



Community Health Analytics: Building a Shared Software and Data Infrastructure: 2019 Update through the lens of Augur



CHAOS





How to track sustainable projects...

How can we know if this open source project is likely to be around in 10 years if we base a product on it?

How can we know if this open source project is ready to be used by another project?

Is there a diverse community of active contributors engaged in the project?

Are there licensing risks in using this open source project?

What is the health of the other projects that this project depends on?

... need a common understanding of open source project health.

CHAOSS Mission



Establish implementation-agnostic metrics for measuring community activity, contributions, and health.

Produce integrated, open source software for analyzing software development in terms of these metrics.

Working in an Open Community...



University of Missouri



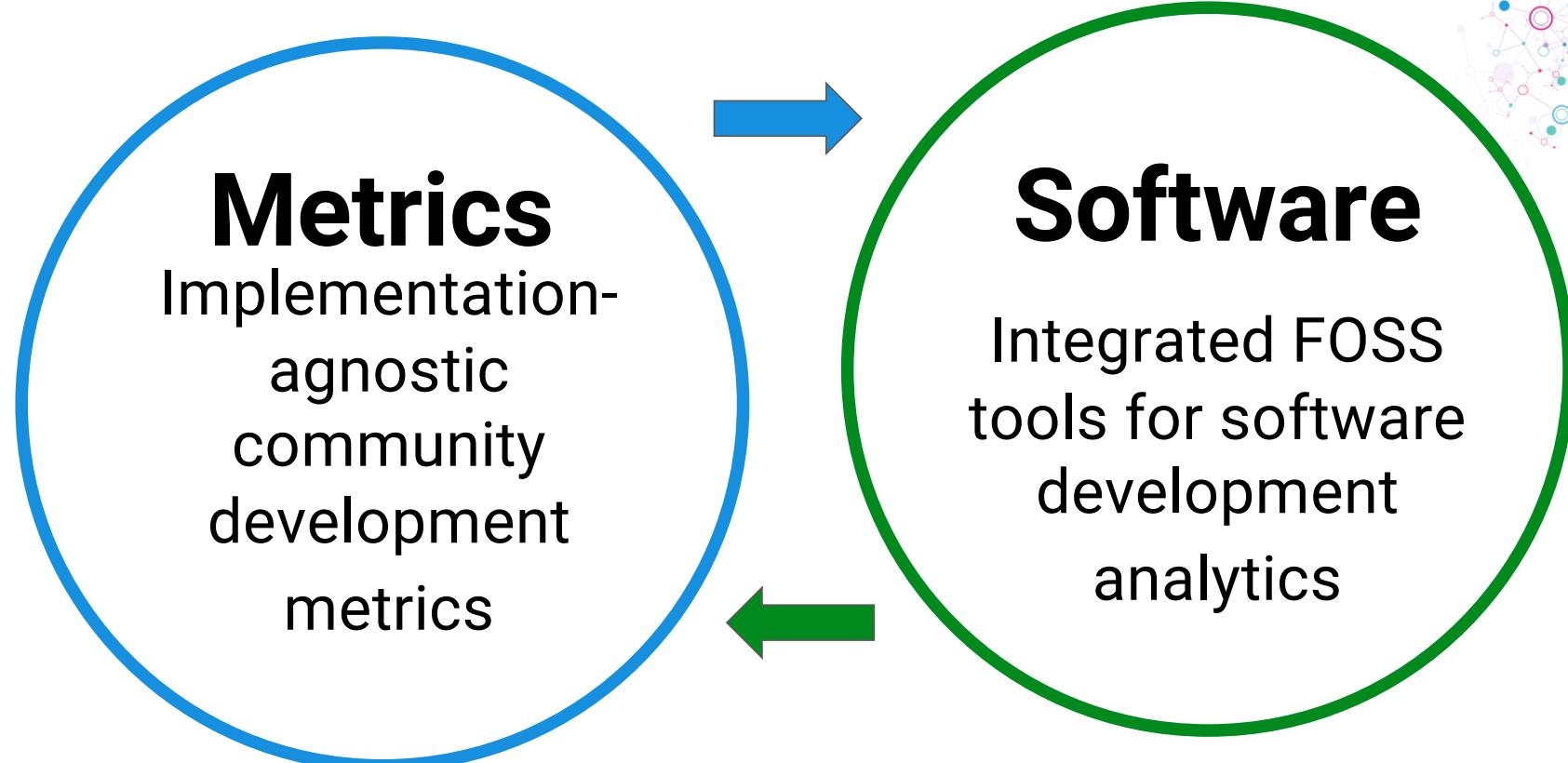
University
of Victoria



JONO BACON CONSULTING

CHAOS

Structure: Focus Around Interests



Working Groups



Diversity-Inclusion	Evolution
Risk	Value
Common	

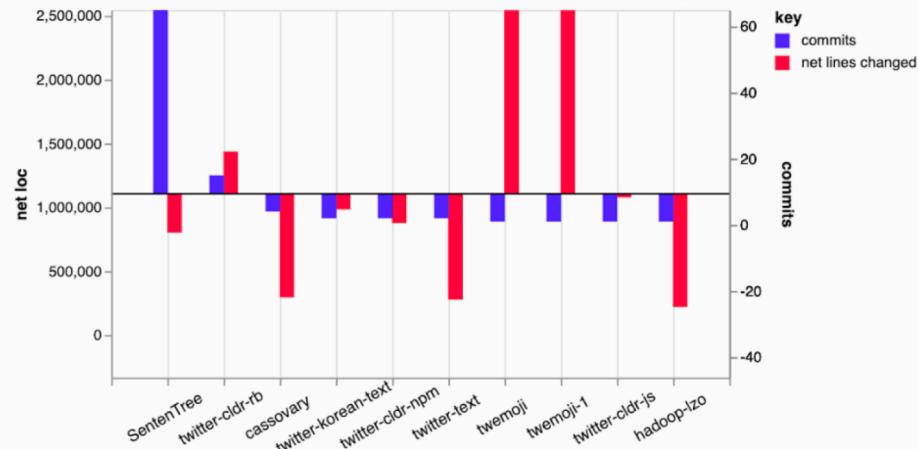
wiki.linuxfoundation.org/chaoss/metrics

Evolution

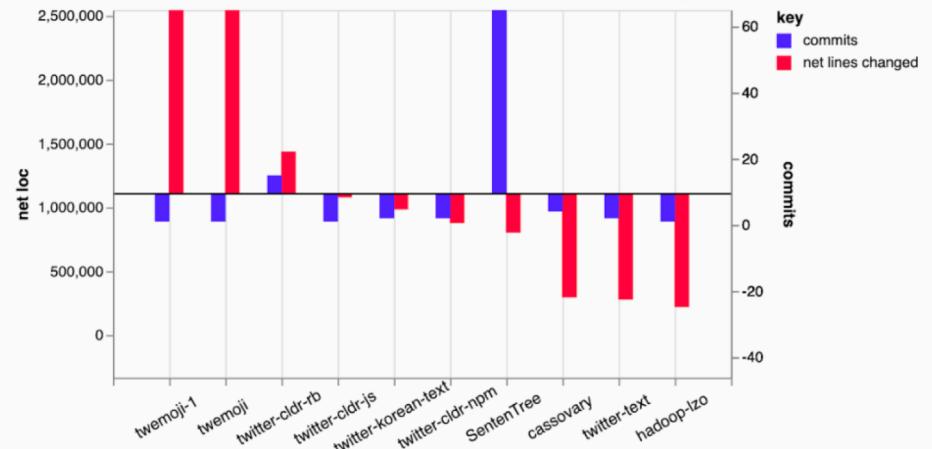


Project Overview: Twitter

Top Repos in 2018 by Commits with Baseline Averages - Sorted



Top Repos in 2018 by Net LoC with Baseline Averages - Sorted



*The black "baseline" represents the averages of both LoC and commits across all repositories within the selected repository's overlying Facade organization during the calendar year shown. Wherever this bar stretches to shows how far above or below the raw value of the statistic is from the regular average.

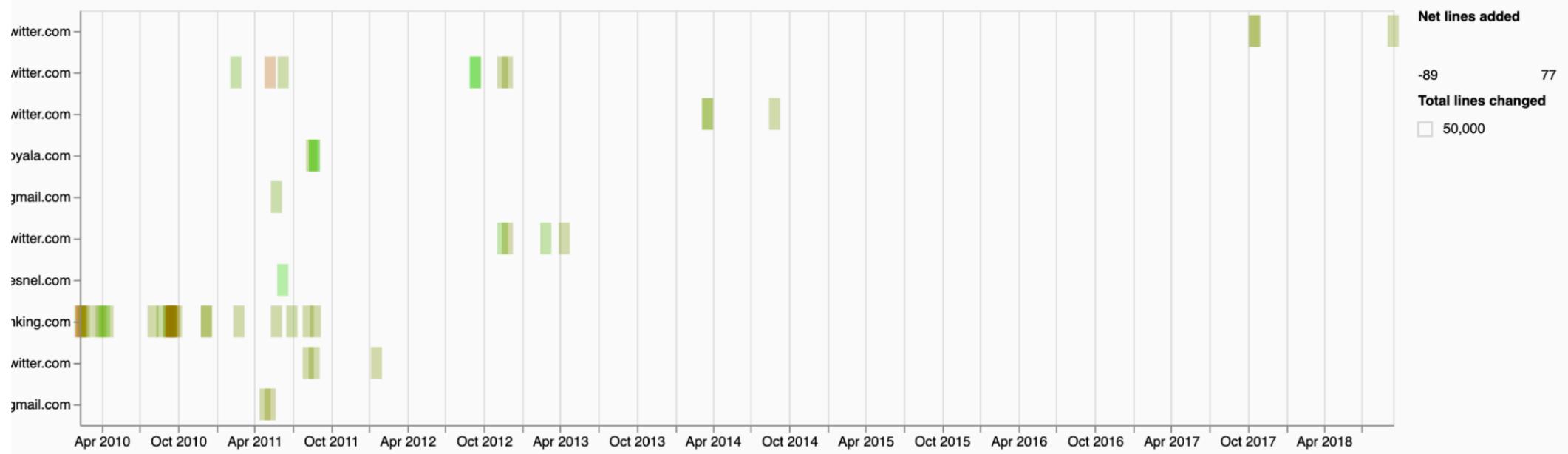
*The black "baseline" represents the averages of both LoC and commits across all repositories within the selected repository's overlying Facade organization during the calendar year shown. Wherever this bar stretches to shows how far above or below the raw value of the statistic is from the regular average.

Evolution

github.com/twitter/thrift_client.git

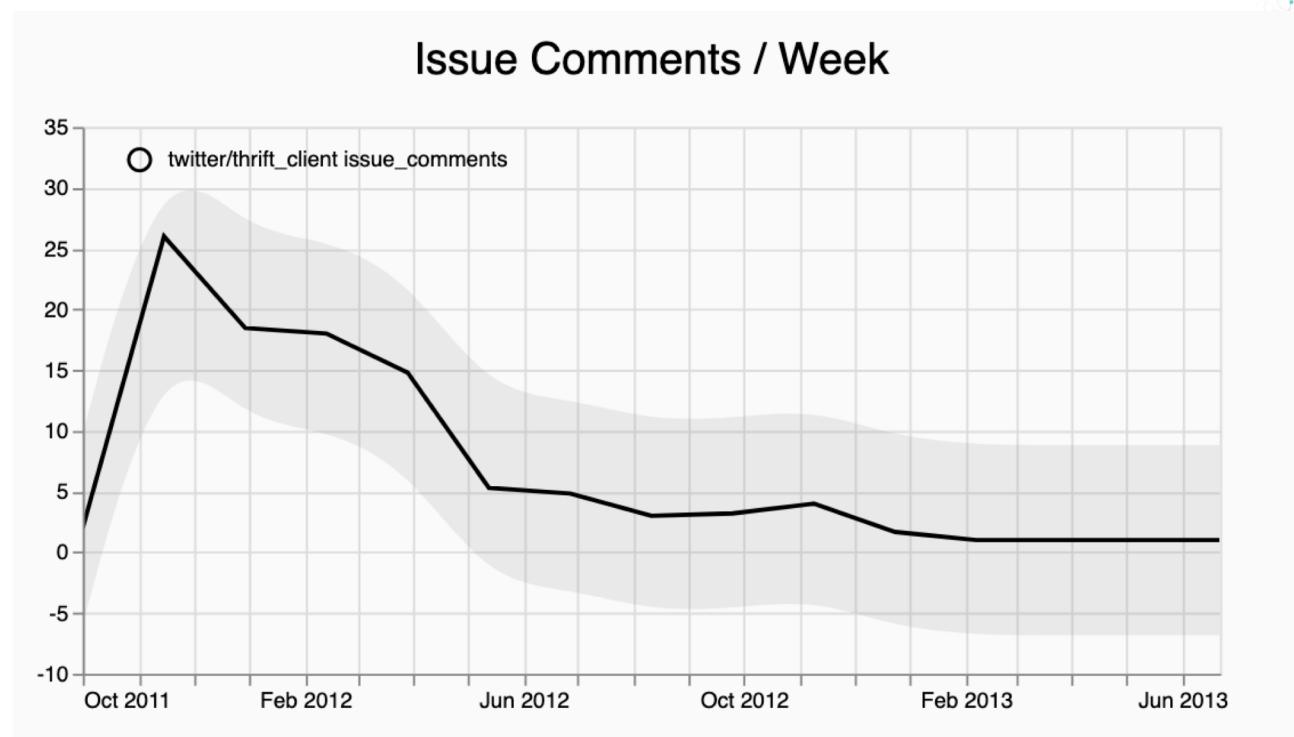


Lines of code added by the top 10 authors visualized



Evolution

github.com/twitter/thrift_client.git



Design Principles



1. Transparency
2. Comparison
3. Easy Modification and Access

Augur API Structure

/repo-groups

Returns a list of all repo groups (IDs and shortnames)

/repo-groups/<repo-group-id>

Returns overview/summary data for repo group w/ given repo group ID

/repo-groups/<repo-group-id>/metrics

Returns a list of all metrics for a repo group w/ given repo group ID

/repo-groups/<repo-group-id>/[metric]

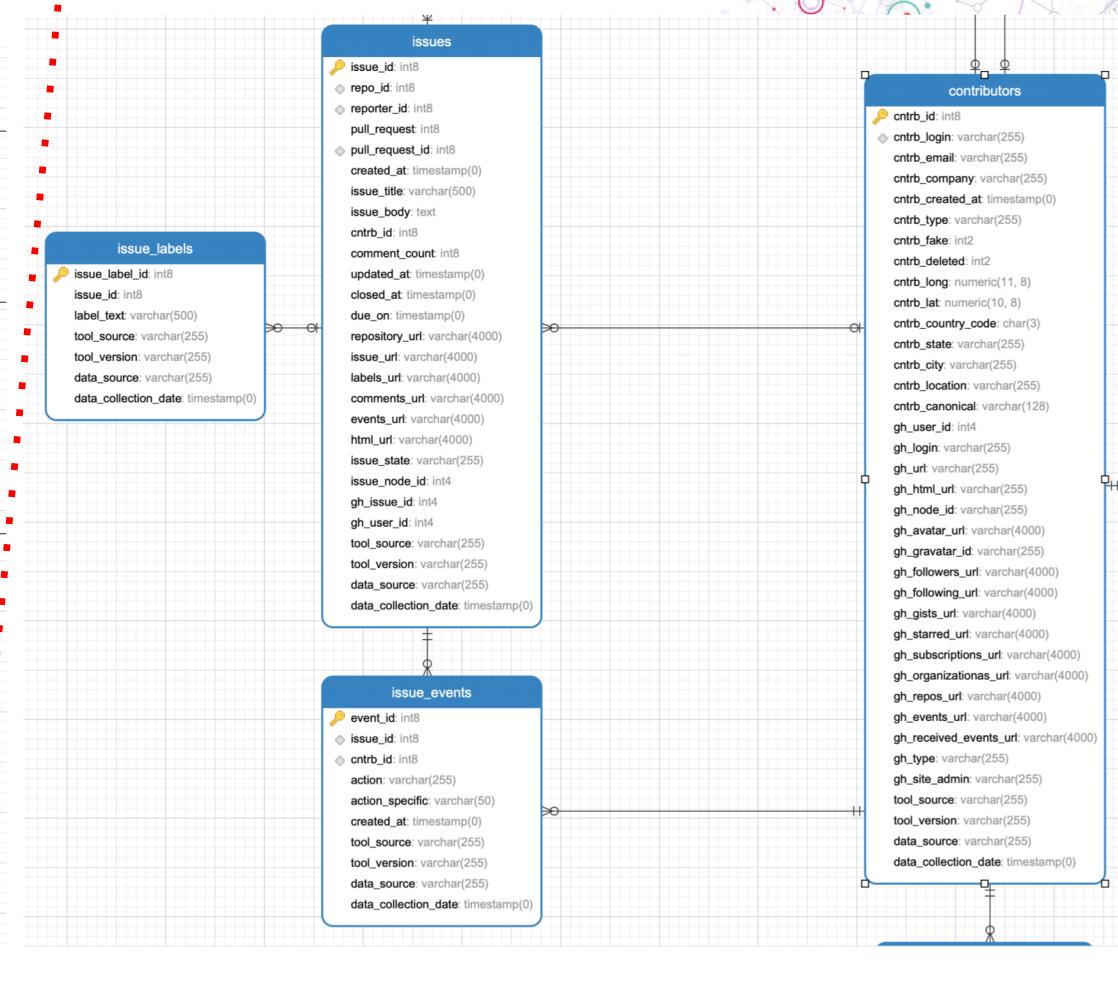
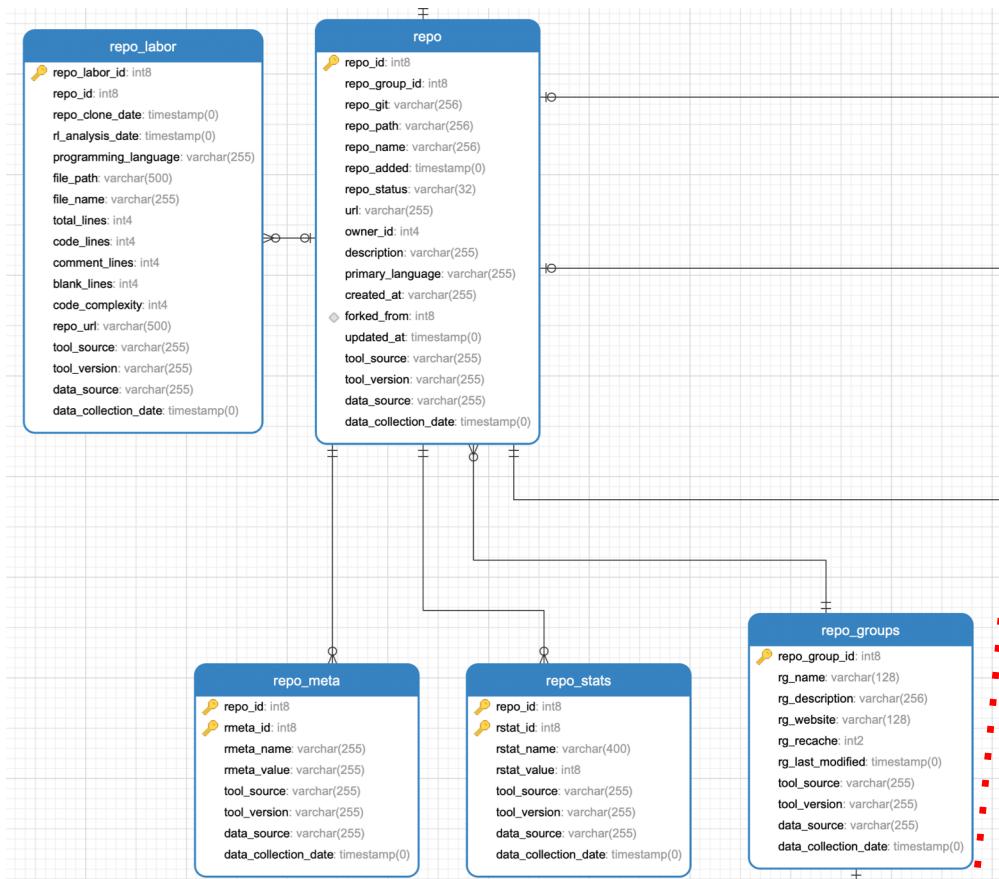
Returns output of metric for repo group w/ given repo group ID

Examples:

/repo-groups/:zephyr/code-changes

/repo-groups/:zephyr/code-changes-lines -

Integrated Data Model



CHAOS



Diversity & Inclusion



Goal: The CHAOSS D&I Working Group establishes and broadly communicates an ethical, peer-validated, research-informed, set of standards and best practices for measuring diversity and inclusion in open source projects, and ensures that CHAOSS project software can implement D&I metrics according to identified goals for inclusion with validation from key industry partnerships in open source. See [2019 objectives](#) for more detail.

Objective #4: Ethical Guidelines

The D&I Working Group has explored key questions of ethics and established ethical guidelines which everyone has to follow before they are allowed to use the brand "CHAOSS D&I Metrics."

- KR 1 - We have crowdsourced and prioritized ethical considerations around data collection, usage, storage, analysis, and reporting.
- KR 2 - We have documented ethical best practices for D&I metrics and embedded them in our project documentation.
- KR 3 - We evaluated solutions for how we could enforce ethical best practices for anyone who wants to use CHAOSS D&I Metrics.
- KR 4 - We have published a blog post about how we developed the ethical guidelines and encourages adoption.

Risk and Value Working Groups Focus Areas

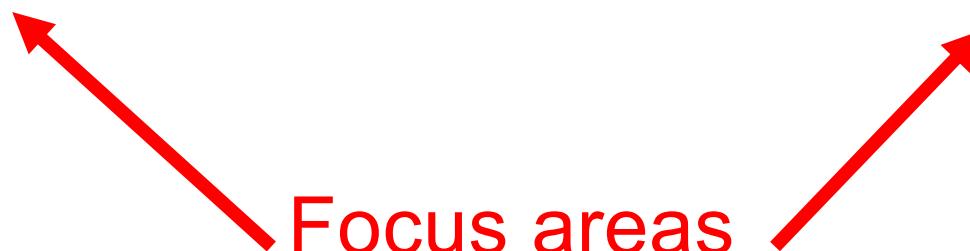


Risk

- Security
- Quality
- Licensing
- Business Risk
- Transparency

Value

- Labor Investment
- Innovation Value
- Downstream Value
- Ecosystem Value
- Living Wage

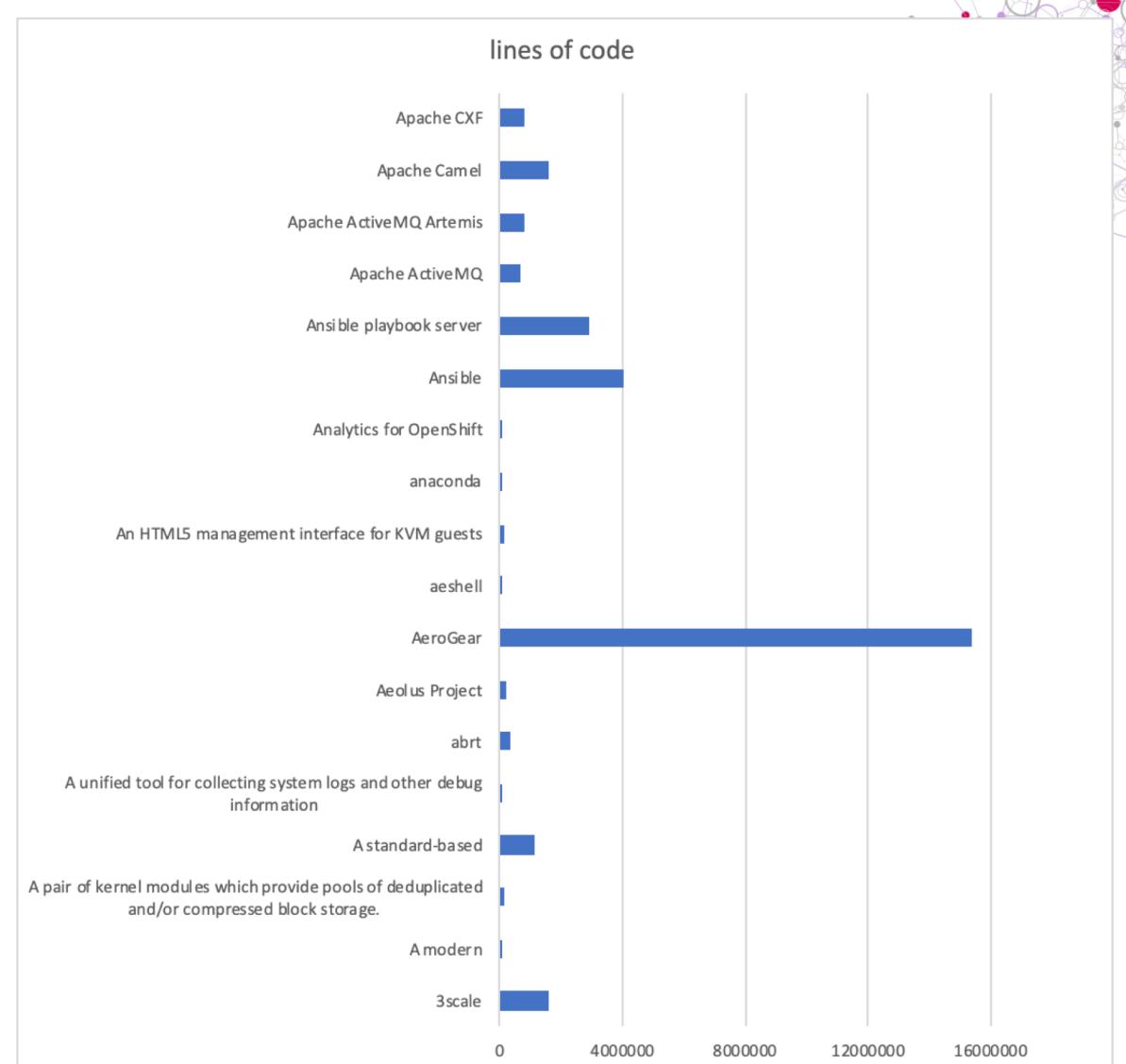


Focus areas

Value



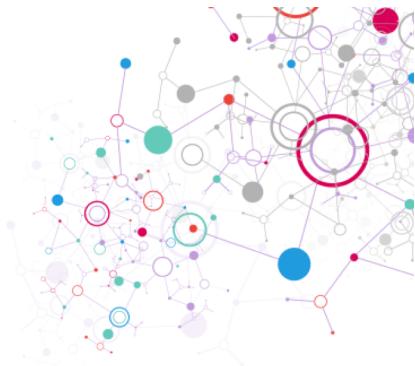
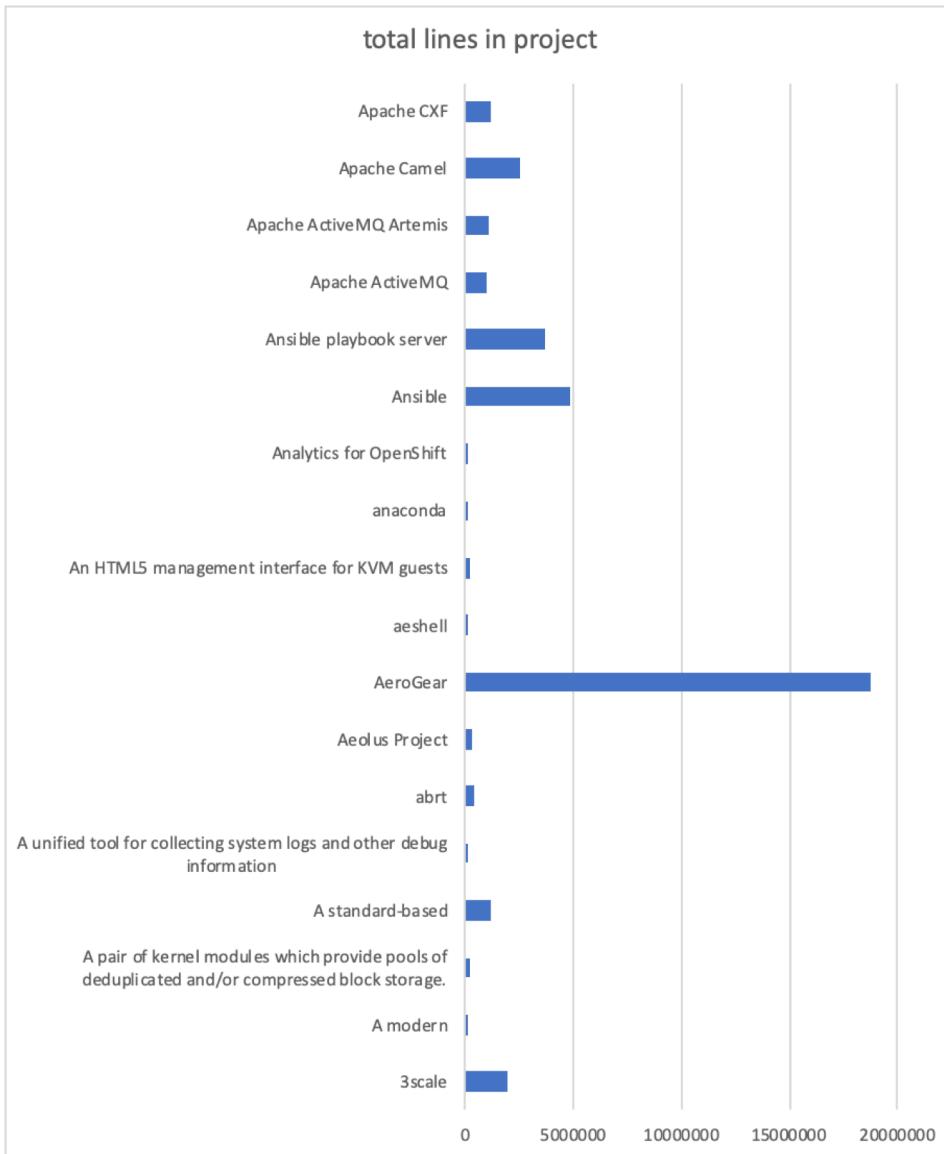
CHAOS



Value



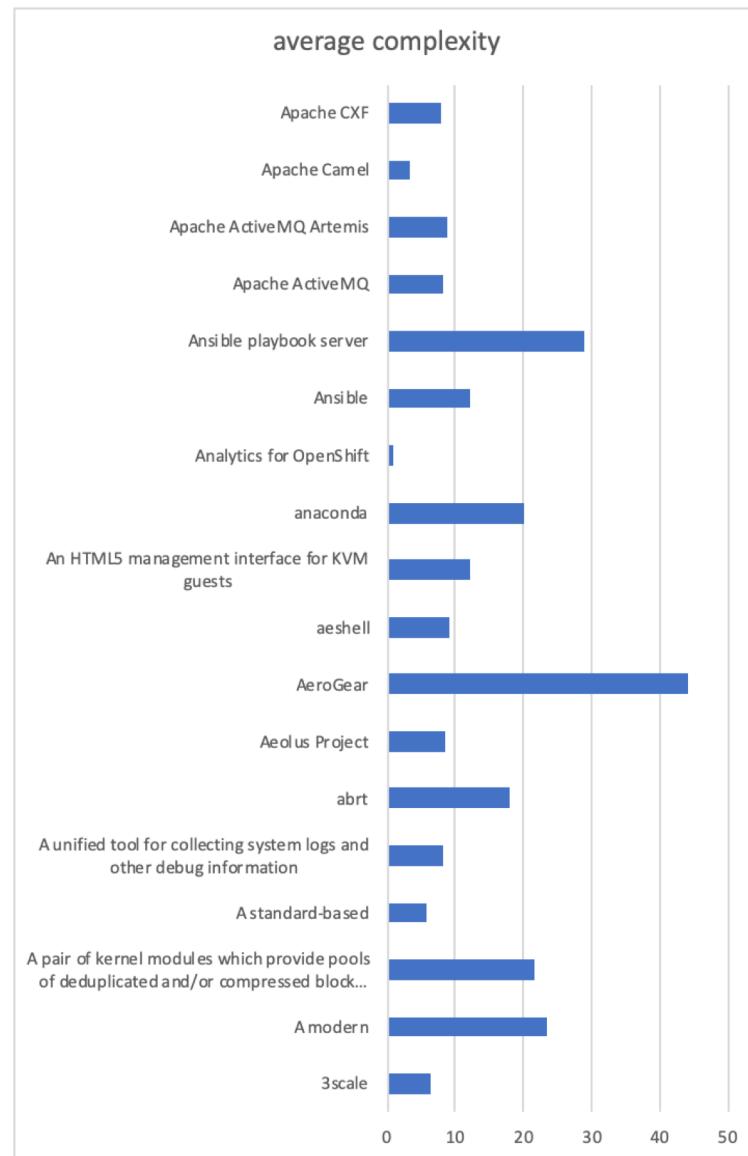
CHAOS



Value



CHAOS

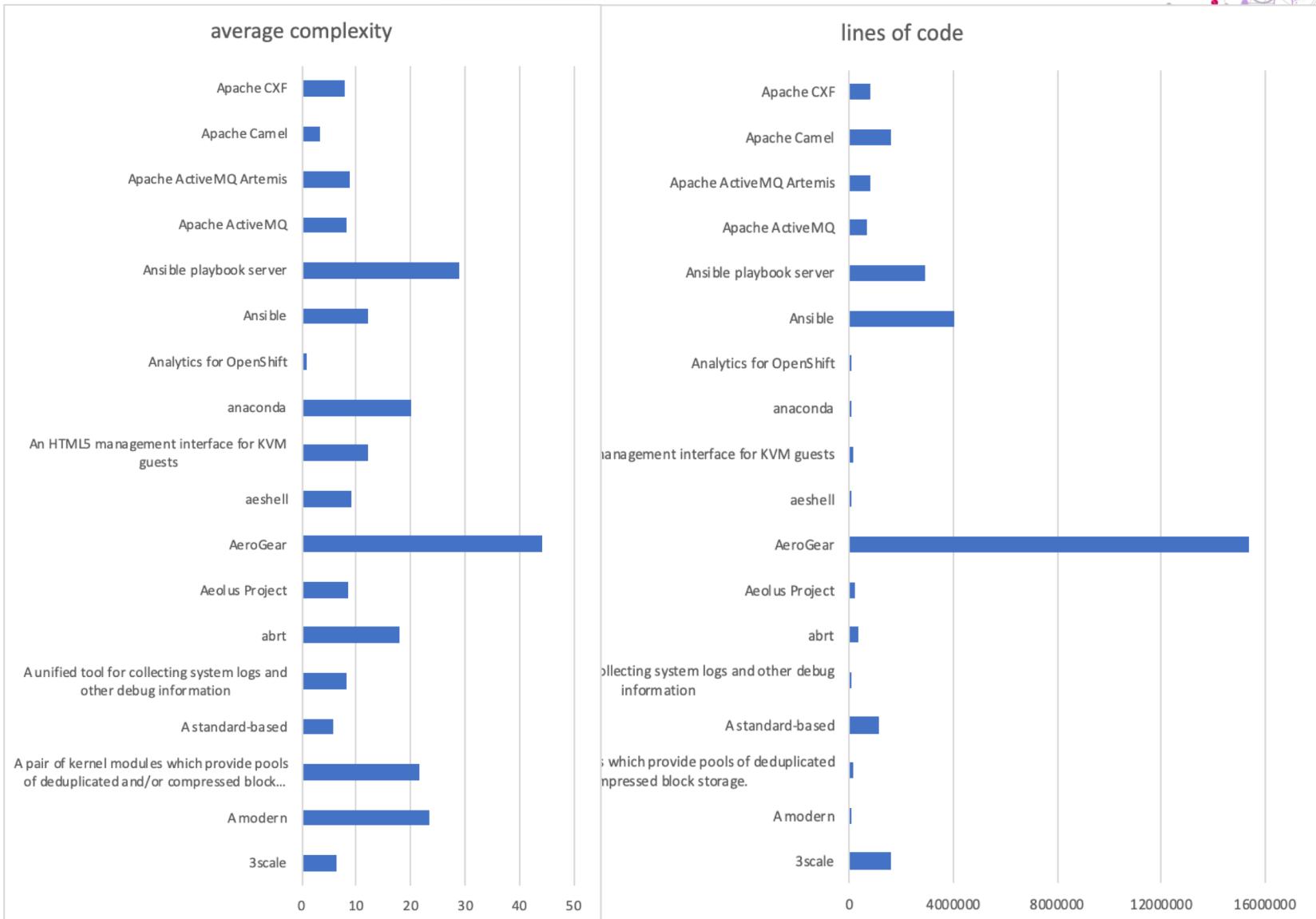




Value



CHAOS

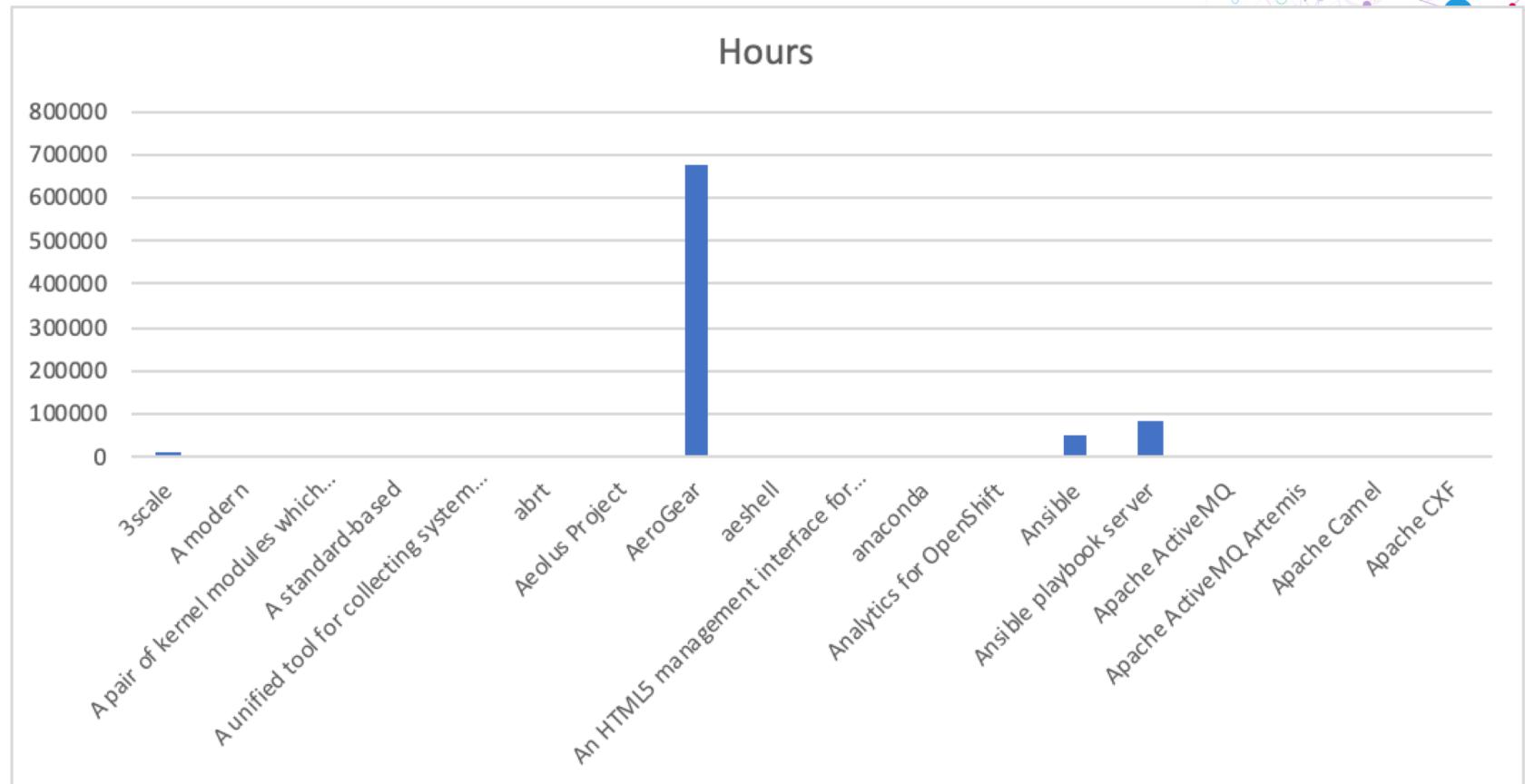


Value

name	Hours
3scale	10,173
A modern	1,162
A pair of kernel modules which...	2,631
A standard-based	6,723
A unified tool for collecting system...	166
abrt	5,600
Aeolus Project	1,862
AeroGear	678,080
aeshell	708
An HTML5 management interface for...	1,857
anaconda	1,651
Analytics for OpenShift	0
Ansible	48,892
Ansible playbook server	83,666
Apache ActiveMQ	5,560
Apache ActiveMQ Artemis	7,026
Apache Camel	5,370
Apache CXF	6,512



Simple Formula for Illustration: Lines of Code times Complexity
You can parameterize this using Augur, and put in your own formulas

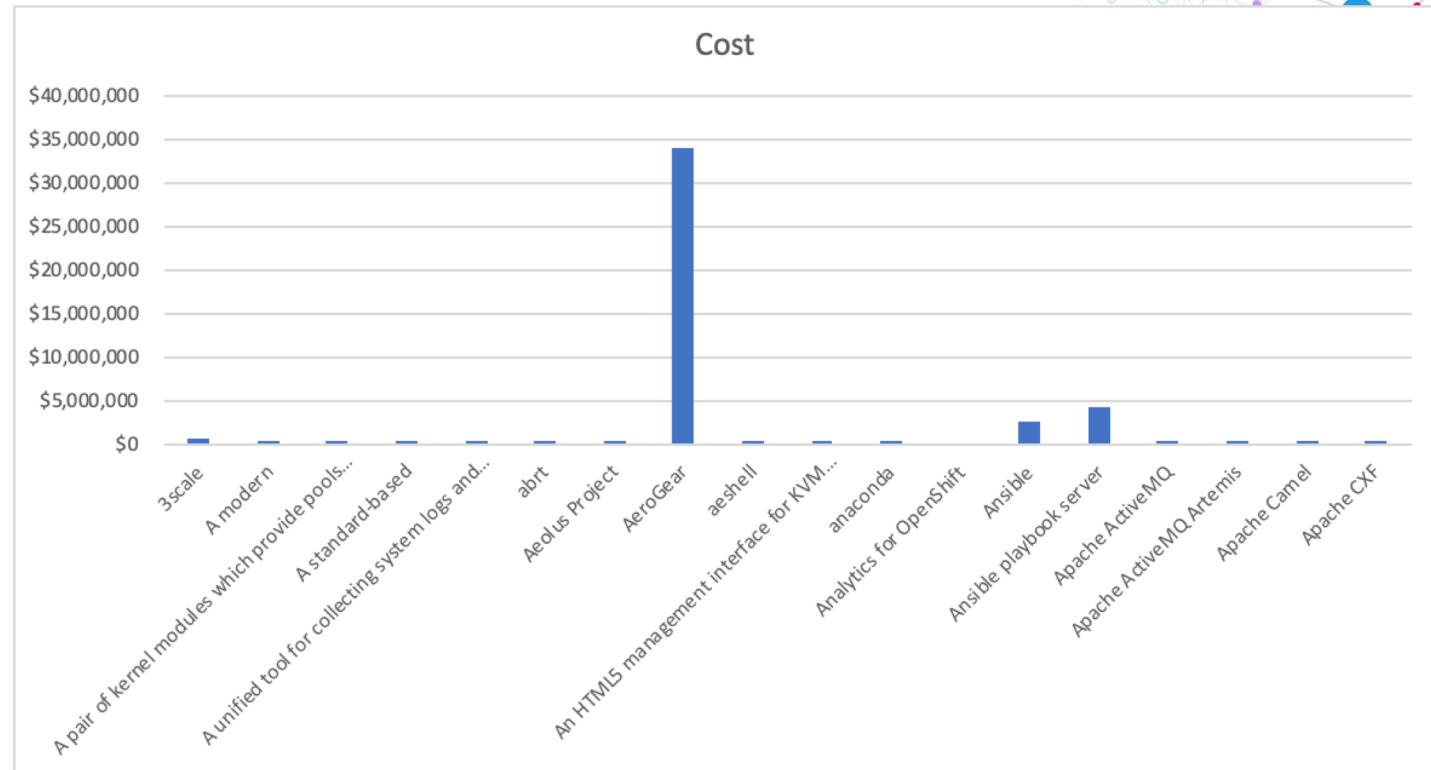


Value

Simple Formula for Illustration: (Hours: Lines of Code times Complexity) **times \$50/hr**
You can parameterize this using Augur, and put in your own formulas



name	Cost
3scale	508,671
A modern	58,096
A pair of kernel modules which provide pools...	131,532
A standard-based	336,158
A unified tool for collecting system logs and...	8,317
abrt	280,015
Aeolus Project	93,115
AeroGear	33,904,017
aeshell	35,396
An HTML5 management interface for KVM...	92,863
anaconda	82,575
Analytics for OpenShift	8
Ansible	2,444,603
Ansible playbook server	4,183,313
Apache ActiveMQ	278,006
Apache ActiveMQ Artemis	351,324
Apache Camel	268,516
Apache CXF	325,616



Value

What languages are used in a repository and to what extent?
Do the complexity estimations have face validity?

name	Cost
3scale	508,671
A modern	58,096
A pair of kernel modules which p	131,532
A standard-based	336,158
A unified tool for collecting syste	8,317
abrt	280,015
Aeolus Project	93,115
AeroGear	33,904,017
aeshell	35,396
An HTML5 management interface	92,863
anaconda	82,575
Analytics for OpenShift	8
Ansible	2,444,603
Ansible playbook server	4,183,313
Apache ActiveMQ	278,006
Apache ActiveMQ Artemis	351,324
Apache Camel	268,516
Apache CXF	325,616

3scale	16082	407.5
JSON	506336	0
Ruby	364529	2.1926
JavaScript	270803	83.7365
PHP	83434	7.5479
YAML	77829	0
Lua	55800	7.3698
Markdown	38926	0
CSS	32519	0
Go	27205	21.29
HTML	24425	0
Sass	15699	0.122
XML Schema	15391	0
Gherkin Specification	13148	1.3604
TypeScript	8692	8.4125



Risk

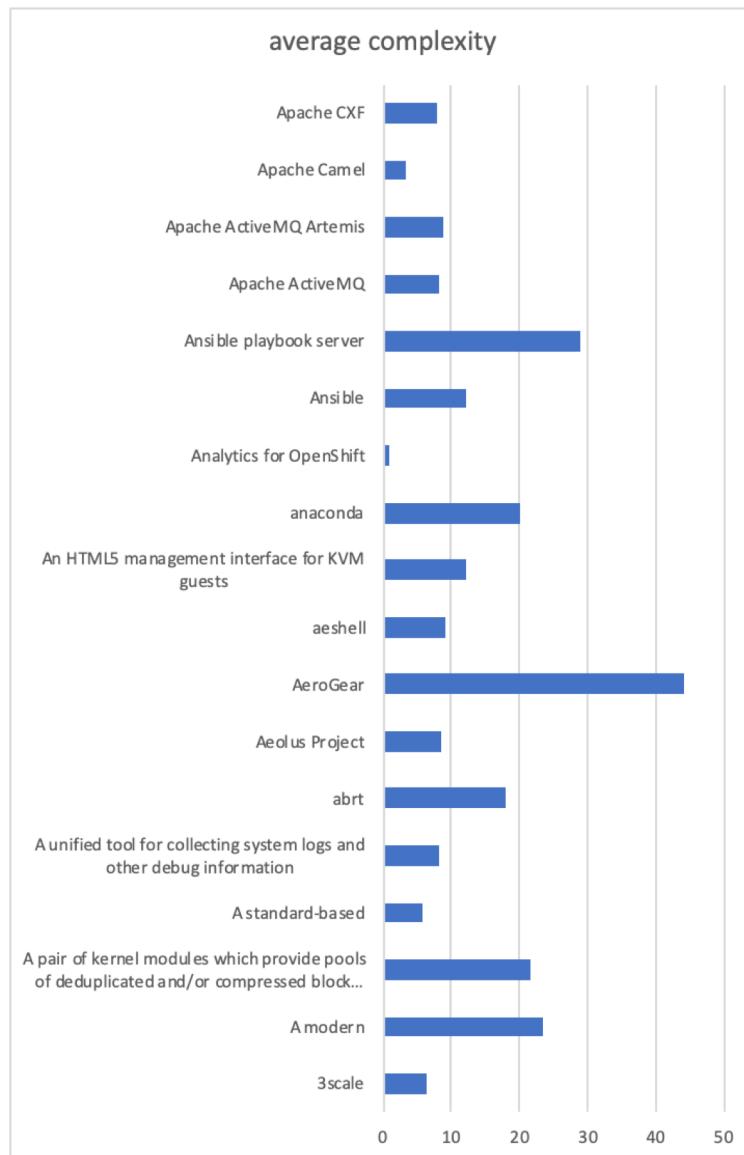


Risk Metrics Prototyping Initiation List (Draft)

1. Security
 - a. CII Best Practices Badge status (Developing. Included in Augur data model)
 - b. Programming Language Analysis (Developing. Included in Augur Data model)
2. Code Quality
 - a. Code complexity (Developing. Included in Augur Data model)
 - b. Test Coverage (Exploring tools)
 - c. Pull request approval rules in place (GitHub)
 - i. Metric: Exploring if a PR has a high number of comments Unique comments might be something to consider
3. Licensing
 - a. License SBOM analysis (DoSocs) (Developing. We have a prototype, but DoSocs is complex to deploy)
 - b. License declaration in repository (Developing. Included in Augur Data model)
4. Transparency
 - a. SBOM (DoSocs) (Developing. We have a prototype, but DoSocs is complex to deploy)
5. Business Risk
 - a. Community activity metrics
 - i. Code
 1. Commits (Available in Augur)
 2. LoC (Developing. We have a prototype, but DoSocs is complex to deploy. We do have “Top Ten Committer” data in the UI, but I think this analysis is different.)
 3. Committers (Developing. We have a prototype, but DoSocs is complex to deploy. We do have “Top Ten Committer” data in the UI, but I think this analysis is different.)
 - ii. Issues
 1. Issue resolution time (Developing. Included in Augur Data model)
 2. Issue volume and status (Developing. Included in Augur Data model)
 - b. Elephant Factor (Developing. Included in Augur Data model. We do have a data endpoint for this already, but are going to put it in the UI.)
 - c. Bus Factor (Developing. Included in Augur Data model)

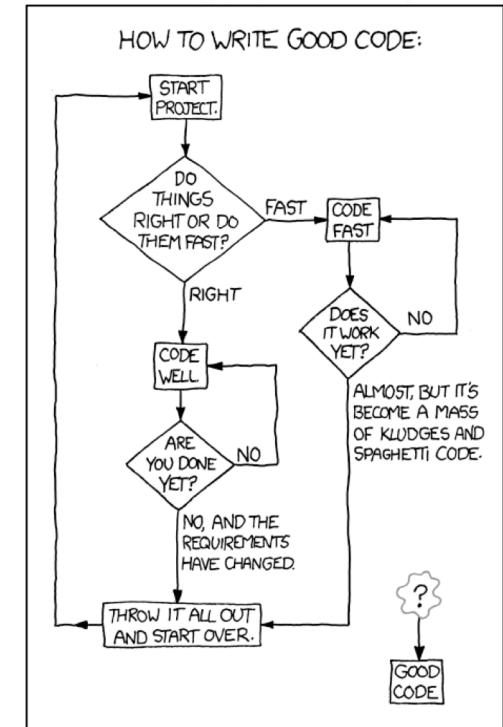
Risk

Estimating complexity with COCOMO inspired analysis

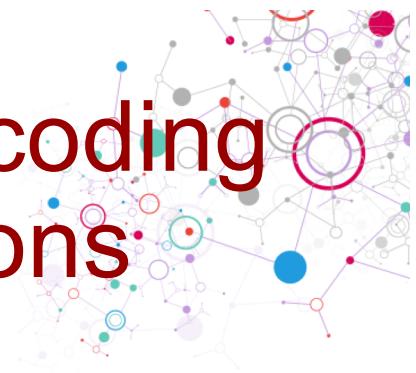


Why Risk and Value?

- No single health determination can be made across all open source projects, however:
 - **We can start to understand what composite metrics signal and how they can be related to actions**
- We aim to provide insight as local interpretations are done on the metrics
 - **Provide guideposts for what others have done in similar contexts and how peer communities compare**
- Develop an understanding of development processes based on facts



Common Working Group: Focus on Decoding on Identity in Compliance with Regulations



Contributors



Contributor
aliases



Organizational
Affiliation

Getting Engaged with the CHAOSS Community



Mailing lists

- General discussions: <https://lists.linuxfoundation.org/mailman/listinfo/chaoss-discussions>
- Metrics Committee: <https://lists.linuxfoundation.org/mailman/listinfo/oss-health-metrics>
- Software Committee: <https://lists.linuxfoundation.org/mailman/listinfo/chaoss-software>
- CHAOSS Board archives: <https://lists.linuxfoundation.org/pipermail/chaoss-members/>

IRC Channel: #chaoss-community on freenode

Wiki: <https://wiki.linuxfoundation.org/chaoss>

Participating in CHAOS



Code

- <https://github.com/chaoss>

Meetings

- Metrics committee
 - [Weekly hangouts](#)
 - [Monthly calls](#)
- Software committee

Events

- <https://wiki.linuxfoundation.org/chaoss/events>

A Growing Community



The screenshot shows the CHAOSS website homepage. At the top, there is a navigation bar with links for Projects, News, About, Wiki, Contact, and a search icon. The main header features the text "Community Health Analytics Open Source Software" and a "LEARN MORE" button. Below the header is a large, dark blue banner with a network graph background. The banner text reads "Community Health Analytics Open Source Software". On the left side of the page, there is a "Recent Posts" section. Two recent posts are listed: "Come help define the metric of community growth/maturity/decline in BOF" by Alex Contini (News) and "Learn more about Community Health Analytics Open Source Software (CHAOSS)" by Alex Contini (News). The overall design is clean and modern, emphasizing the theme of community and connectivity.

Discussion: Mailing Lists, Periodic Meetings, Weekly Hangouts, IRC Channels

Code: Issues, Pull requests welcome

CHAOS

chaoss.community

How to Get Started with CHAOSS



How to get started

- Join our [mail list](#) and introduce yourself.
- Look through the [mail list archive](#) for past discussions and self-introductions of people involved.
- Join our [monthly phone call](#) on your calendar.
- Join our [weekly hangout](#) on your calendar.

How to contribute

- Join our weekly hangouts and monthly calls.
- Share your experience with metrics.
- Let us know what you would like CHAOSS to do for you.
- Just participate in the conversation.

Community Health Analytics: Building a Shared Software and Data Infrastructure



Health Metric Software and Data Infrastructure:

1. How, to what extent and in what ways does your current research stand to benefit from leveraging and contributing to a shared software infrastructure?
2. How and to what extent are you gathering novel data or reusing current data in the course of your research?
3. What to do about metrics that run counter to open source culture but are becoming mainstream in the broader culture?

Join us to extract knowledge from



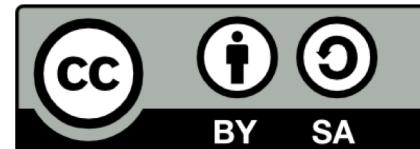
CHAOS

chaoss.community

CHAOS

License & Credits

(c) 2017 CHAOSS. Some rights reserved. This presentation is distributed under the “Attribution-ShareAlike 4.0” license, by Creative Commons, available at creativecommons.org/licenses/by-sa/4.0/



Credits:

- Good code”, Comic by Randall Munroe, XKCD 844
License: Creative Commons Attribution-NonCommercial 2.5
xkcd.com/844/