Measuring the Health of the LLVM Community

Jeremy Bennett

Copyright © 2025 Embecosm. Freely available under a Creative Commons Attribution-ShareAlike license.

The *git* Repository

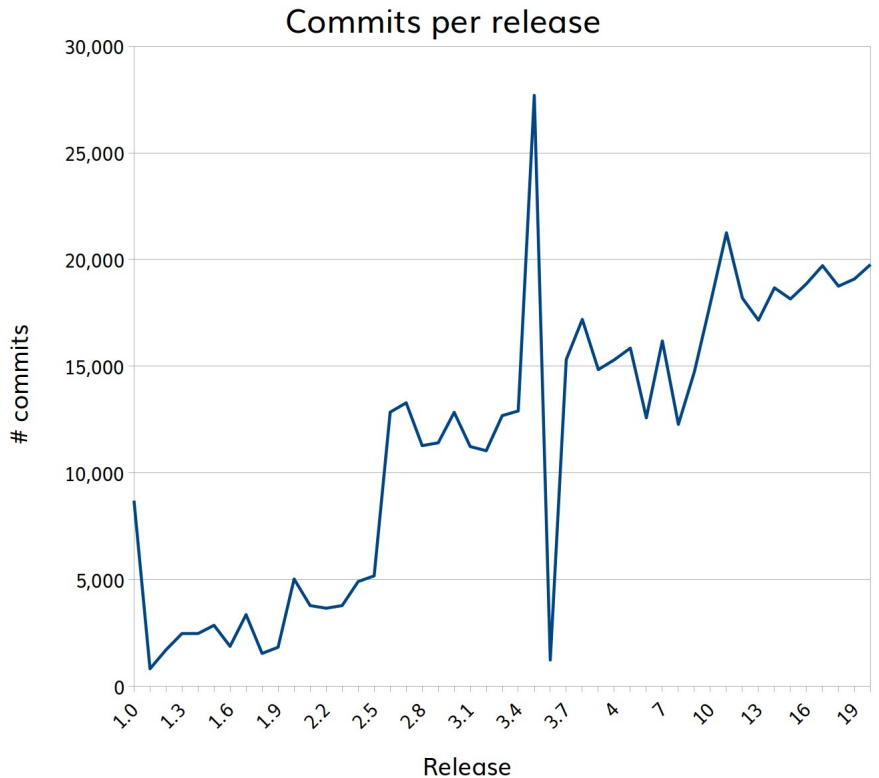
Copyright © 2025 Embecosm. Freely available under a Creative Commons Attribution-ShareAlike license.

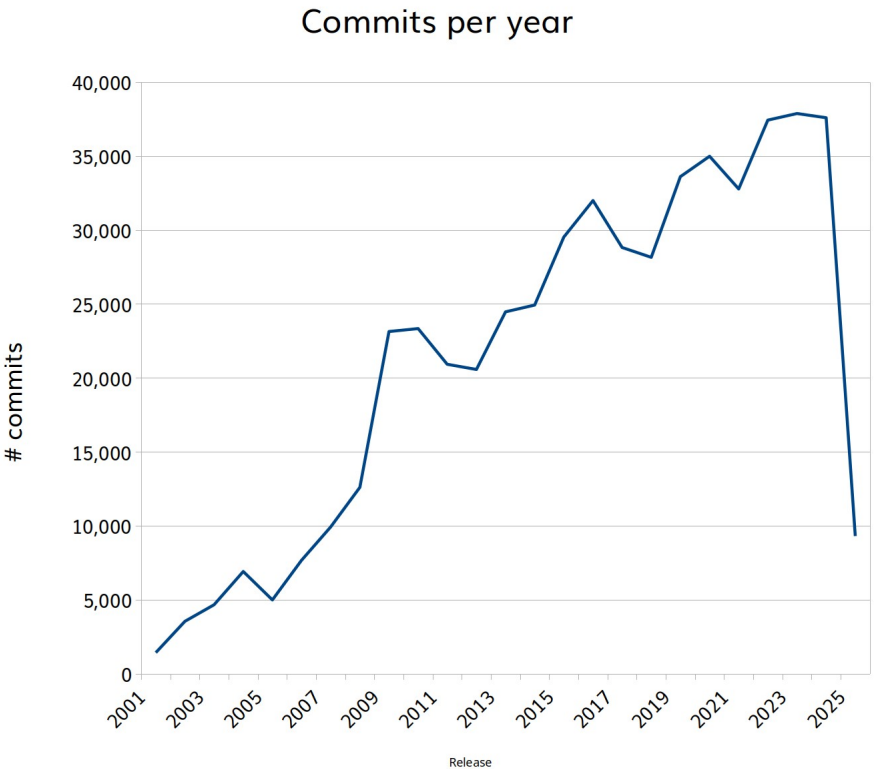
Counting commits

```
git log --oneline --no-merges remotes/upstream/release/20.x \
        ^remotes/upstream/release/19.x | wc -l

git log --oneline --no-merges --since-as-filter="2024-01-01" \
        --until="2024-12-31" | wc -l
```

CC BY SA

Counting Commits

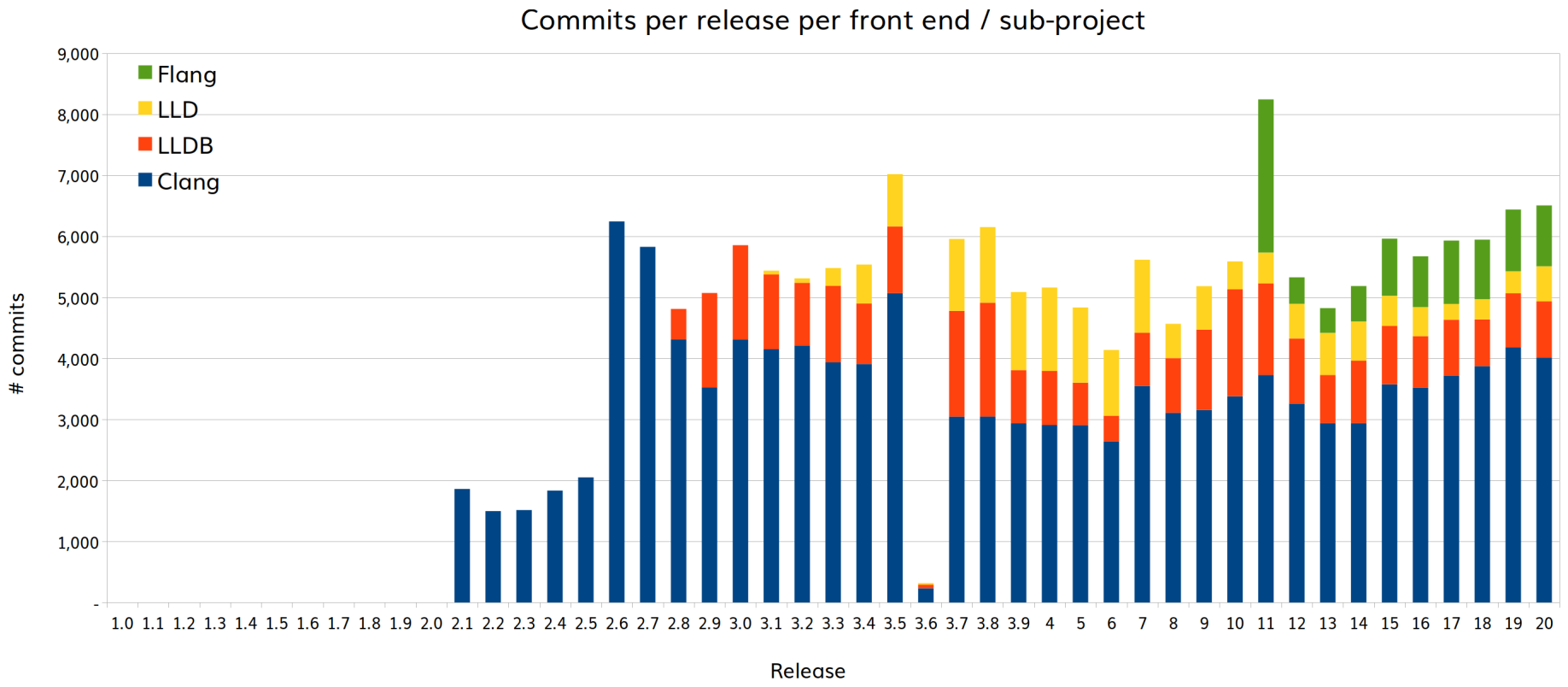




CC BY SA

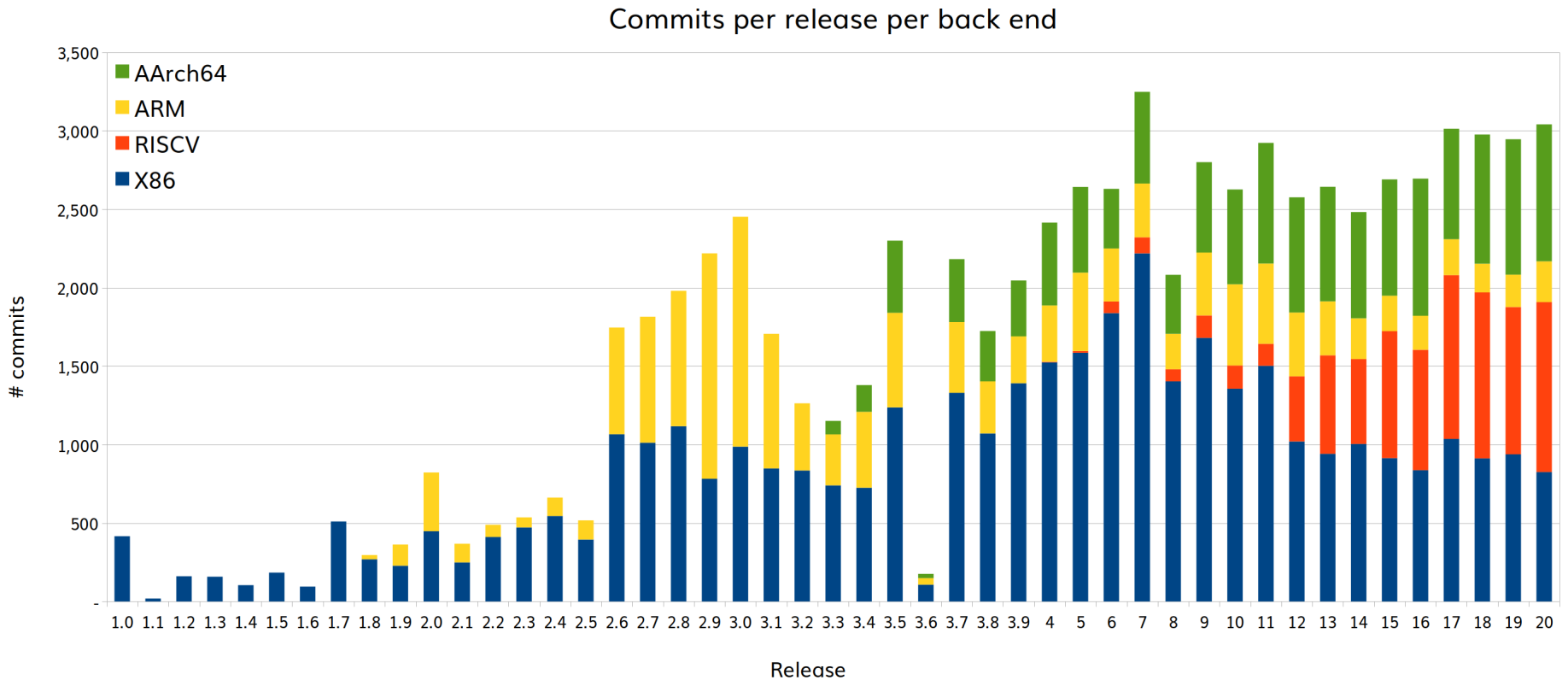
EMBECOSM®

Front End & Sub-Project Activity (Commits)



CC BY SA

Back End Activity (Commits)



CC BY SA

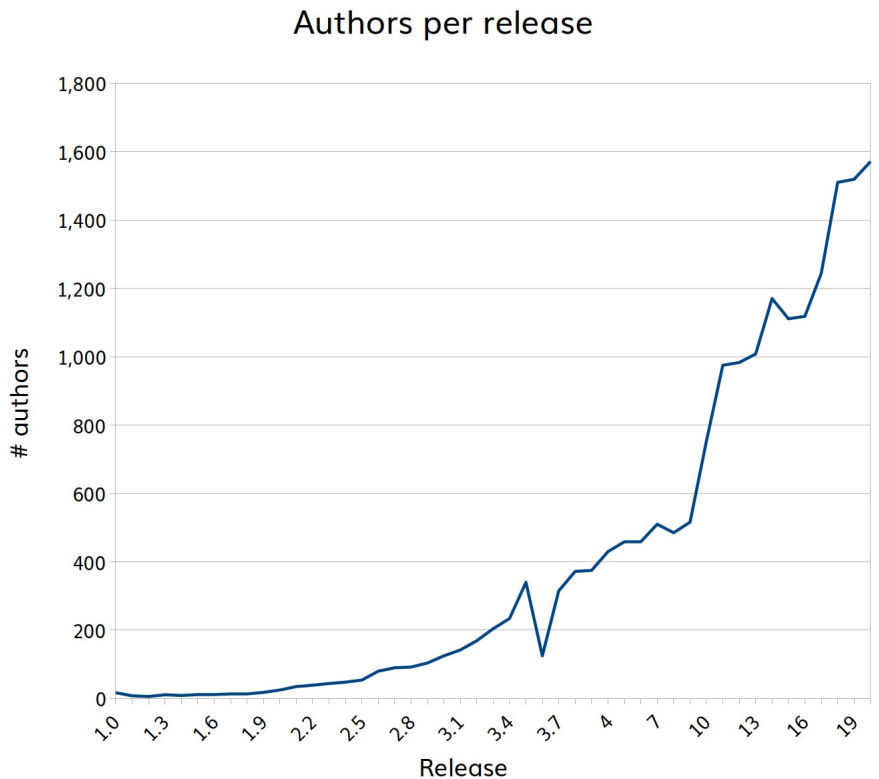
Counting Contributors

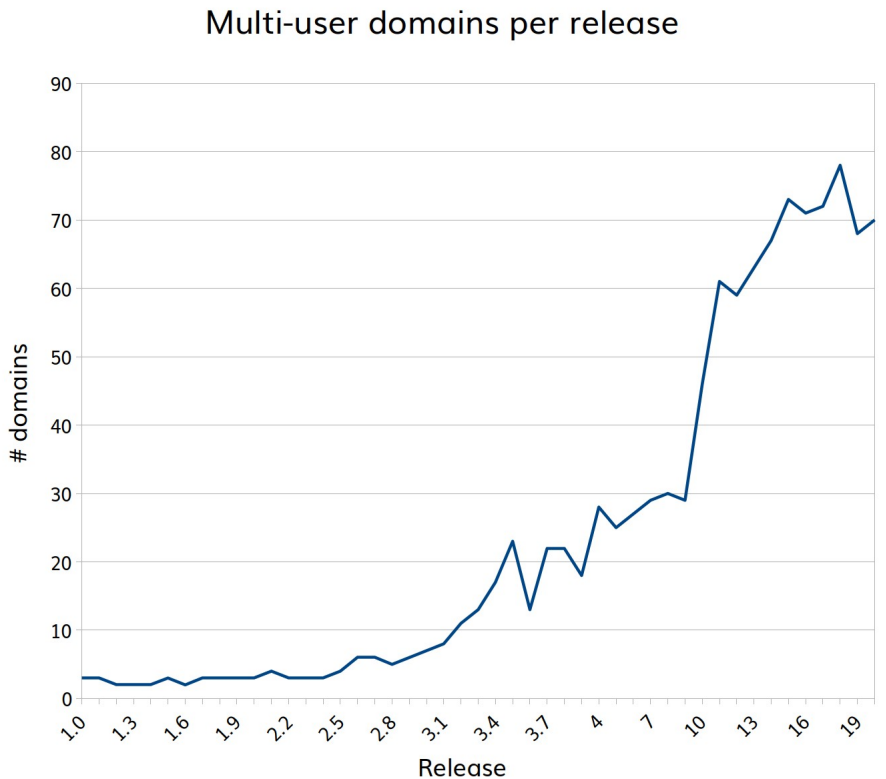
```
git log --no-merges remotes/upstream/release/20.x \
        ^remotes/upstream/release/19.x \
        --pretty="format:%an" | sort -u | wc -l

git log --no-merges remotes/upstream/release/20.x \
        ^remotes/upstream/release/19.x \
        --pretty="format:%ae" | sed -e 's/^.\+@//' | sort -u
```

Then post-process to remove generic email domains (gmail etc) and any domain with a single contributors

CC BY SA

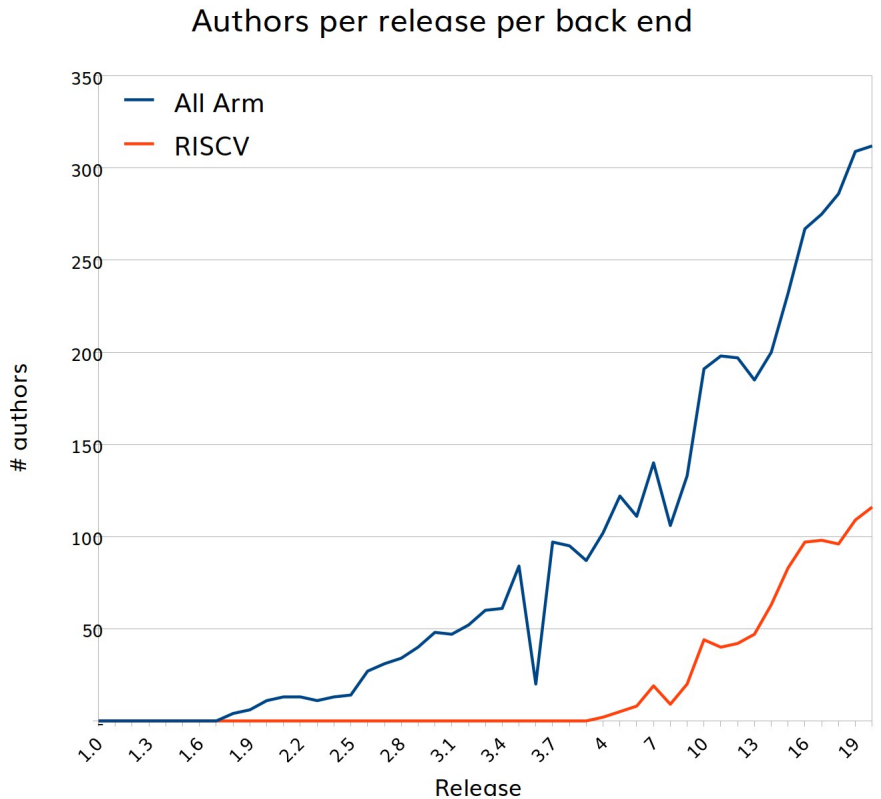
Counting Contributors

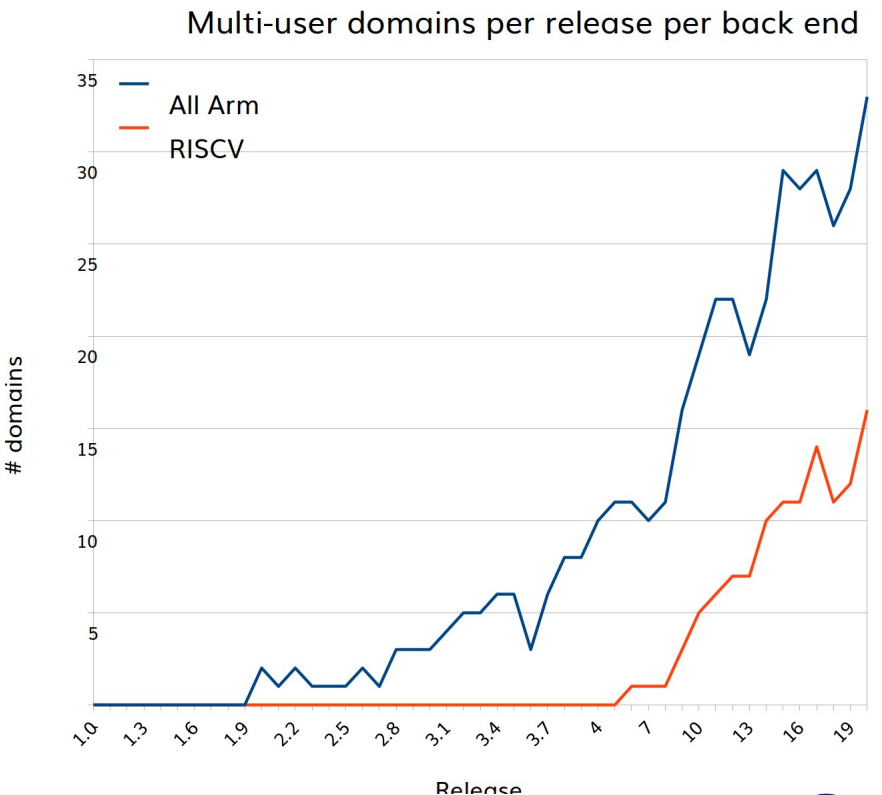




CC BY SA

The Detail: Comparing Back Ends





CC BY SA

The Detail: LLVM Supercommitters

Release 20

- 1,571 individuals
- 70 “corporate” domains
- 63 individuals committed more than once per week
- Highest individual: 767 commits

CC BY SA

The Detail: LLVM Supercommitters

Release 20

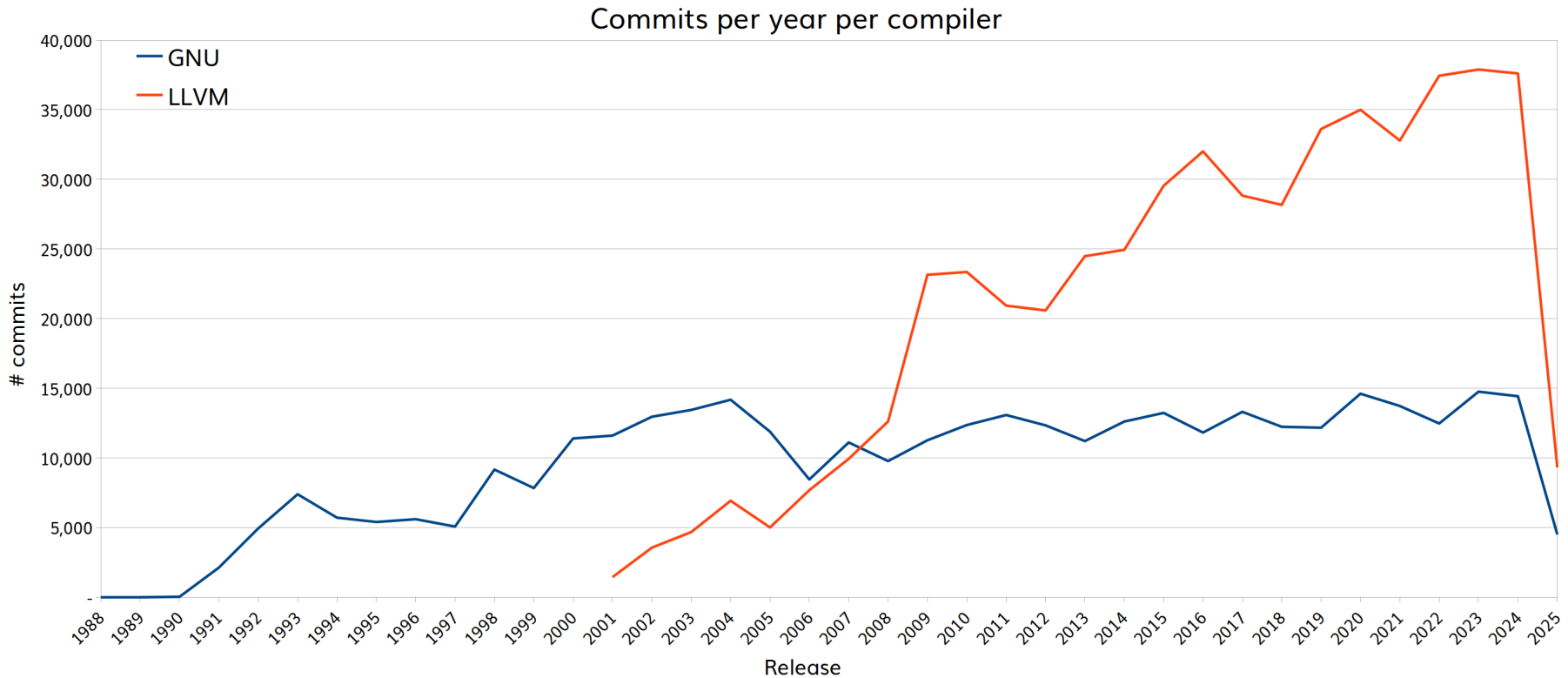
- 1,571 individuals
- 70 “corporate” domains
- 63 individuals committed more than once per week
- Highest individual: 767 commits

Release 1.0

- 17 individuals
- 2 “corporate” domains
- 7 individuals committed more than once per week
- Highest individual: 6,511 commits

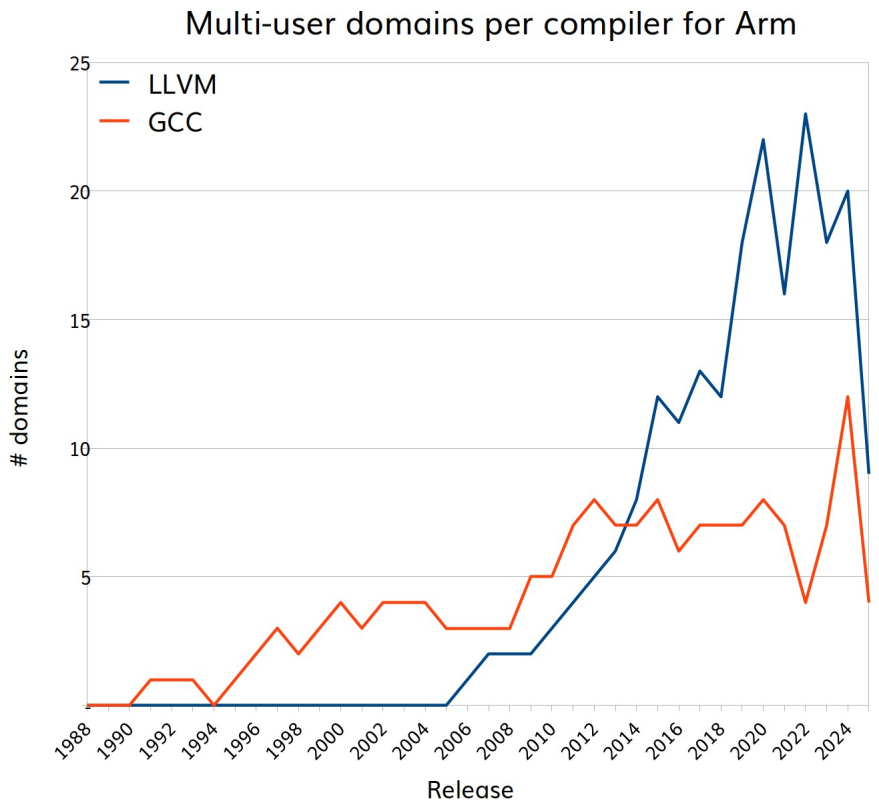
CC BY SA

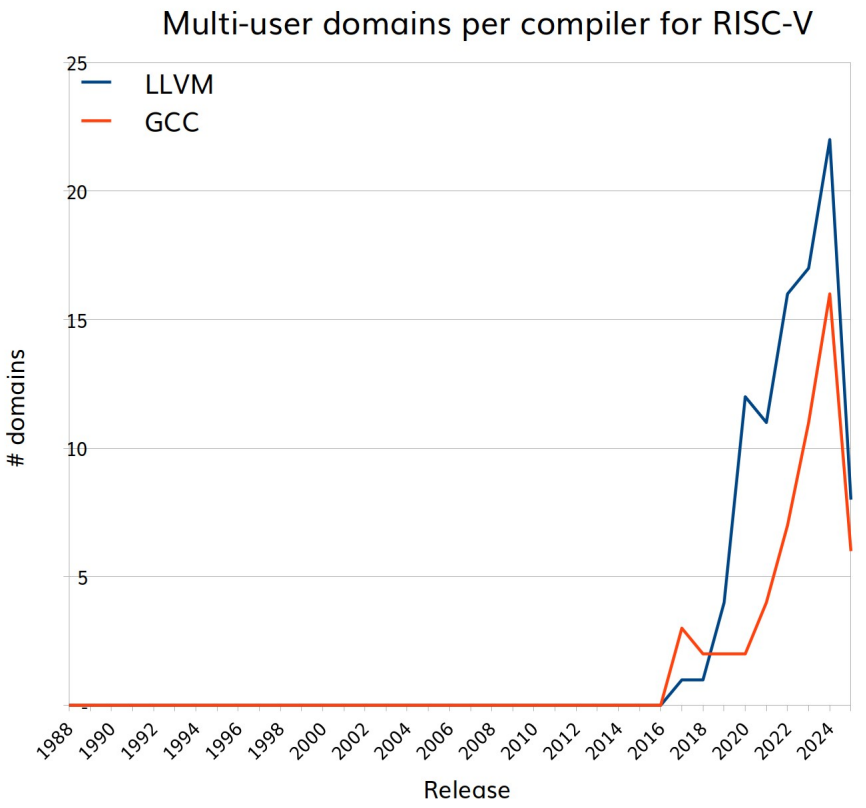
Comparing Compilers: Commits per Year



CC BY SA

The Detail: Comparing Compiler Back Ends





CC BY SA

GitHub using *gh*

Copyright © 2025 Embecosm. Freely available under a Creative Commons Attribution-ShareAlike license.

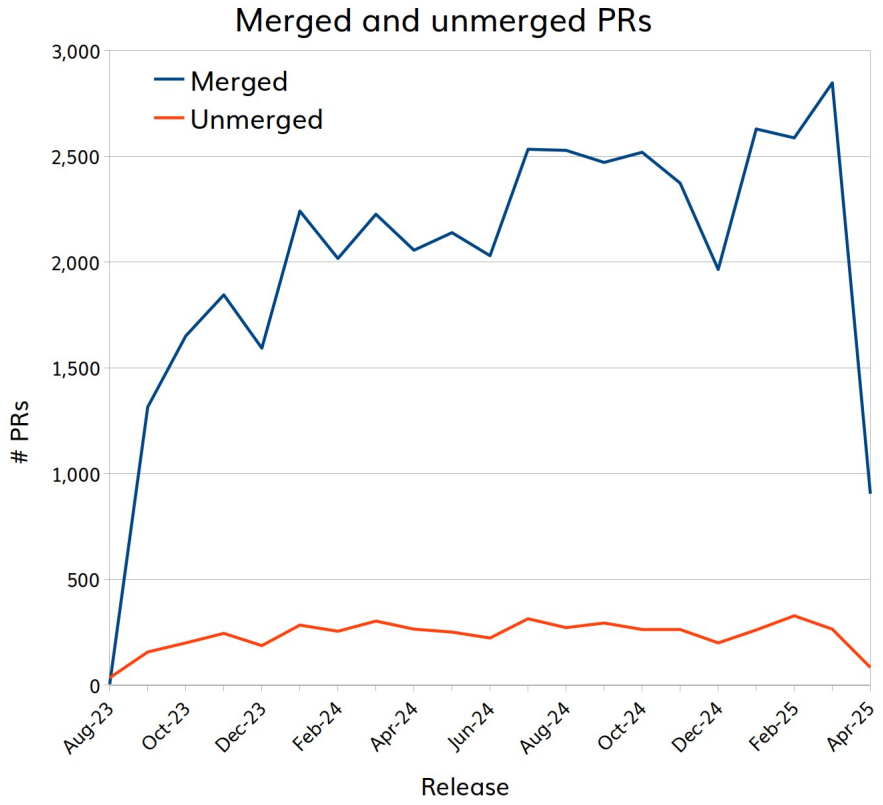
Analysing Pull Requests

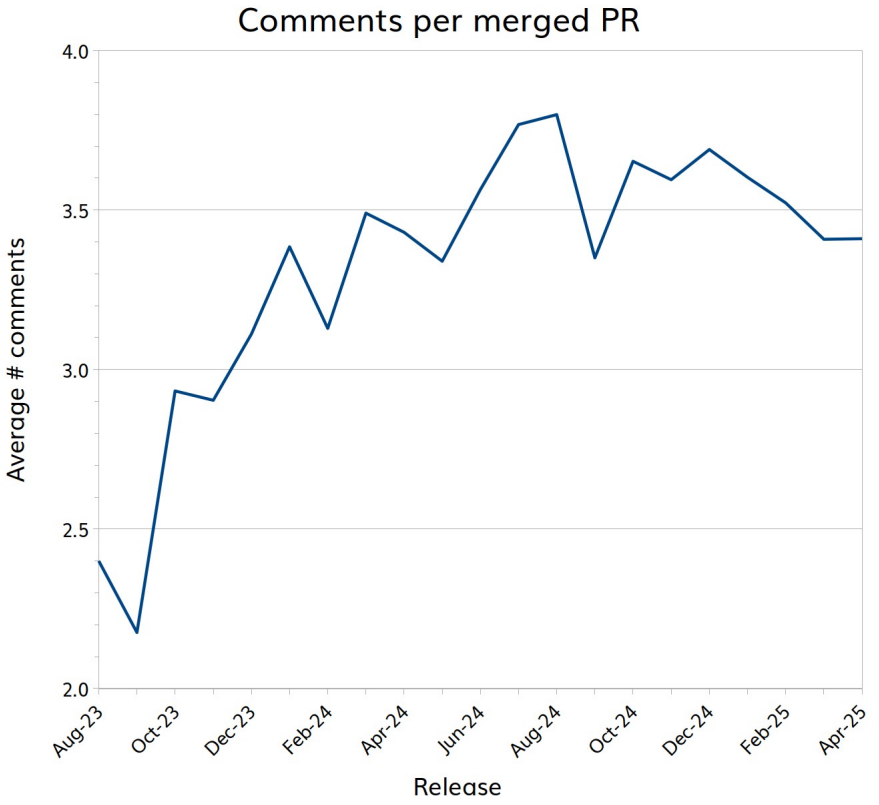
```
gh pr list -L 1000000 -s closed \
           --json id,createdAt,closedAt,state,comments \
           -q '.. | .id?,.createdAt?,.closedAt?,.state?'
```

Post process with an *awk* script to generate a CSV file.

CC BY SA

PRs: Comments per Commit





CC BY SA

More Detailed Analysis: Diversity and Inclusion

Copyright © 2025 Embecosm. Freely available under a Creative Commons Attribution-ShareAlike license.

The Inspiration

Prof Andrea Capiluppi
University of Groningen

- www.rug.nl/staff/a.capiluppi/
- Infer gender activity
  - how robust is this?
- Statistical analysis
- youtu.be/sTAtuSxwCss
- doi.org/10.1093/iwc/iwt047

CC BY SA

Thank You

jeremy.bennett@embecosm.com

github.com/embecosm/toolchain-analyze

embecosm.com

Jeremy Bennett

Copyright © 2025 Embecosm. Freely available under a Creative Commons Attribution-ShareAlike license.