# TIL...: 10+ Findings from 2022 Global FLOSS Research

Kaylea Champion – kaylea@uw.edu https://www.kayleachampion.com https://communitydata.science University of Washington SeaGL—November 4-5, 2022



#! /usr/global/bin/research
import FLOSS

# So Much FLOSS in Research!

Caveats:

1. Most of this is not my work!

# So Much FLOSS in Research!

#### Caveats:

- 1. Most of this is not my work!
- 2. But it is still awesome!

### So Much FLOSS in Research!

#### Caveats:

- 1. Most of this is not my work!
- 2. But it is still awesome!
- 3. Read these papers for full details!



· Who: Xin Tan and Minghui Zhou, China



- · Who: Xin Tan and Minghui Zhou, China
- · Why: FLOSS projects are vital and sometimes get huge



- · Who: Xin Tan and Minghui Zhou, China
- · Why: FLOSS projects are vital and sometimes get huge
- · How: Series of projects: the kernel, node.js, fastlane



- · Who: Xin Tan and Minghui Zhou, China
- · Why: FLOSS projects are vital and sometimes get huge
- · How: Series of projects: the kernel, node.js, fastlane
- Key Finding: 1 + 1 + 1 + 1 = 2; automating and getting organized helps





· Who: Nan Yang, Isabella Ferreira, Alexander Serebrenik, Bram Adams - Netherlands, Canada



- · Who: Nan Yang, Isabella Ferreira, Alexander Serebrenik, Bram Adams Netherlands, Canada
- · Why: Foundations have been a key part of FLOSS growth



- · Who: Nan Yang, Isabella Ferreira, Alexander Serebrenik, Bram Adams Netherlands, Canada
- · Why: Foundations have been a key part of FLOSS growth
- · How: Analyze proposals to join the Apache Incubator program



- · Who: Nan Yang, Isabella Ferreira, Alexander Serebrenik, Bram Adams Netherlands, Canada
- · Why: Foundations have been a key part of FLOSS growth
- · How: Analyze proposals to join the Apache Incubator program
- Key Finding: Projects want community! Also, there's an increase in interest from corporations, Asia





· Who: Mariam Guizani, Thomas Zimmermann, Anita Sarma, Denae Ford – US



- · Who: Mariam Guizani, Thomas Zimmermann, Anita Sarma, Denae Ford US
- · Why: New features are being added to source code hosting platforms do they work?



- · Who: Mariam Guizani, Thomas Zimmermann, Anita Sarma, Denae Ford US
- · Why: New features are being added to source code hosting platforms do they work?
- How: Design a dashboard to implement features, then interview project leaders about them



- · Who: Mariam Guizani, Thomas Zimmermann, Anita Sarma, Denae Ford US
- · Why: New features are being added to source code hosting platforms do they work?
- · How: Design a dashboard to implement features, then interview project leaders about them
- Result: Dashboard is useful! Tagging good first issues, adding "social good" tags, using badges for rising contributors





· Who: Wenxin Xiao, Hao He, Weiwei Xu, Xin Tan, Jinhao Dong, Minghui Zhou – China



- · Who: Wenxin Xiao, Hao He, Weiwei Xu, Xin Tan, Jinhao Dong, Minghui Zhou China
- Why: Good first issue tagging is important for onboarding newcomers but time consuming



- · Who: Wenxin Xiao, Hao He, Weiwei Xu, Xin Tan, Jinhao Dong, Minghui Zhou China
- Why: Good first issue tagging is important for onboarding newcomers but time consuming
- How: Treat newcomer-resolved issues as ground truth data to train machine learning classifier



- · Who: Wenxin Xiao, Hao He, Weiwei Xu, Xin Tan, Jinhao Dong, Minghui Zhou China
- Why: Good first issue tagging is important for onboarding newcomers but time consuming
- How: Treat newcomer-resolved issues as ground truth data to train machine learning classifier
- · Key Finding: Machine learning can help with this problem!





· Who: Xunhui Zhang, Tao Wang, Yue Yu, Qiubing Zeng, Zhixing Li, and Huaimin Wang - China.



- · Who: Xunhui Zhang, Tao Wang, Yue Yu, Qiubing Zeng, Zhixing Li, and Huaimin Wang China.
- · Why: Sponsorship mechanism added to Github



- · Who: Xunhui Zhang, Tao Wang, Yue Yu, Qiubing Zeng, Zhixing Li, and Huaimin Wang China.
- · Why: Sponsorship mechanism added to Github
- · How: Survey sponsors, recipients, and would-be recipients



- · Who: Xunhui Zhang, Tao Wang, Yue Yu, Qiubing Zeng, Zhixing Li, and Huaimin Wang China.
- · Why: Sponsorship mechanism added to Github
- · How: Survey sponsors, recipients, and would-be recipients
- · Key Finding: Short-term impact, low dollar value. Too focused on the individual?





 Who: Hongbo Fang, Hemank Lamba, James Herbsleb, Bogdan Vasilescu – Carnegie Mellon, USA



- Who: Hongbo Fang, Hemank Lamba, James Herbsleb, Bogdan Vasilescu Carnegie Mellon, USA
- · Why: What's the impact of social media on FLOSS?



- Who: Hongbo Fang, Hemank Lamba, James Herbsleb, Bogdan Vasilescu Carnegie Mellon, USA
- · Why: What's the impact of social media on FLOSS?
- How: Studying Twitter and GitHub



- Who: Hongbo Fang, Hemank Lamba, James Herbsleb, Bogdan Vasilescu Carnegie Mellon, USA
- · Why: What's the impact of social media on FLOSS?
- How: Studying Twitter and GitHub
- Key Finding: Tweeting works especially if you want folks who are newcomers to FLOSS.





· Who: Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill – US



Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill. 2022. ADetecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed-Source Approaches. In Software Engineering in Society (ICSE-SEIS'22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 15 pages. https://cmustrudel.github.io/papers/seis22pushback.pdf

- · Who: Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill US
- · Why: Toxicity and unnecessary pushback are stressful and wasteful



Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill. 2022. ADetecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed-Source Approaches. In Software Engineering in Society (ICSE-SEIS'22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 15 pages. https://cmustrudel.github.io/papers/seis22pushback.pdf

- · Who: Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill US
- · Why: Toxicity and unnecessary pushback are stressful and wasteful
- · How: analyzing text and log data



Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill. 2022. ADetecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed-Source Approaches. In Software Engineering in Society (ICSE-SEIS'22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 15 pages. https://cmustrudel.github.io/papers/seis22pushback.pdf

- · Who: Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill US
- · Why: Toxicity and unnecessary pushback are stressful and wasteful
- · How: analyzing text and log data
- Key Finding: Usage of "you" is a signal a review might be toxic. Toxicity signals unnecessary pushback.



Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kastner, Carolyn Egelman, Ciera Jaspan, Emerson Murphy-Hill. 2022. ADetecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed-Source Approaches. In Software Engineering in Society (ICSE-SEIS'22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 15 pages.

https://cmustrudel.github.io/papers/seis22pushback.pdf



· Who: Stefano Lambiase, Gemma Catolino, Damian A. Tamburri, Alexander Serebrenik, Fabio Palomba, and Filomena Ferrucci (Italy, the Netherlands)



- · Who: Stefano Lambiase, Gemma Catolino, Damian A. Tamburri, Alexander Serebrenik, Fabio Palomba, and Filomena Ferrucci (Italy, the Netherlands)
- · Why: FLOSS Communities are geographically and culturally diverse does it matter?



- Who: Stefano Lambiase, Gemma Catolino, Damian A. Tamburri, Alexander Serebrenik, Fabio Palomba, and Filomena Ferrucci (Italy, the Netherlands)
- Why: FLOSS Communities are geographically and culturally diverse does it matter?
- · How: Mining Github: contributor network analysis, location analysis



- · Who: Stefano Lambiase, Gemma Catolino, Damian A. Tamburri, Alexander Serebrenik, Fabio Palomba, and Filomena Ferrucci (Italy, the Netherlands)
- Why: FLOSS Communities are geographically and culturally diverse does it matter?
- · How: Mining Github: contributor network analysis, location analysis
- Key Finding: Geographical dispersion introduces organizational trade-offs; high concentration of participants from individualistic cultures may increase risk of "lone wolves"



CORRESPONDENCE

POST CARD

ADDRESS ONLY

\* KODAK

· Who: Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu – China, Ireland



- · Who: Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu China, Ireland
- · Why: Commits are important but under-examined



- · Who: Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu China, Ireland
- · Why: Commits are important but under-examined
- · How: 1600 commit messages, 5 projects



- · Who: Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu China, Ireland
- · Why: Commits are important but under-examined
- · How: 1600 commit messages, 5 projects
- Key Finding: 44% of commits don't explain "what" they do or "why" the work was done (...or are missing both elements!).





· Who: Elgun Jabrayilzade, Mikhail Evtikyhiev, Eray Tuzun, Vladimir Kovalenko—Turkey, Russia, The Netherlands



Elgun Jabrayilzade, Mikhail Evtikhiev, Eray Tüzün, Vladimir Kovalenko. 2022. Bus Factor in Practice. In 44th International Conference on Software Engineering (ICSE '22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 10 pages. https://arxiv.org/abs/2202.01523

- · Who: Elgun Jabrayilzade, Mikhail Evtikyhiev, Eray Tuzun, Vladimir Kovalenko—Turkey, Russia, The Netherlands
- · Why: Projects are at risk when key people leave



Elgun Jabrayilzade, Mikhail Evtikhiev, Eray Tüzün, Vladimir Kovalenko. 2022. Bus Factor in Practice. In 44th International Conference on Software Engineering (ICSE '22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 10 pages. https://arxiv.org/abs/2202.01523

- · Who: Elgun Jabrayilzade, Mikhail Evtikyhiev, Eray Tuzun, Vladimir Kovalenko—Turkey, Russia, The Netherlands
- · Why: Projects are at risk when key people leave
- · How: Comparing measures of activity with survey results



Elgun Jabrayilzade, Mikhail Evtikhiev, Eray Tüzün, Vladimir Kovalenko. 2022. Bus Factor in Practice. In 44th International Conference on Software Engineering (ICSE '22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 10 pages. https://arxiv.org/abs/2202.01523

- · Who: Elgun Jabrayilzade, Mikhail Evtikyhiev, Eray Tuzun, Vladimir Kovalenko—Turkey, Russia, The Netherlands
- · Why: Projects are at risk when key people leave
- · How: Comparing measures of activity with survey results
- Key Finding: Key people do more than just write code—and losing them can indeed stall projects.



Elgun Jabrayilzade, Mikhail Evtikhiev, Eray Tüzün, Vladimir Kovalenko. 2022. Bus Factor in Practice. In 44th International Conference on Software Engineering (ICSE '22), May 21–29, 2022, Pittsburgh, PA, USA. ACM, New York, NY, USA, 10 pages. https://arxiv.org/abs/2202.01523



#### In Other News....:

• Funding + Community? https://www.codeforsociety.org/incubator/



#### In Other News....:

- Funding + Community? https://www.codeforsociety.org/incubator/
- Huge influx of \$\$\$ for Software Supply Chain Security



#### In Other News....:

- Funding + Community? https://www.codeforsociety.org/incubator/
- Huge influx of \$\$\$ for Software Supply Chain Security
- · Machine learning everywhere





## 2022 FLOSS Research In Summary....



Scale? Automate, review, document!



Foundations for community! (???)



Tag it and badge it!



Good first issues are detectable!



Money isn't everything!



Tweeting helps!



Toxicity is a sign of unnecessary pushback!



Cultural diversity: hard, helpful!



Good commits state what and why!



Bus factor: real & more than code!



Incubator Opportunity - \$\$\$supply

chain - Machine Learning!

# Thank You!