From Data Tsunami to Actionable Insights

SCALE 2024

Cali Dolfi, Red Hat
Dawn Foster, CHAOSS

https://chaoss.community/
https://github.com/chaoss
@chaoss@fosstodon.org

Introductions

Cali Dolfi
Senior Data Scientist, Red Hat
www.linkedin.com/in/calidolfi

Dawn Foster
Data Science, CHAOSS
fastwonderblog.com/
hachyderm.io/@geekygirldawn
www.linkedin.com/in/dawnfoster

Data Tsunami

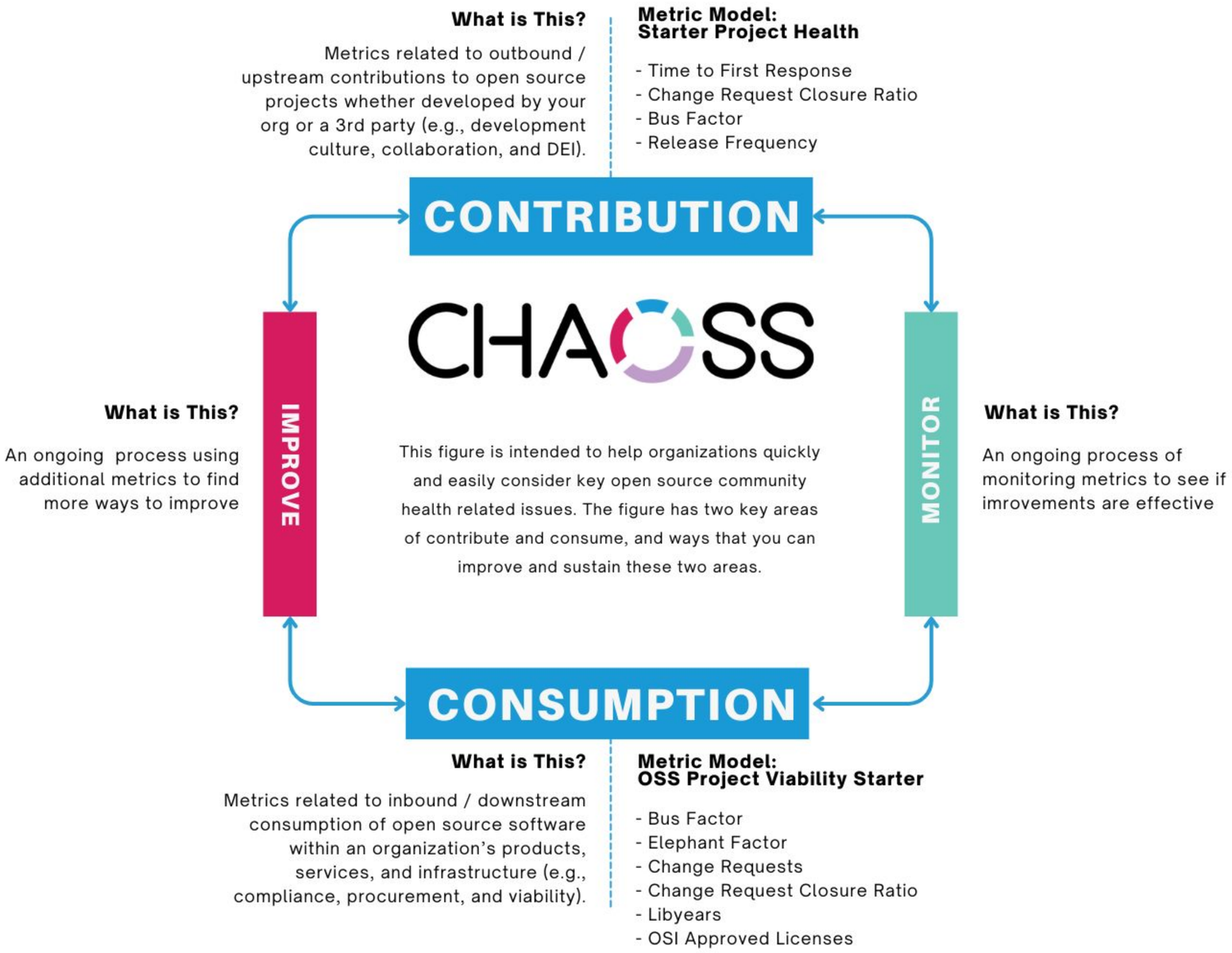
People can be overwhelmed
by a wall of metrics
and unsure how to start

Photo by Matt Paul Catalano on Unsplash

CHAOSS

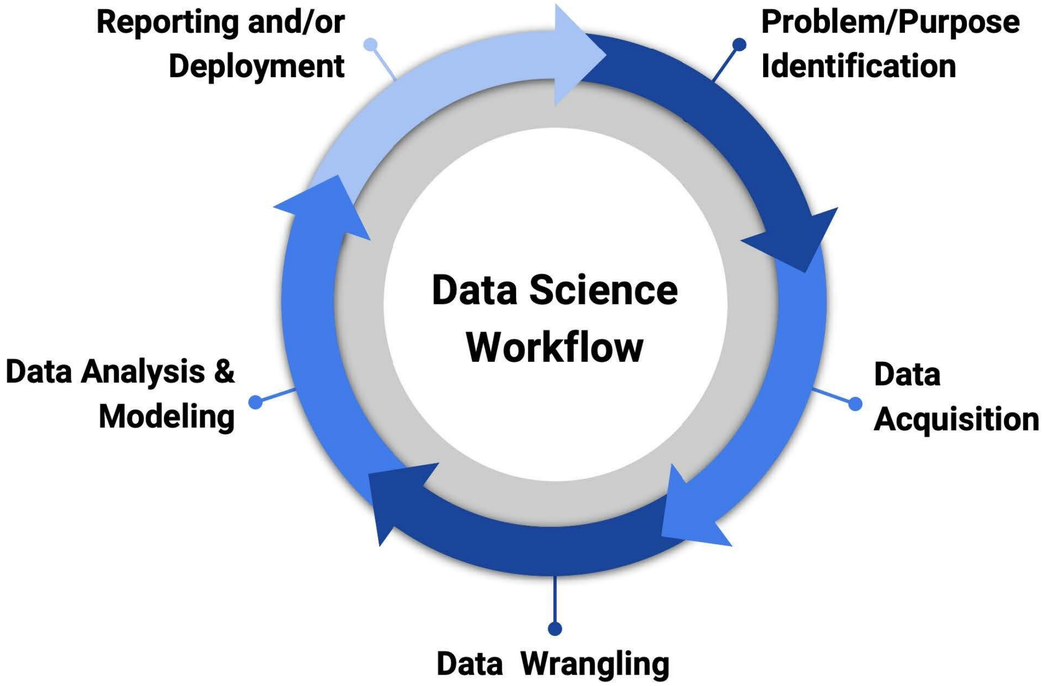
Metrics Models

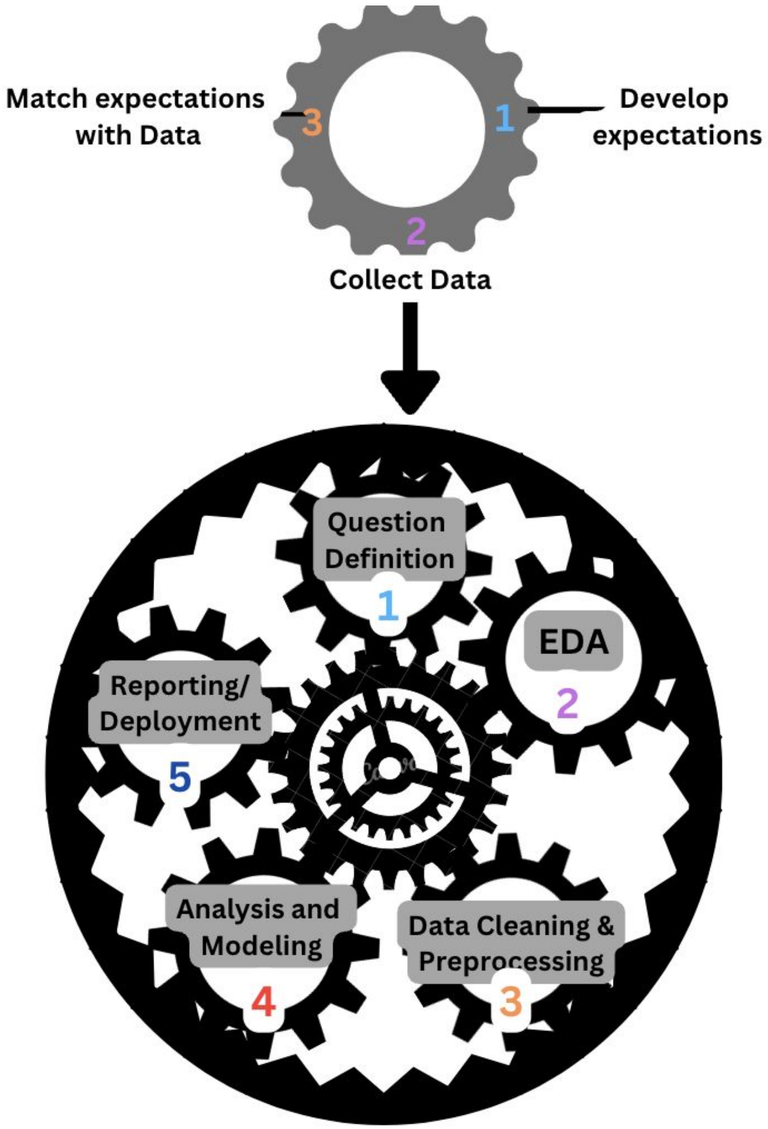
Using collections of metrics to focus on a particular topic



CHAOSS https://chaoss.community/kbtopic/all-metrics-models/

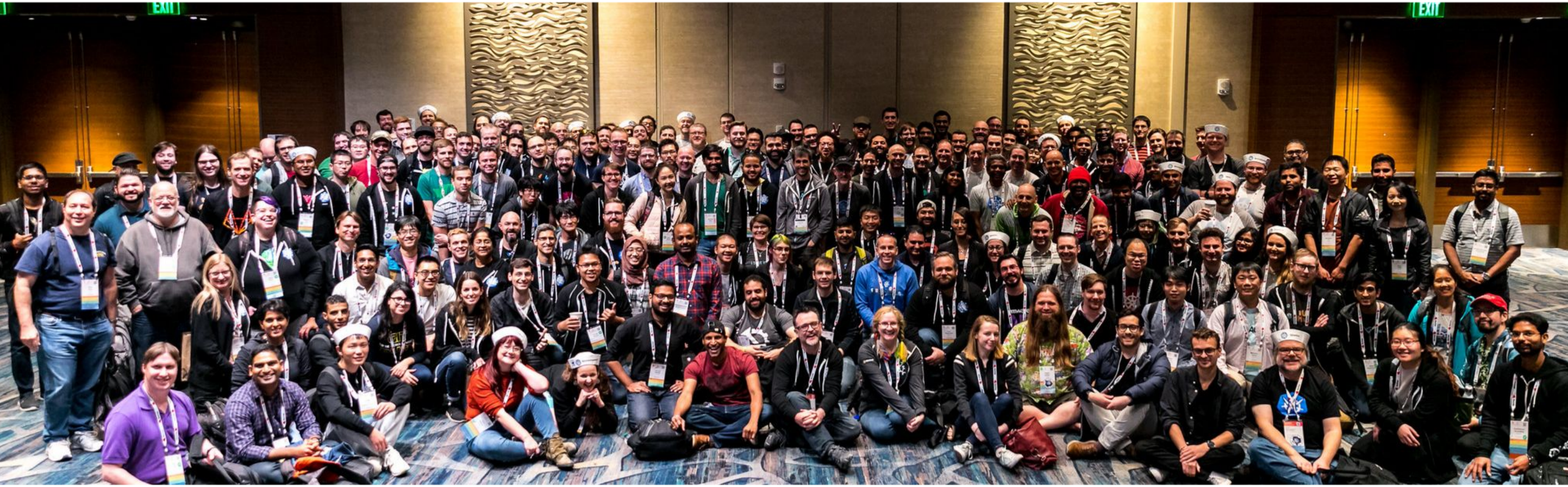
General Data Science Workflow





CHAOSS

What do you want to learn about your community?



CHAOSS

Converting a question to a metric

- ❑ Research on established metrics that could relate to the question
- ❑ Specific data points needed
- ❑ Visualization to represent the data
- ❑ Potential insights and actions

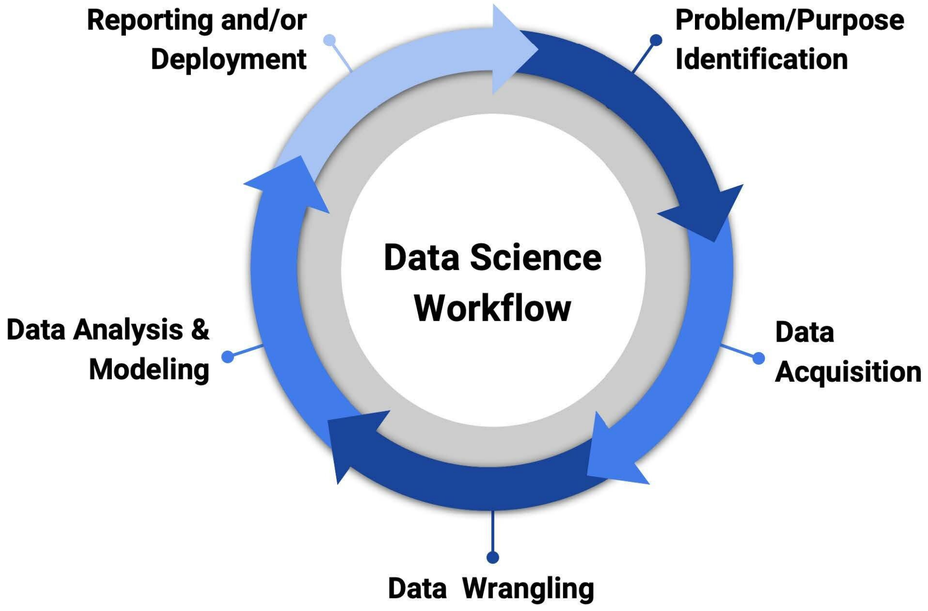
WIP Metric

Community Feedback

CHAOSS

Let’s talk data

- **Rarely:**
  - are you going to need a single data point or metric about a community
  - End with the same question, metrics, or visualization concept you started with
- **Always:**
  - Space leaves room for a concept to grow



Common OSS Community Data Sources

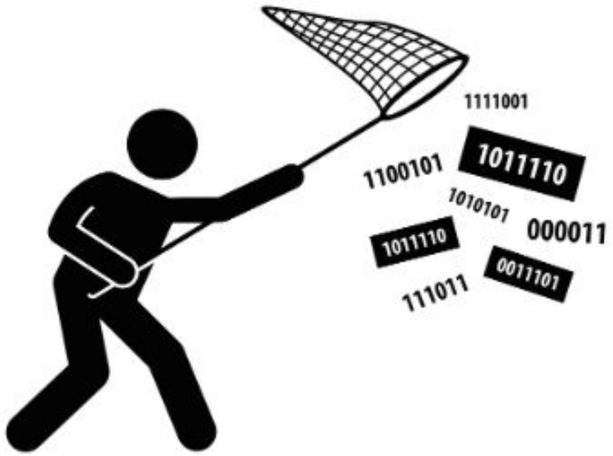
Community OSS Data Collection

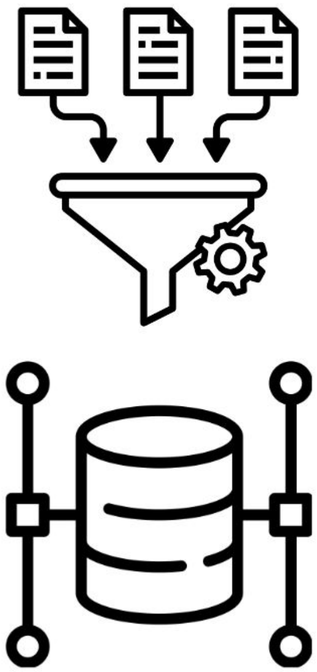
Retrieval Method

Output Format

In one time use cases or EDA, this format can be sufficient, but that data tsunami grows quick....

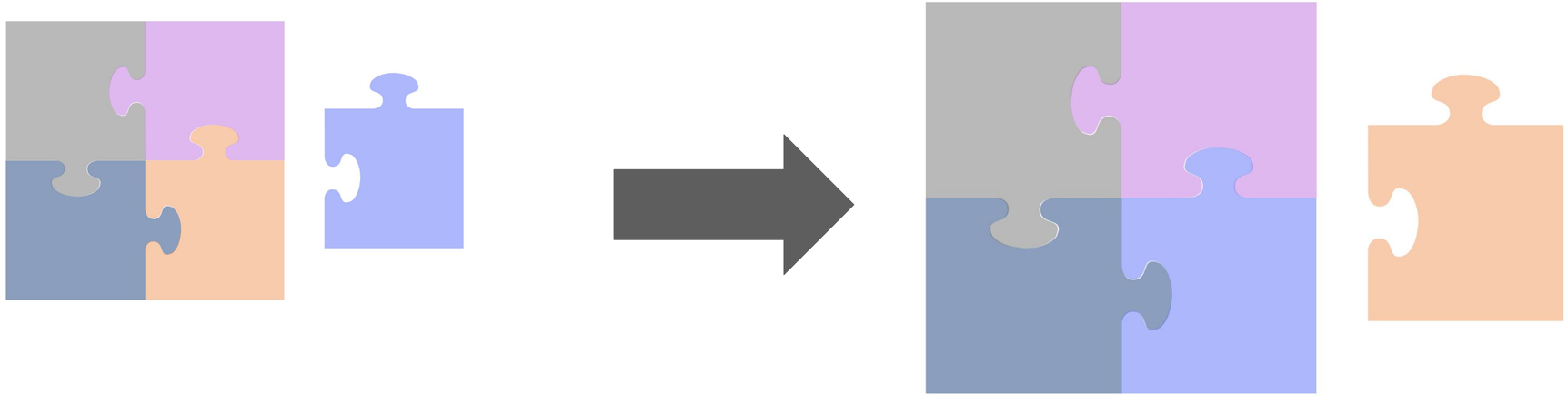
Setting up for a general use case - Relational Databases



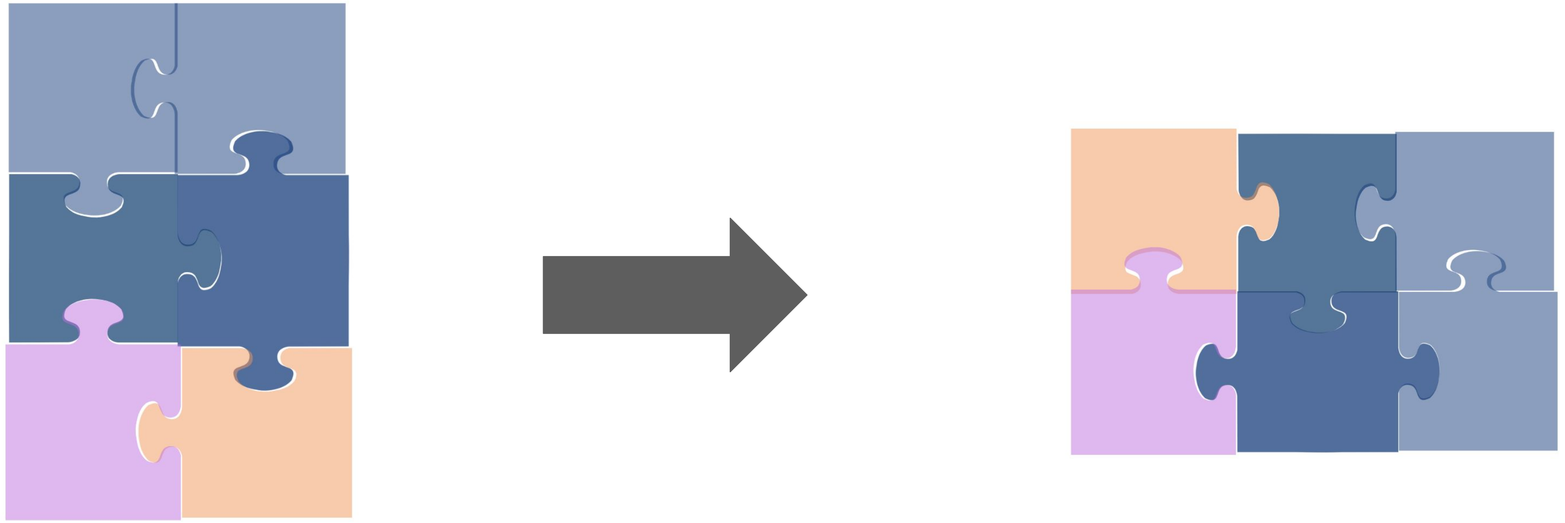


Standardized data structure - preprocessing out of the way and get to reuse work across different visualizations

<u>Structural similarity:</u>
Time to first response:
Issues vs PRs
Staleness:
Issues vs PRs



<u>Data Similarity:</u>
PR Review Assignments per contributor vs status counts
Fly By Contributors by month vs Repeat Contributors by month



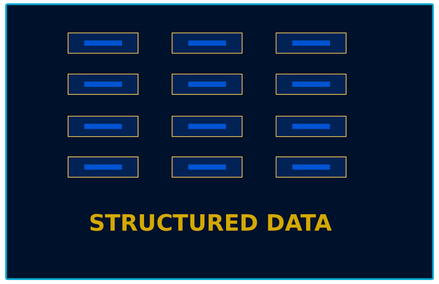
CHAOSS

How does this look in action?

Augur: A path to Data Science through a relational DB

Mountains of Data

6 Years of Data Carpentry

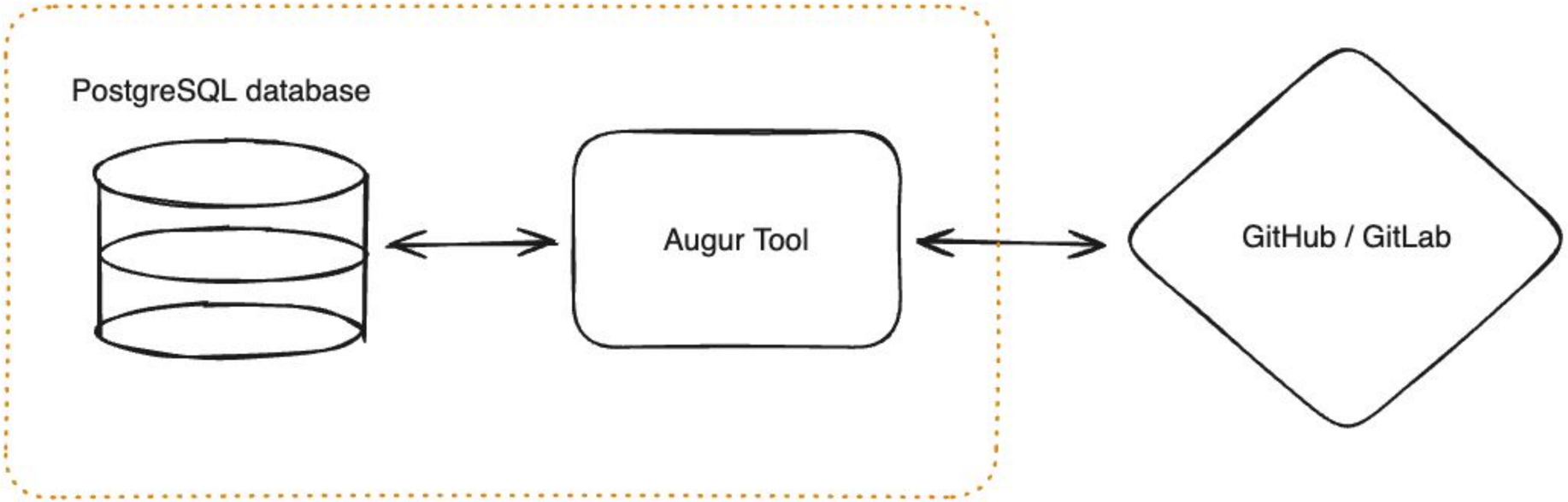


Structured Data

Validated Data

CHAOSS

Augur High-level Architecture



Augur Database

Relational database with organized Git Platform data and an enforced data validation

8Knot Dashboard

Data Science Tools: Dash-Ploty dashboard with the structure to visualize any analysis of the Augur data

8Knot/Augur Demo

CHAOSS

Interpretation:

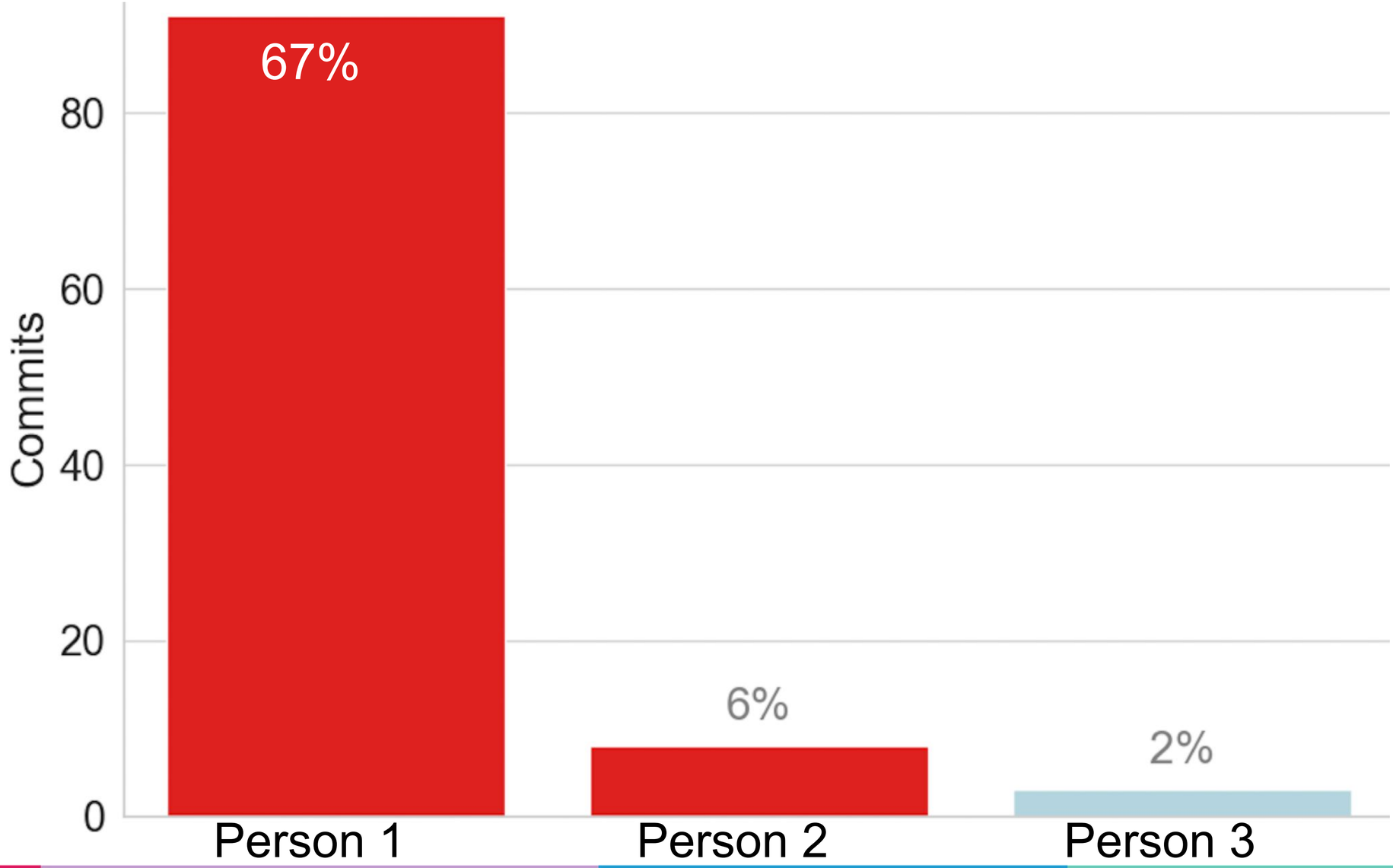
Improving your project and community

CHAOSS

https://opensource.net/measure-open-source-project-health/

Bus Factor for Contributor Sustainability

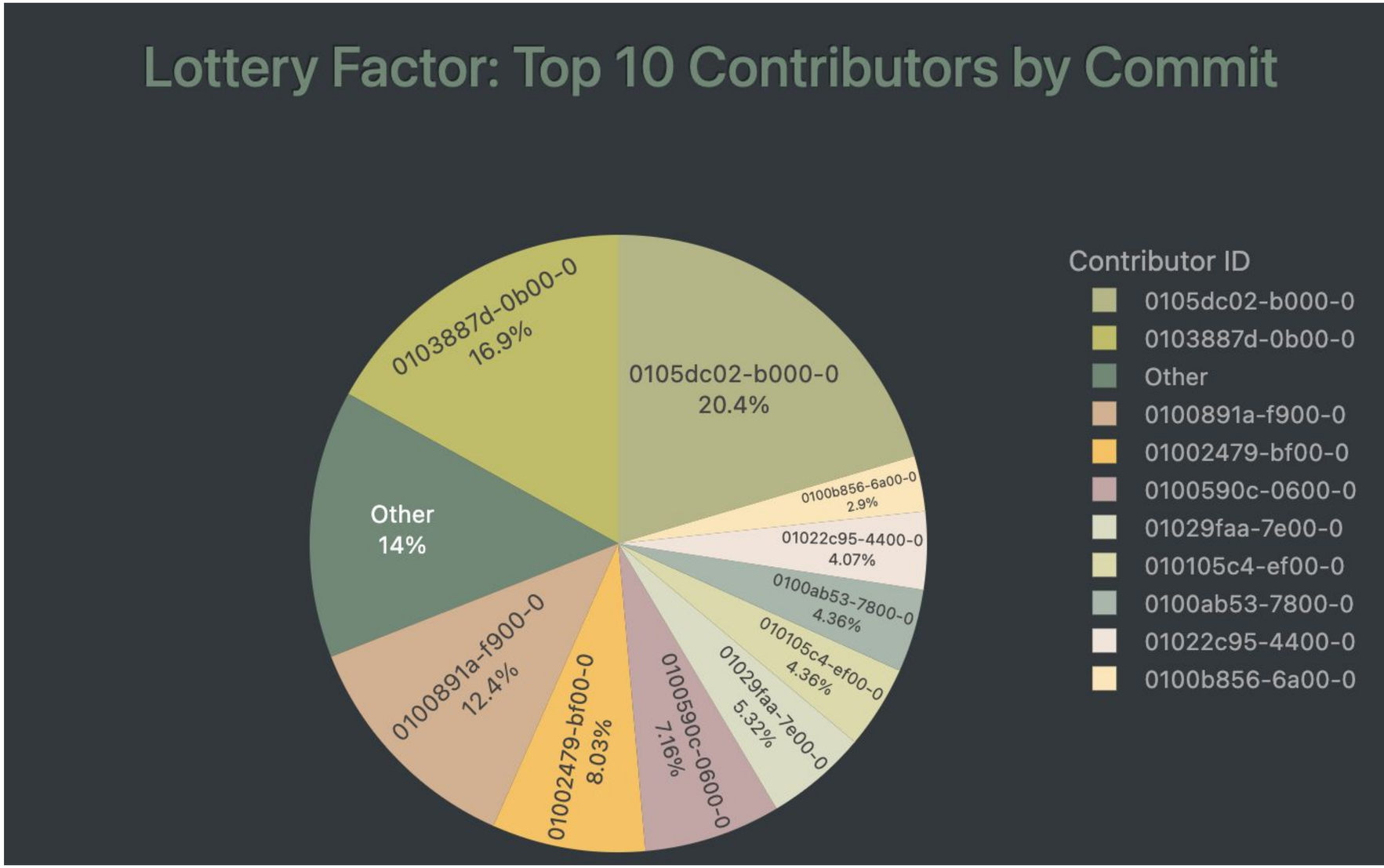
How big of an issue is it?



CHAOSS

Bus Factor for Contributor Sustainability

Who might be ready to move into a leadership position?

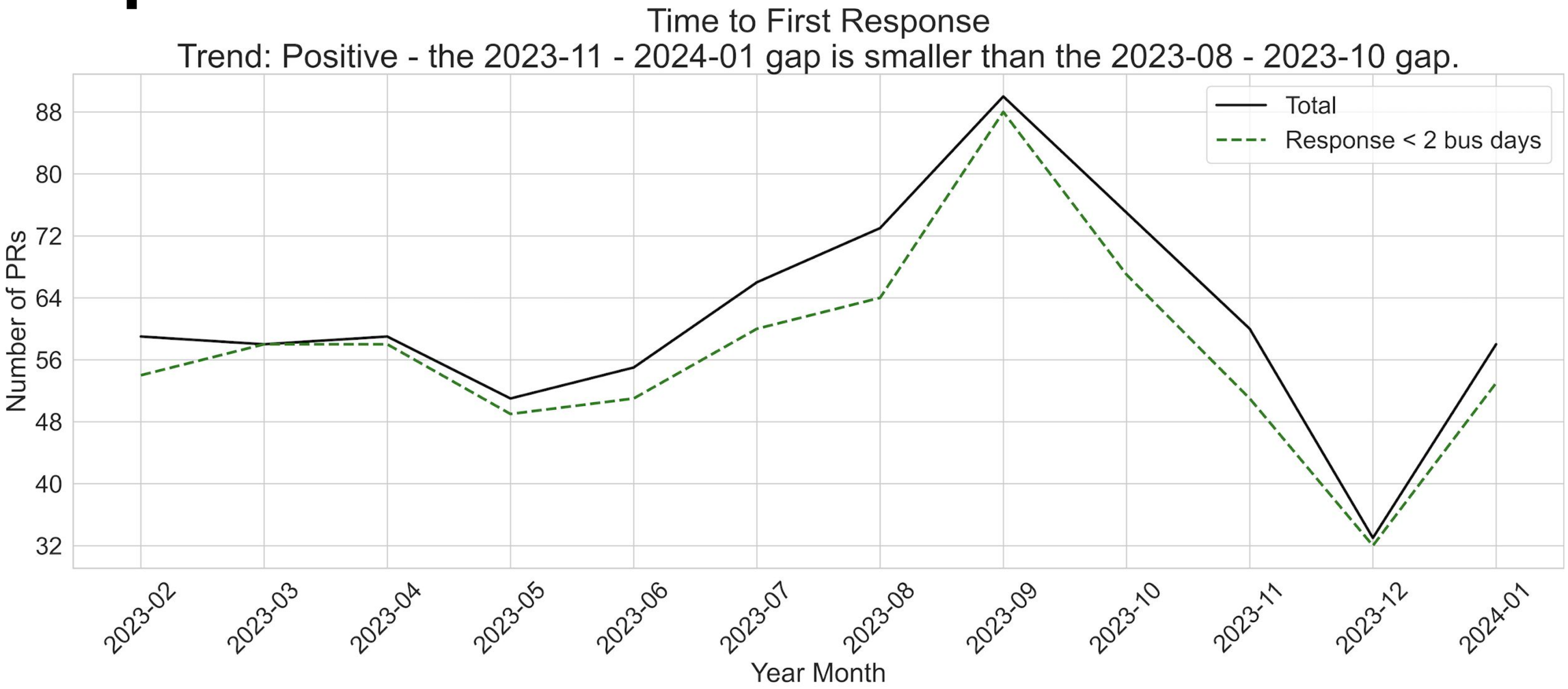


CHAOSS

Responsiveness: Closure Ratio

CHAOSS

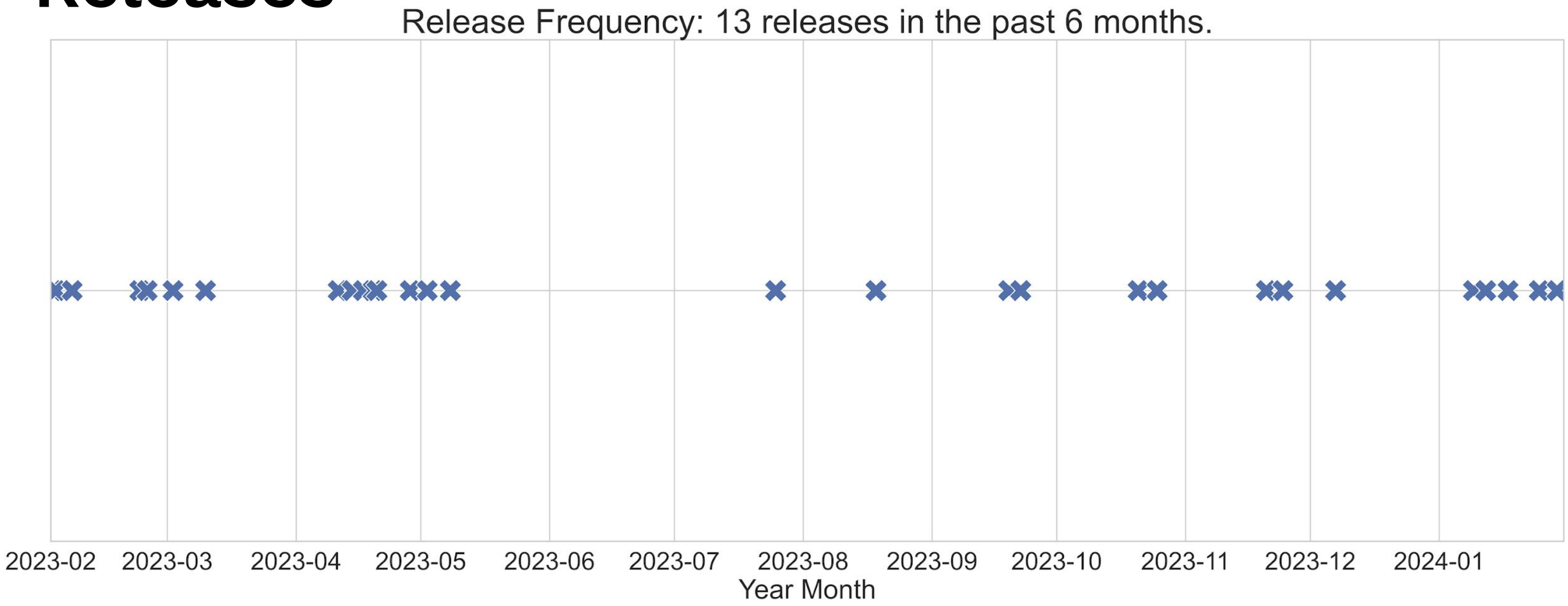
Responsiveness



Interpretation: Healthy projects will have little or no gap. A large or increasing gap requires attention.

CHAOSS

Releases

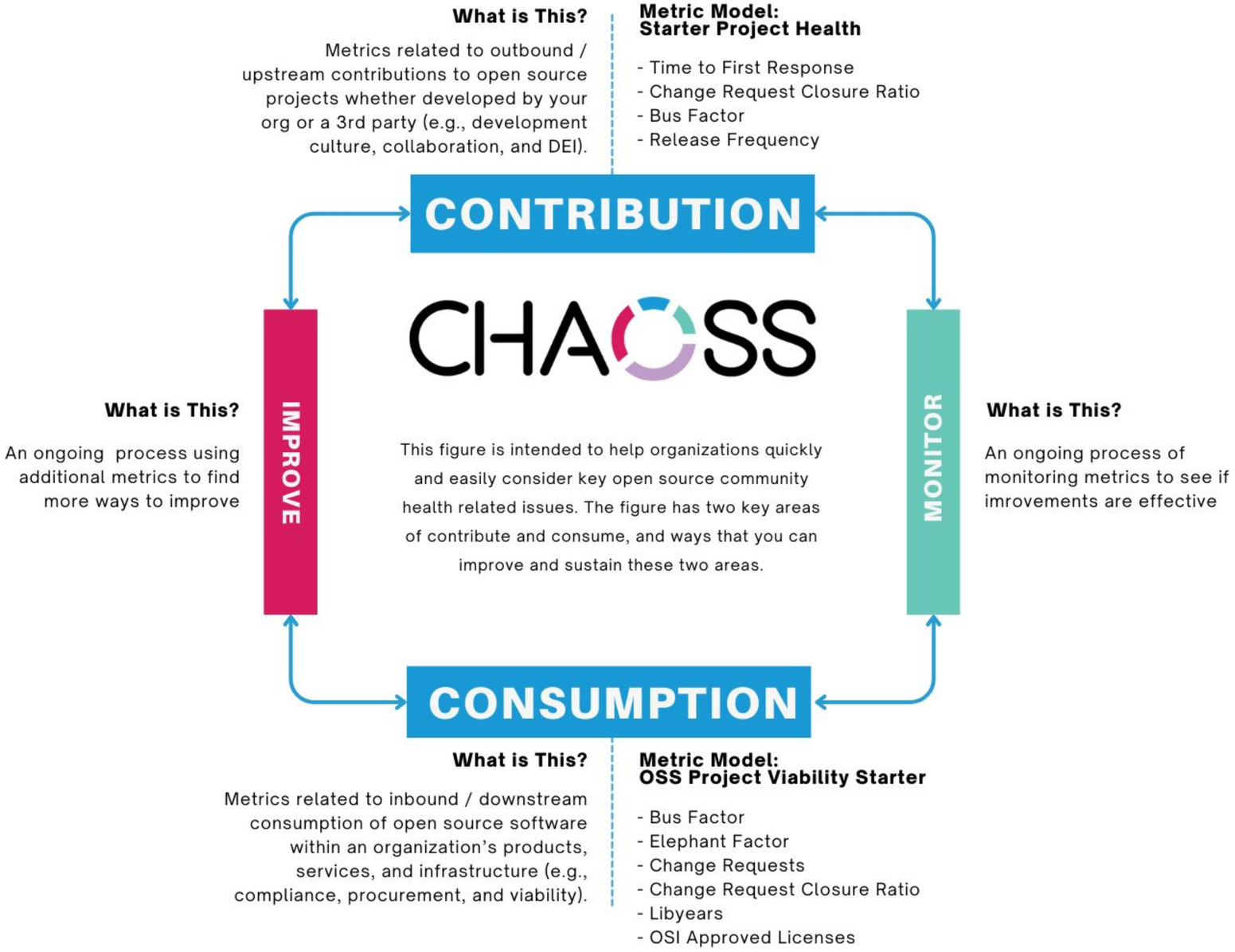


Interpretation: Healthy projects will have frequent releases with security updates, bug fixes, and features.

CHAOSS

Ongoing Cycle

Measure, improve, monitor, and repeat



Resources

https://chaoss.community

https://metrix.chaoss.io

https://github.com/oss-aspen/8Knot

https://cacm.acm.org/practice/beyond-the-repository

Final Thoughts

Use a data science workflow to convert the tsunami of data into actionable insights to improve your project.

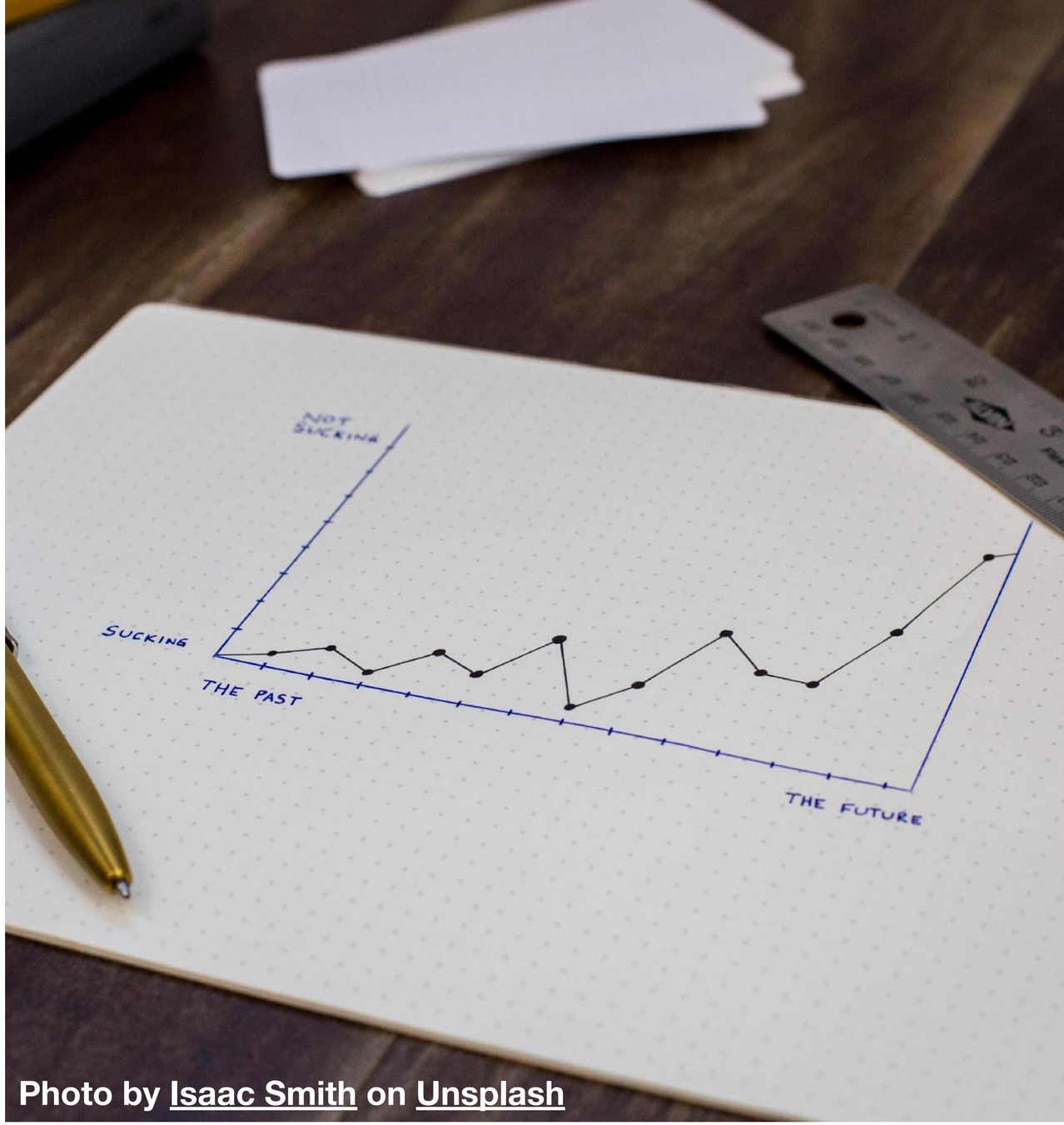


Photo by Isaac Smith on Unsplash

CHAOSS

THANK YOU!

Any Questions?

https://chaoss.community/
https://github.com/chaoss
@chaoss@fosstodon.org

CHAOSS