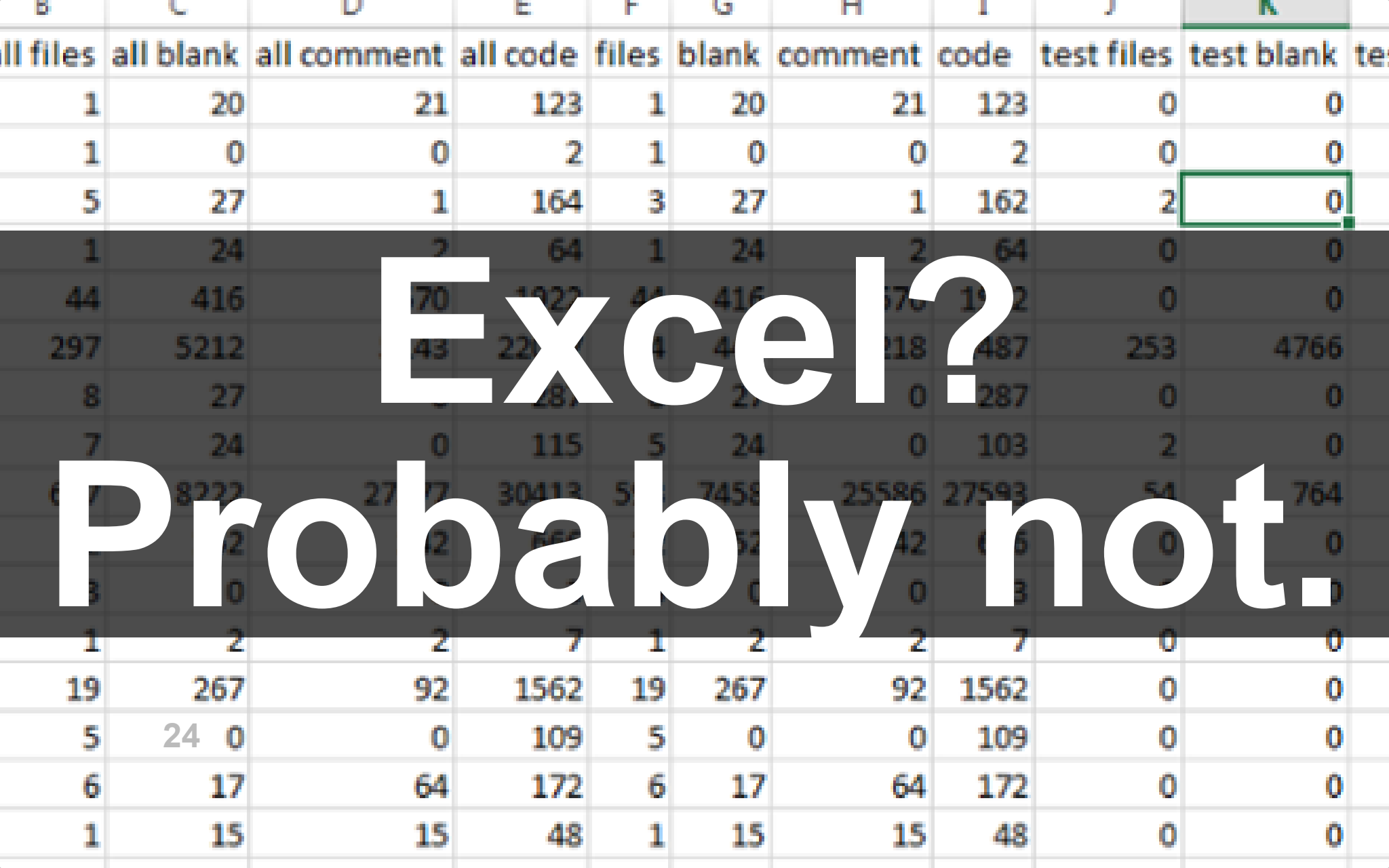
```
cloc --csv
     --quite
     --progress-rate=0
     --ignored=files.ignored
     --exclude-dir=test,build
     --read-lang-def=lang.defs
     --out=data.csv
     .
```

Language definitions

```
Gradle
    filter remove_matches ^\s*//
    filter remove_inline //.*$
    filter call_regexp_common C
    extension gradle
    3rd_gen_scale 4.10
```

Where are the pictures?

We have stats, let's plot them!

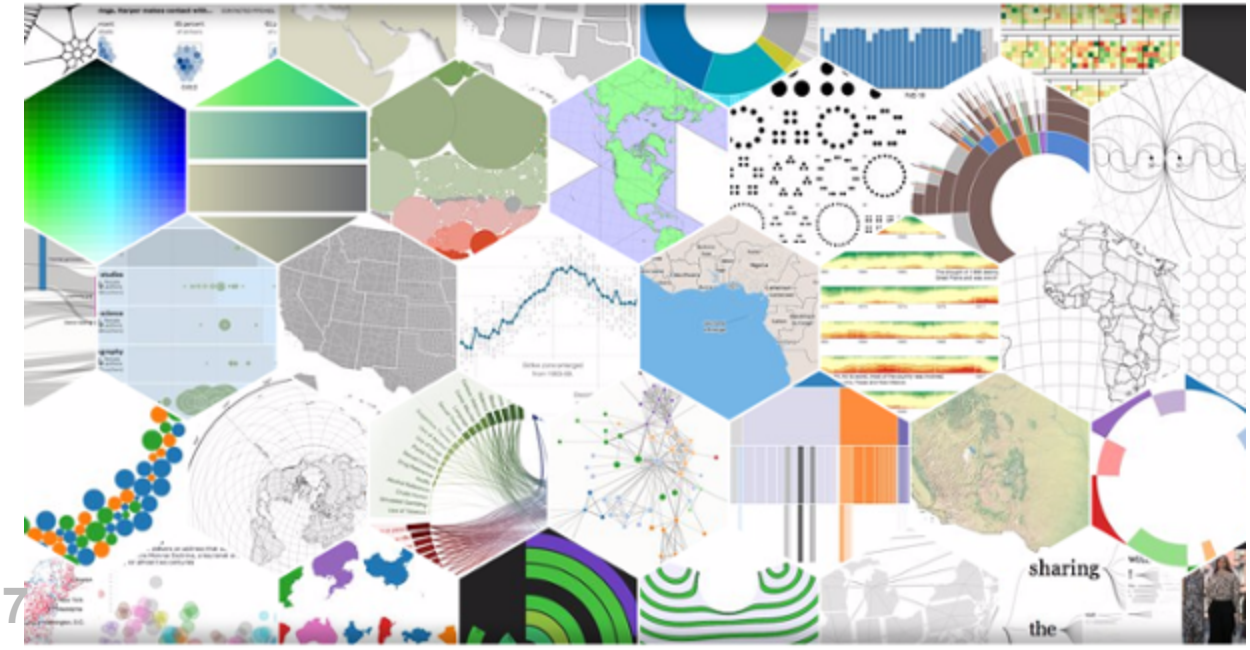

Excel?
Probably not.

Pie charts are boring!

Infographics!

d3.js

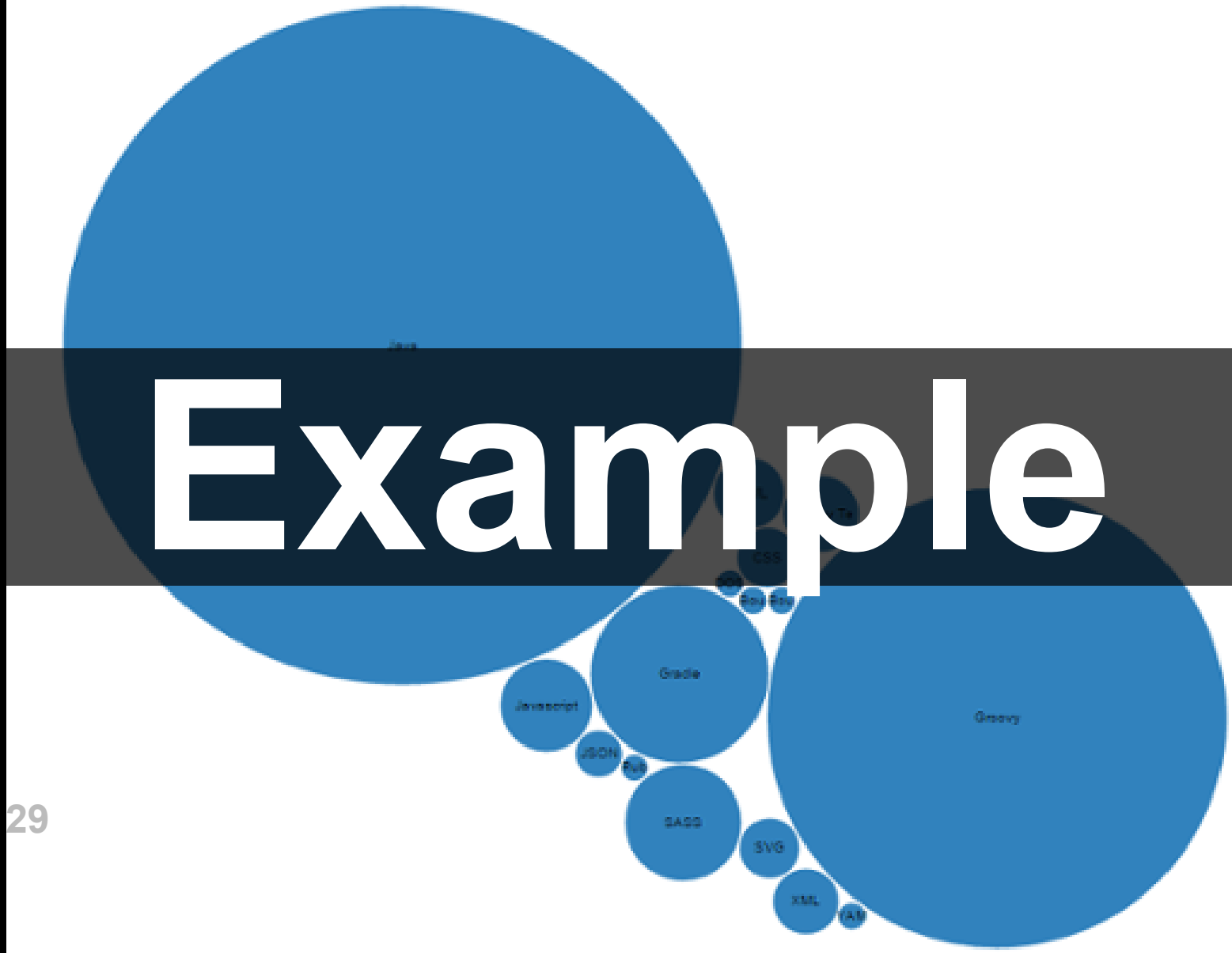
Data-Driven Documents



d3.js

- Javascript library for data visualizations
- Tons of examples
- Many libraries built on top of d3.js
- Several books

Example



Many alternatives

- vis.js
- raphael.js
- sigma.js
- many, many, many more

Tableau Public

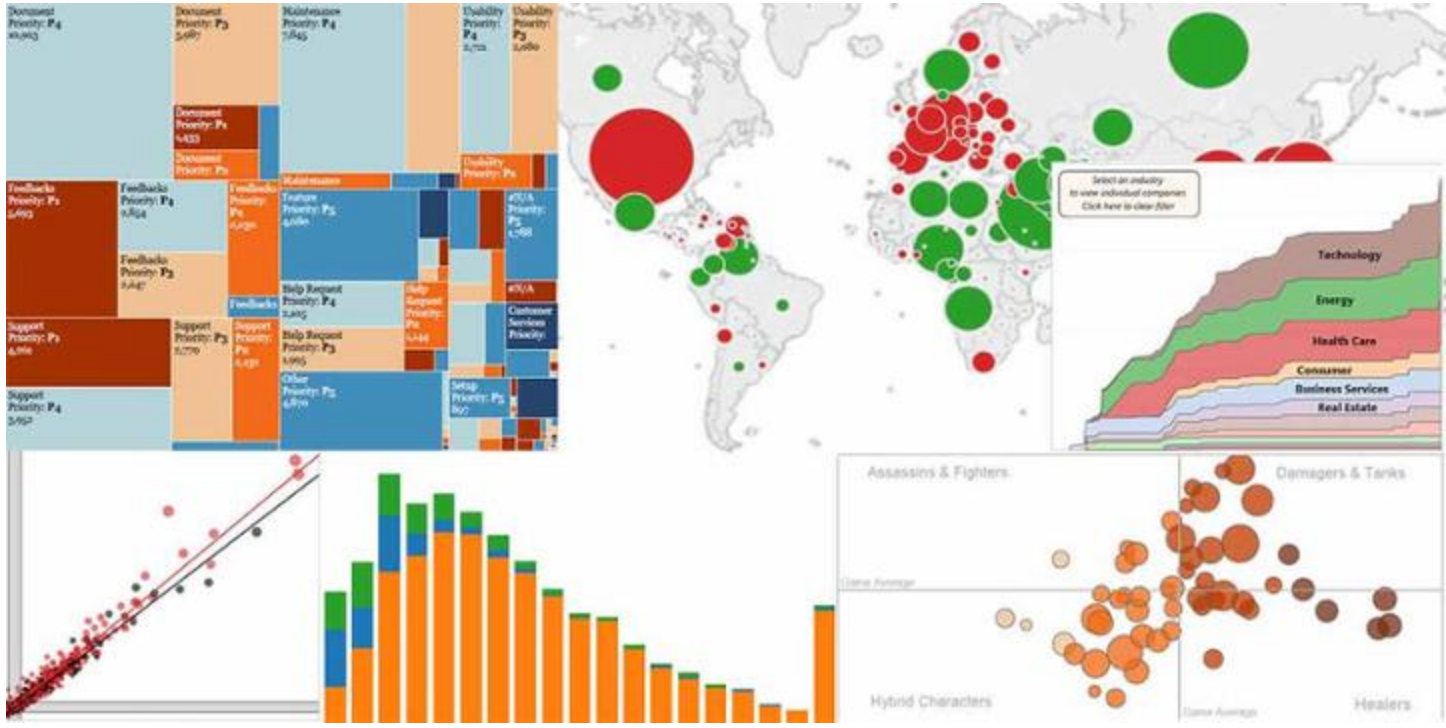


Tableau Public

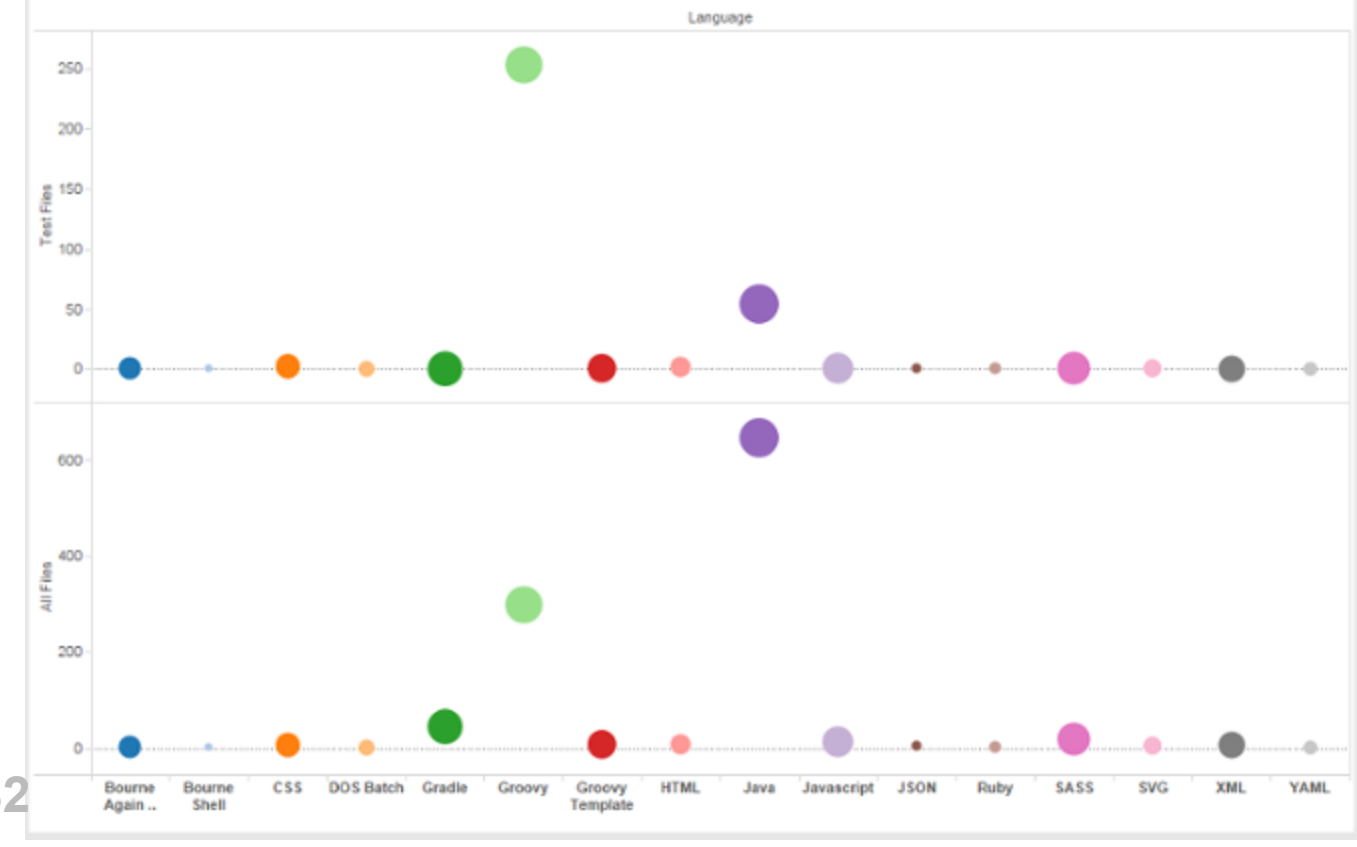
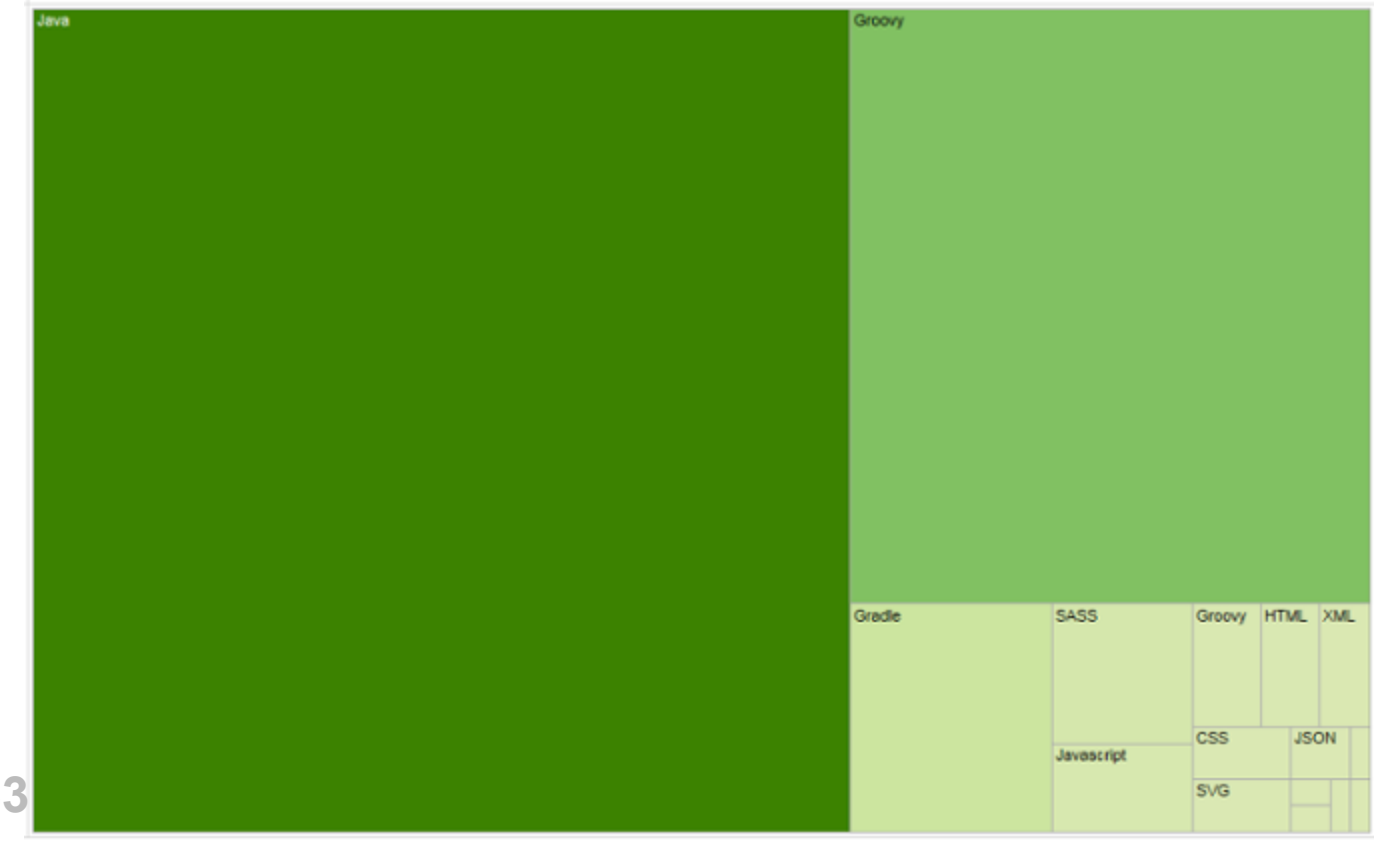


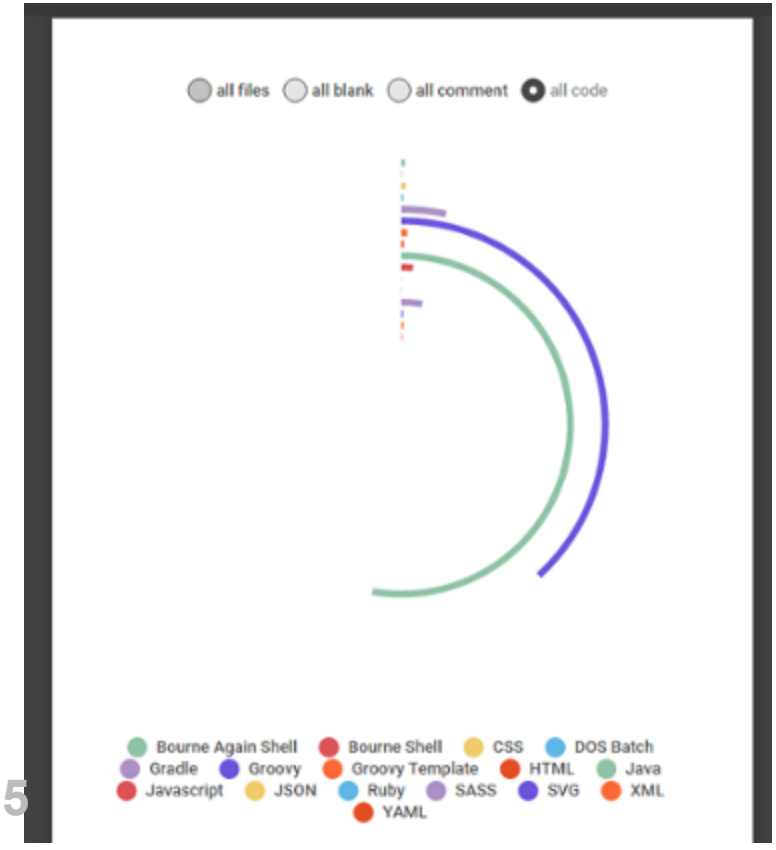
Tableau Public



infogr.am


infogr.am
EU countries by area
Austria
Belgium
Bulgaria
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Poland
Spain
Italy
Portugal
Romania
Sweden
United Kingdom
Latvia

infogr.am

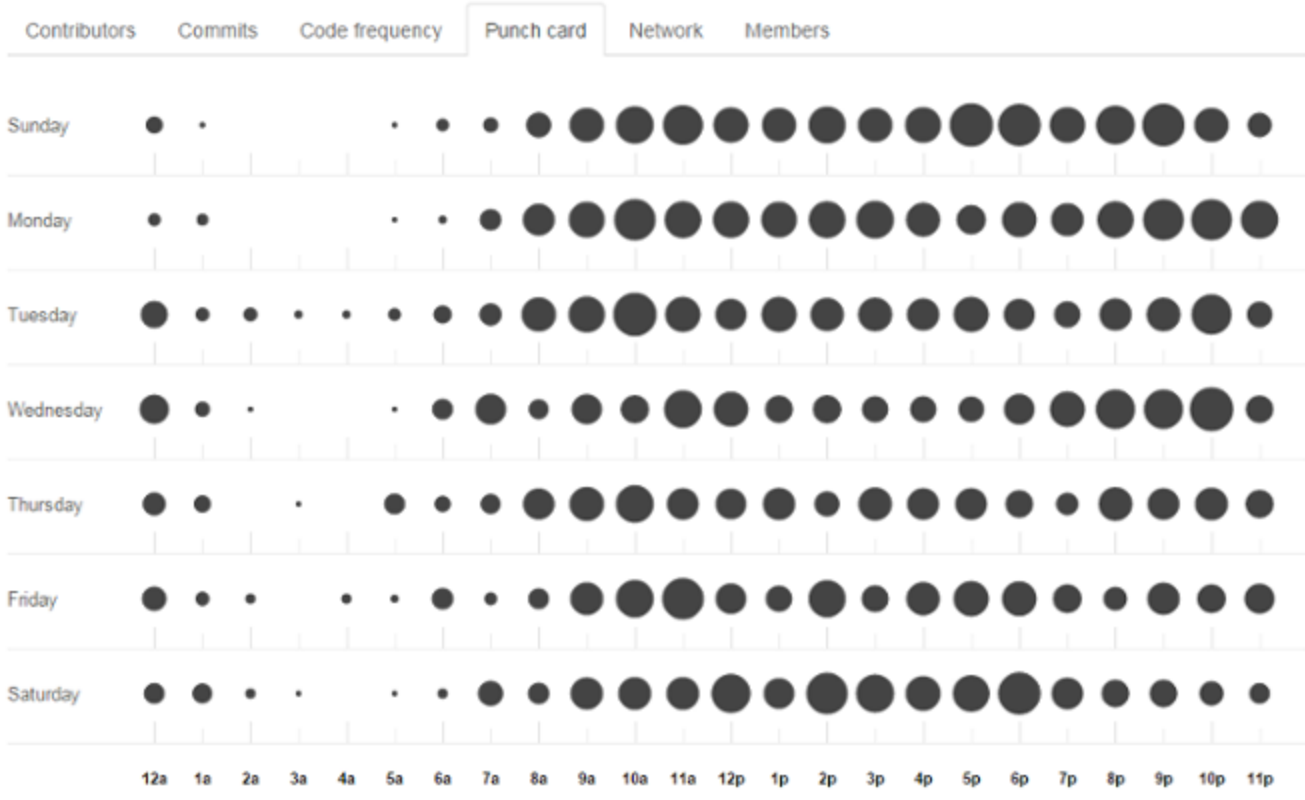

all files
all blank
all comment
all code
Bourne Again Shell
Bourne Shell
CSS
DOS Batch
Gradle
Groovy
Groovy Template
HTML
Java
Javascript
JSON
Ruby
SASS
SVG
XML
YAML

Temporal analysis

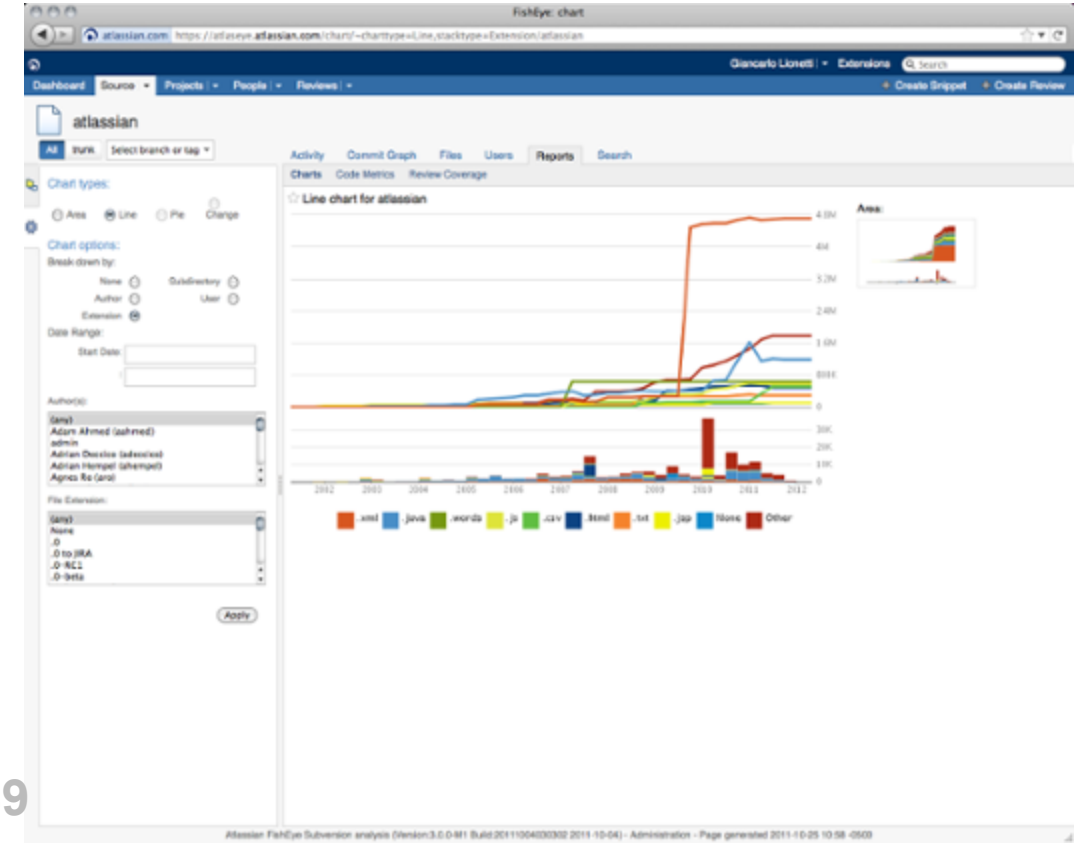
GitHub


Contributors
Commits
Code frequency
Punch card
Network
Members
Use ← and → to navigate
80
60
40
20
0
11/09
11/30
12/21
01/11
02/01
02/22
03/15
04/05
04/26
05/17
06/07
06/28
07/19
08/09
08/30
09/20
10/11
11/01

GitHub


Contributors
Commits
Code frequency
Punch card
Network
Members
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
12a 1a 2a 3a 4a 5a 6a 7a 8a 9a 10a 11a 12p 1p 2p 3p 4p 5p 6p 7p 8p 9p 10p 11p

FishEye


FishEye: chart
Dashboard
Source
Projects
People
Reviews
Create Snippet
Create Review
atlassian
Activity
Commit Graph
Files
Users
Reports
Search
Charts
Code Metrics
Review Coverage
Chart types:
Area
Line
Pie
Change
Chart options:
Break down by:
Date Range:
Start Date:
Author(s):
File Extension:
Apply
Line chart for atlassian
Area:

Gource

Software projects are displayed by Gource as an animated tree with the root directory of the project at its centre.

Directories appear as branches with files as leaves.

Developers can be seen working on the tree at the times they contributed to the project.

Launch gource

```
01. gource -s 0.1 -1280x720
02. gource --output-custom-log log1.txt
03. gource -s 0.1 -1280x720 log1.txt
```

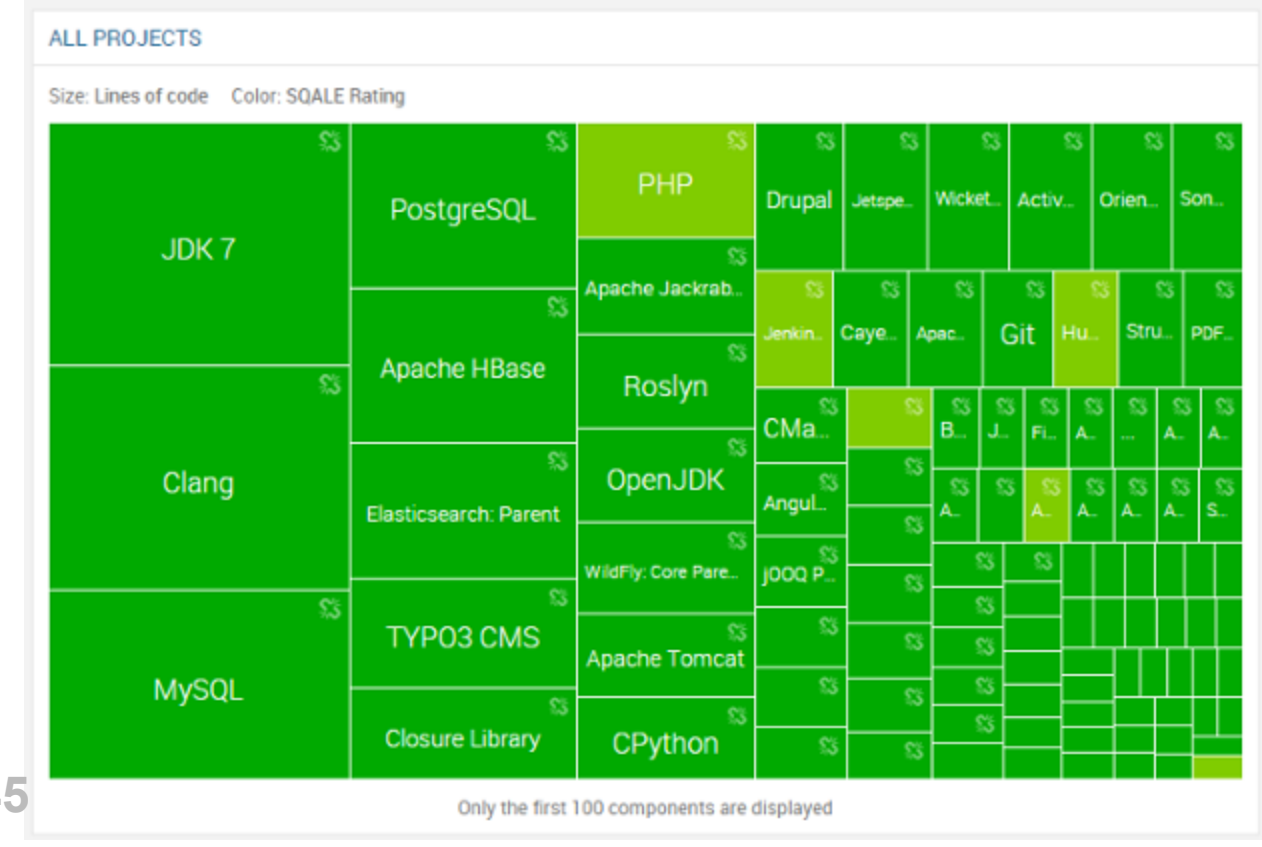
Log format

```
01. 1444125624|Luke Daley|M|/ratpack-core/.../WiretapPublisher.java
02. 1444125624|Luke Daley|M|/ratpack-core/.../YieldingPublisher.java
03. 1444306114|Stian Lindhom|M|/ratpack-manual/.../13-http.md
04. 1444312172|Andrey Antukh|M|/ratpack-core/.../WebSocketEngine.java
```

Demo

Metrics

SonarQube


ALL PROJECTS
Size: Lines of code
Color: SQALE Rating
JDK 7
PostgreSQL
PHP
Drupal
Jetspe...
Wicket...
Activ...
Orien...
Son...
Apache Jackrab...
Jenkin...
Caye...
Apac...
Git
Hu...
Stru...
PDF...
Apache HBase
Roslyn
CMa...
B...
J...
Fi...
A...
...
A...
A...
Clang
OpenJDK
Angul...
A...
A...
A...
A...
A...
S...
Elasticsearch: Parent
WildFly: Core Pare...
jOOQ P...
TYPO3 CMS
Apache Tomcat
MySQL
Closure Library
CPython
Only the first 100 components are displayed

Demo

Supported langauges I

- ABAP (comercial)
- Android
- C/C++ (comercial)
- C#
- COBOL (comercial)
- Flex

Supported langauges II

- Groovy
- Java
- JavaScript
- Objective-C (comercial)
- PHP

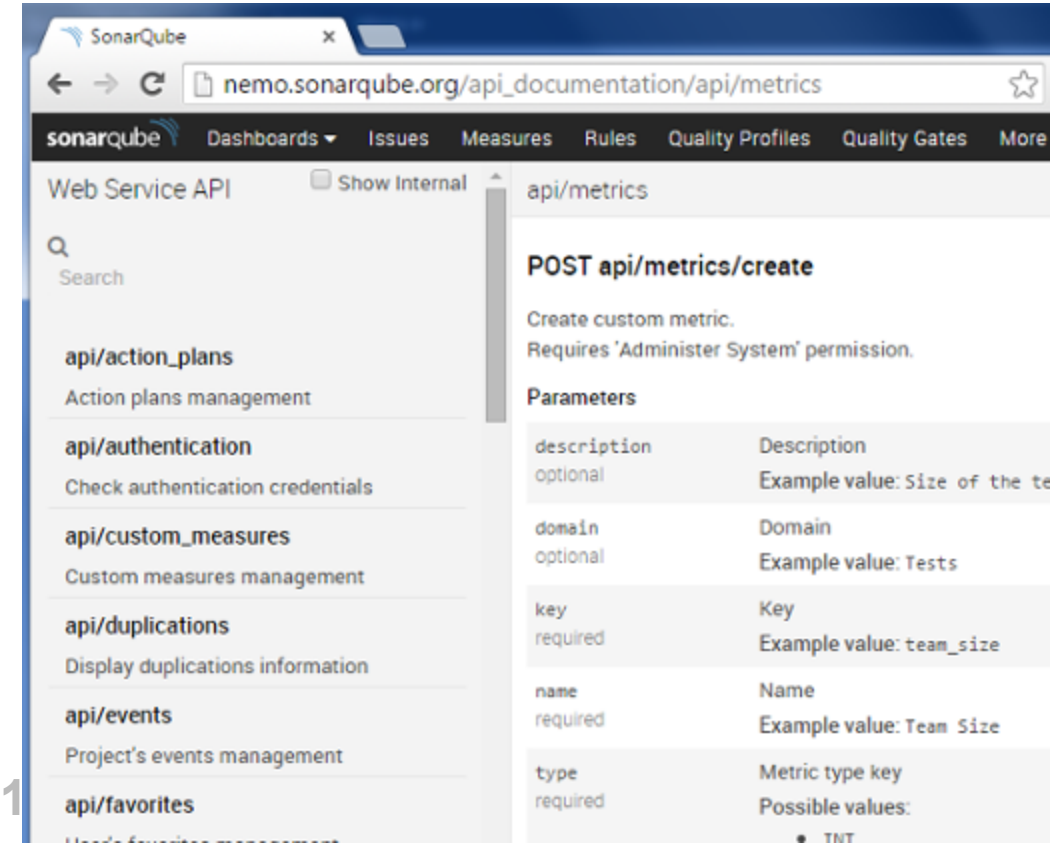
Supported langauges II

- PL/I (comercial)
- PL/SQL (comercial)
- Python
- RPG (comercial)
- Swift (comercial)

Supported langauges II

- VB.NET (comercial)
- Visual Basic 6 (comercial)
- Web (HTML, JSP/JSF)
- XML

Extracting data

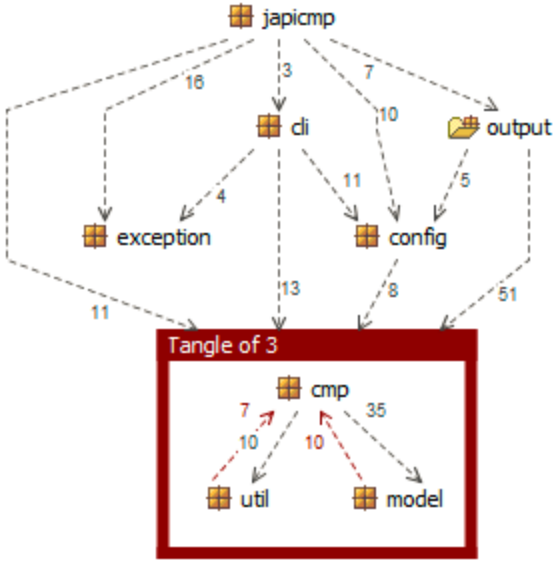


Structure

Saikuro

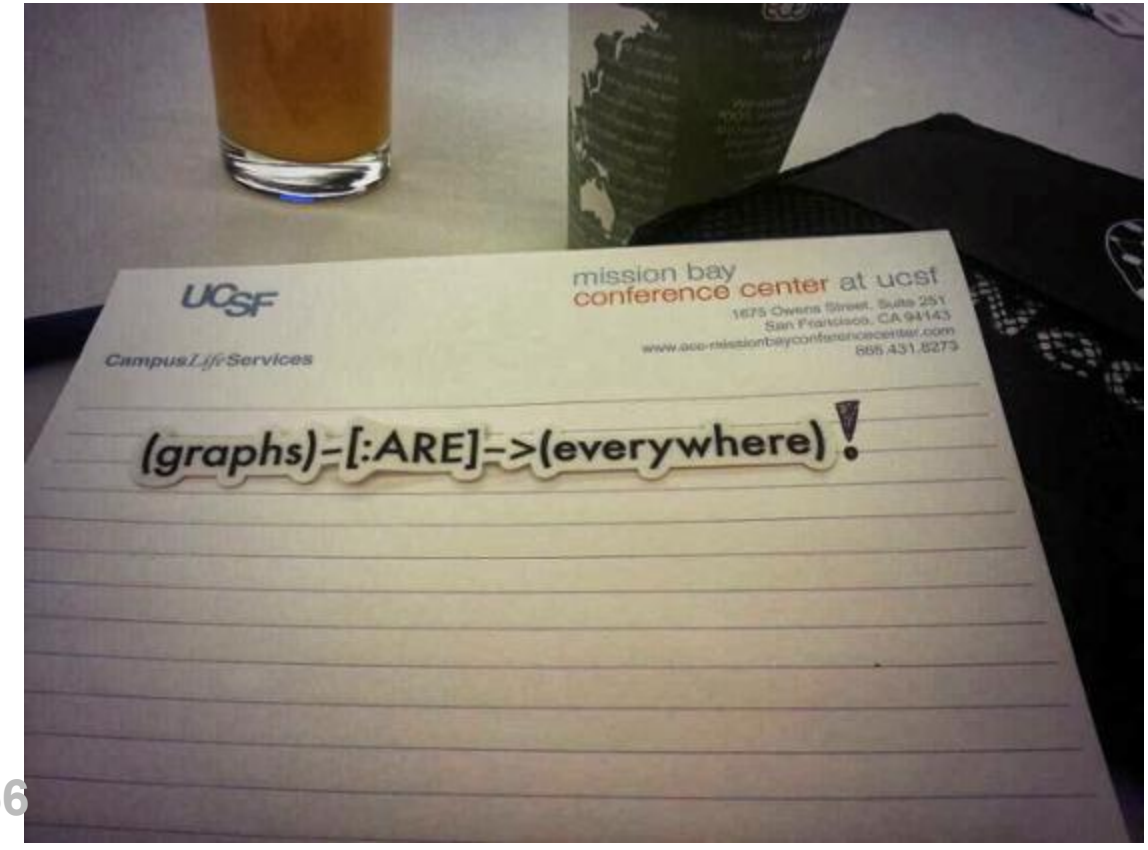
- https://github.com/ThoughtWorksStudios/saikuro_treemap
- https://www.youtube.com/watch?v=iiIytERhV9o

Structure101



Stan4j

Graphs are everywhere!



AST


statement sequence
while
return
condition
body
compare op: ≠
variable name: a
variable name: b
constant value: 0
branch
condition
if-body
else-body
compare op: >
assign
assign
variable name: a
variable name: b
variable name: a
bin op op: −
variable name: b
bin op op: −
variable name: a
variable name: b
variable name: b
variable name: a

jQAssistant

jQAssistant is a QA tool which allows the definition and validation of project specific rules on a structural level.

It is built upon the graph database Neo4j and can easily be plugged into the build process to automate detection of constraint violations and generate reports about user defined concepts and metrics.

Query your code

```
01. MATCH
02.   (class:Class)-[:DECLARES]->(method:Method)
03. RETURN
04.   class.fqn, count(method) as Methods
05. ORDER BY
06.   Methods DESC
07. LIMIT 20
```

Demo

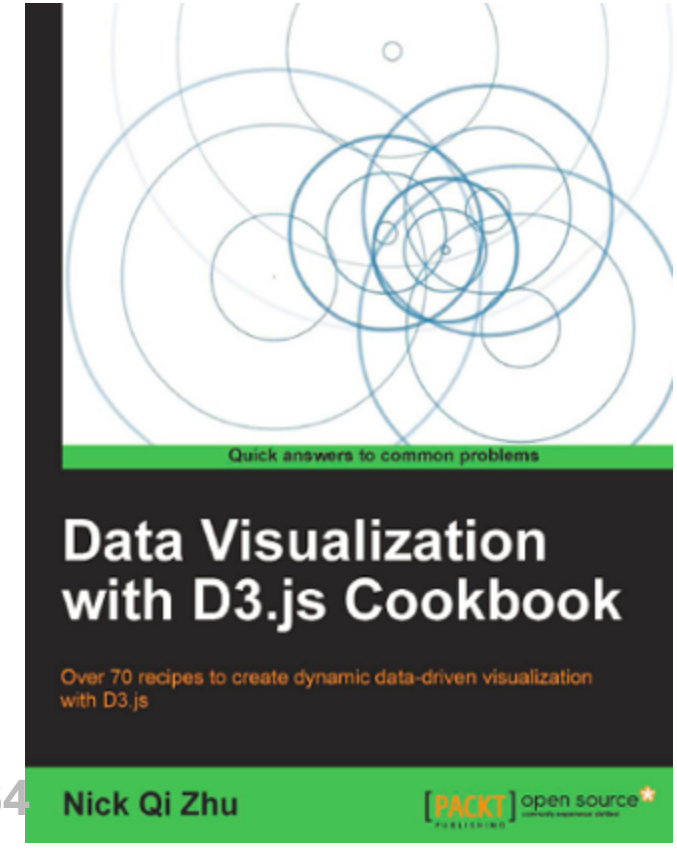
That's it!

Conclusion

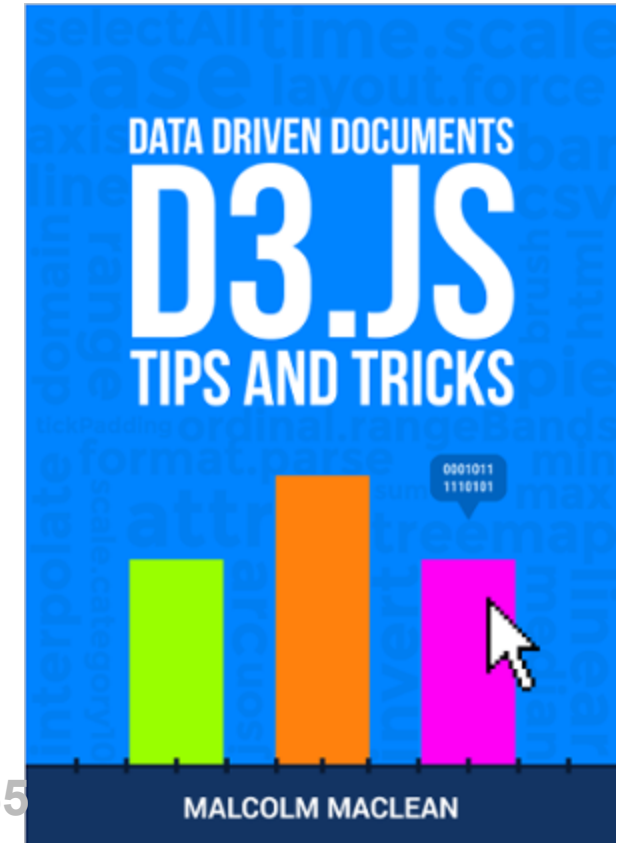
- Extract data from your code!
- Visualize it!
- Search for new facts and knowledge!

Reading material

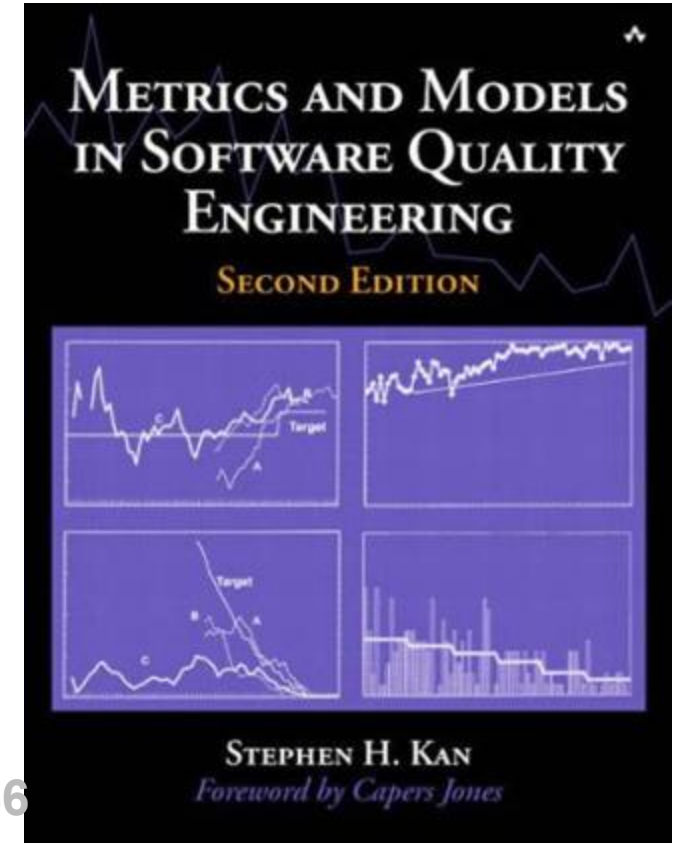
D3 Cookbook


Quick answers to common problems
Data Visualization with D3.js Cookbook
Over 70 recipes to create dynamic data-driven visualization with D3.js
Nick Qi Zhu
[PACKT] open source

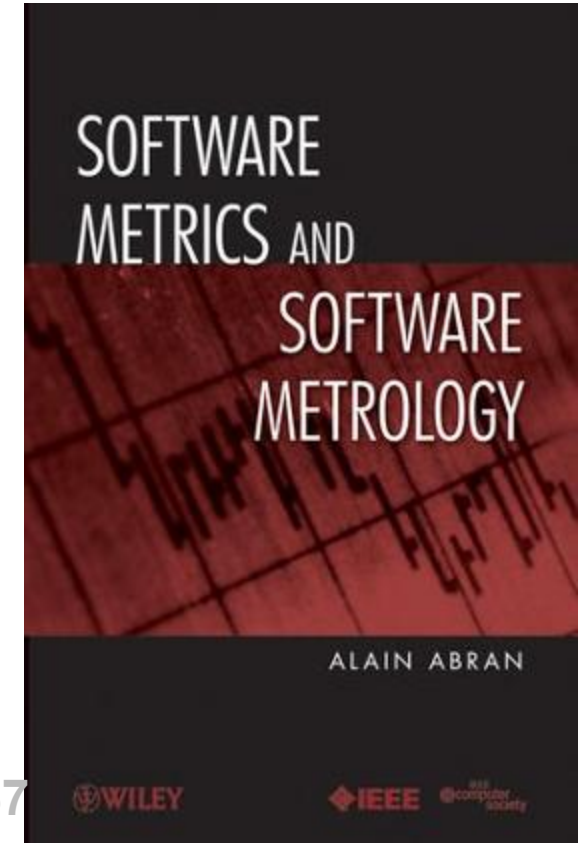
D3 Tips & Tricks



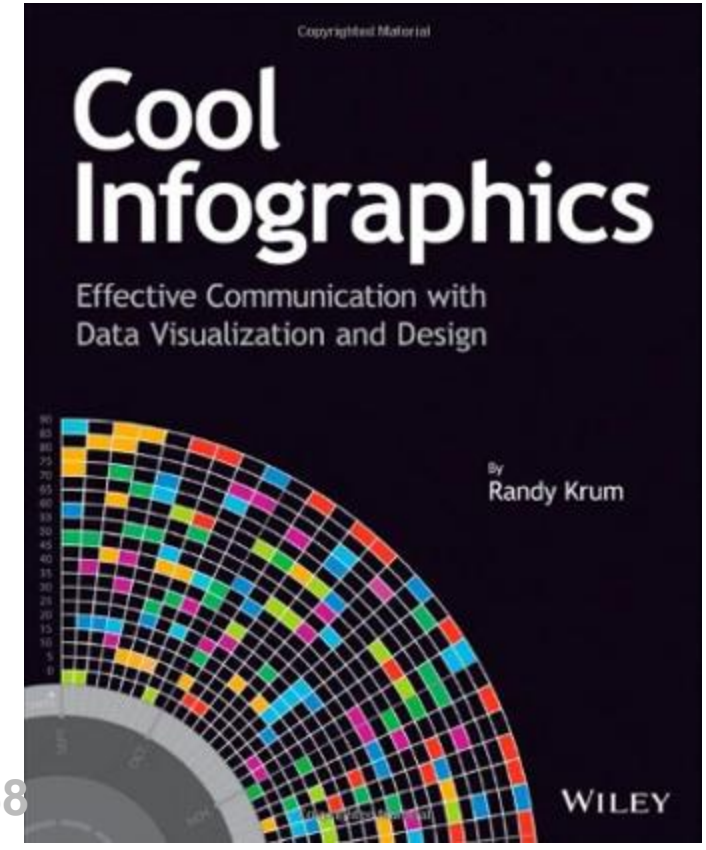
Metrics and models in SQE



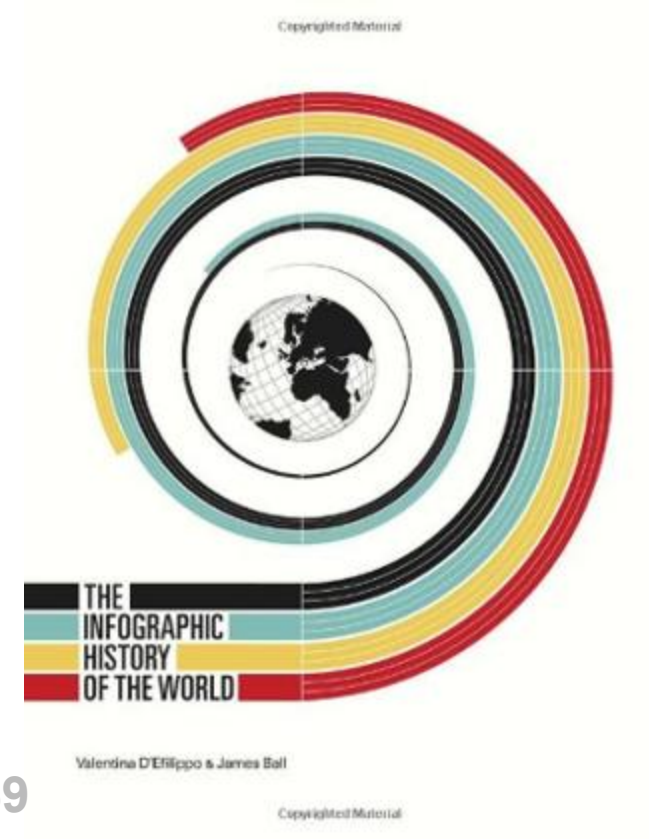
Software metrics and metrology



Cool infographics

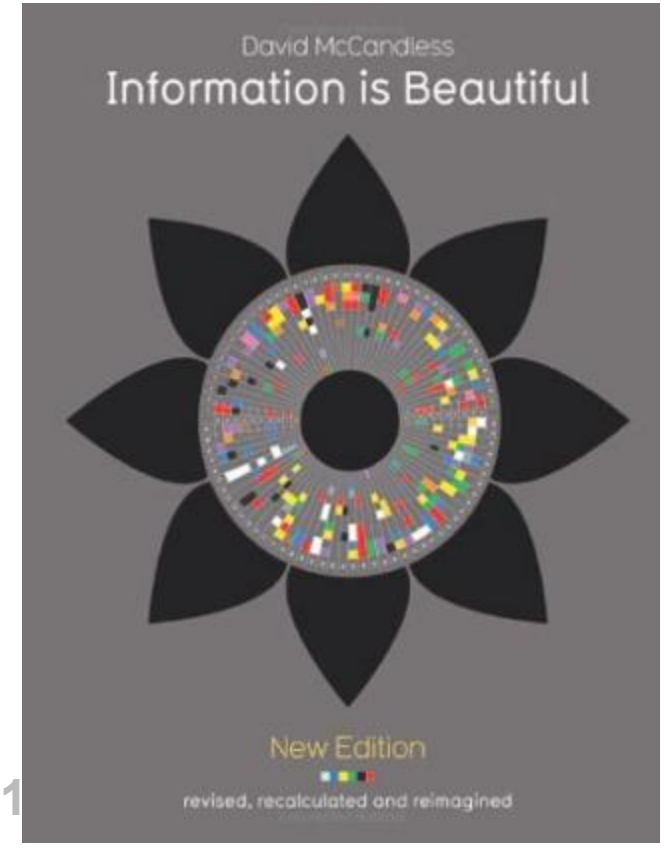


The infographic history of the world

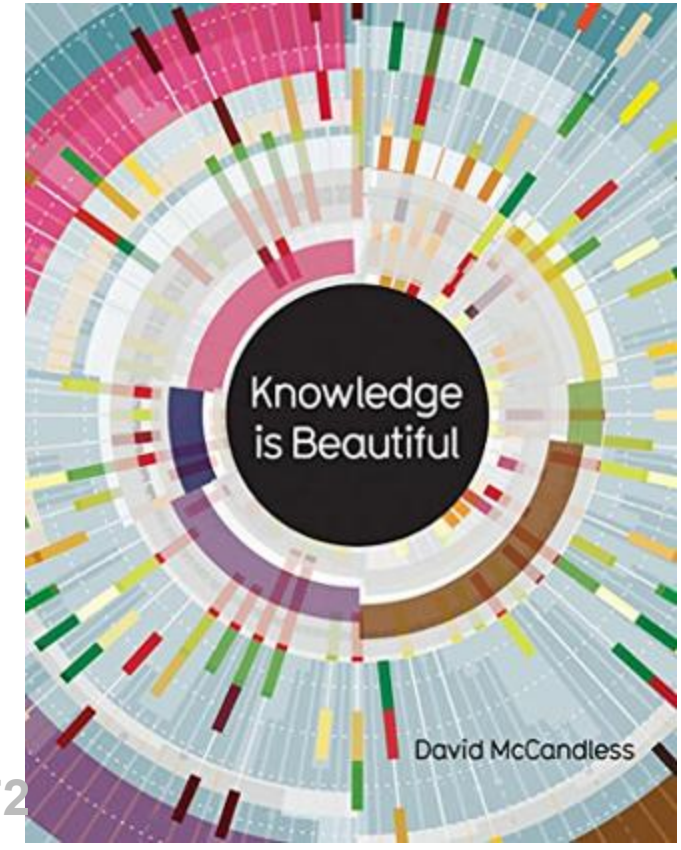


Infographica

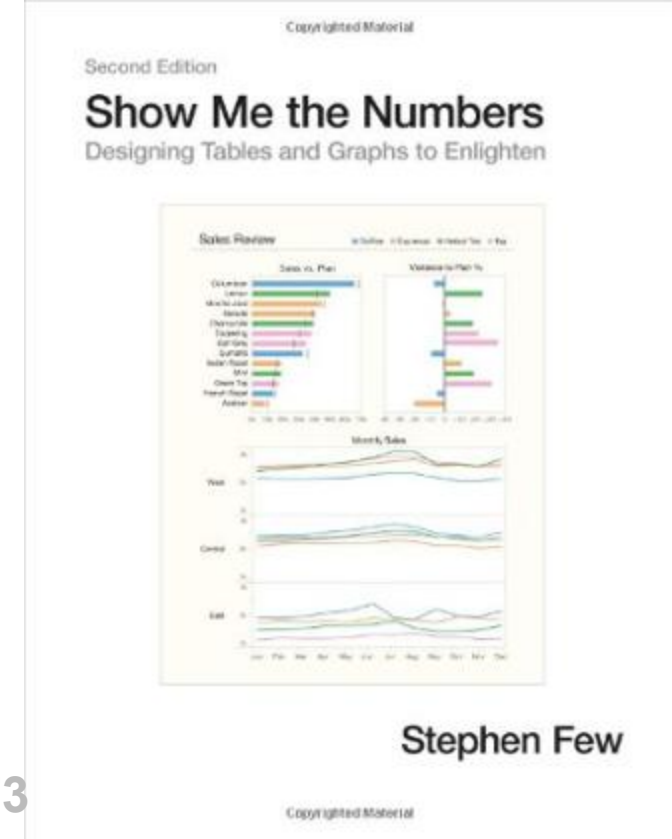
Information is beautiful



Knowledge is beautiful



Show me numbers



Links: images

- http://abstrusegoose.com/432
- http://camarenaphoto.tumblr.com/post/112238079516/its-life-jim-but-not-as-we-know-it-spock
- http://technology.ie/big-data-looks-like/
- http://www.informationisbeautiful.net/visualizations/million-lines-of-code/
- http://githut.info/
- http://emmanueloga.com/2013/10/07/Graphs-are-Everywhere-An-overview-of-GraphConnect-San-Francisco-2013.html

Links: tools

- https://github.com/AlDanial/cloc
- http://gource.io/
- http://d3js.org/
- http://visjs.org/

Links: tools

- http://www.sonarqube.org/
- https://www.atlassian.com/pt/software/fisheye/overview
- http://jqassistant.org/
- http://neo4j.com/

Links: tools

- https://github.com/ThoughtWorksStudios/saikuro_treemap
- https://www.youtube.com/watch?v=iiIytERhV9o

Danke schön!

Thank you!

Questions?