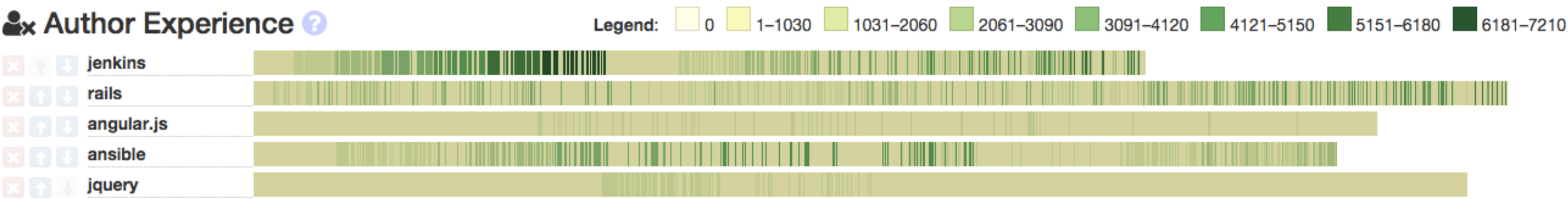


Comparing Repositories Visually with RepoGrams

http://repograms.net

Daniel Rozenberg, Ivan Beschastnikh, Fabian Kosmale,
Valerie Poser, Heiko Becker, Marc Palyart, Gail C. Murphy

University of British Columbia

Saarland University

Big (SE) data

GitHub

Bitbucket

- Millions of projects
- Open APIs
- Meticulously tracked and archived activity

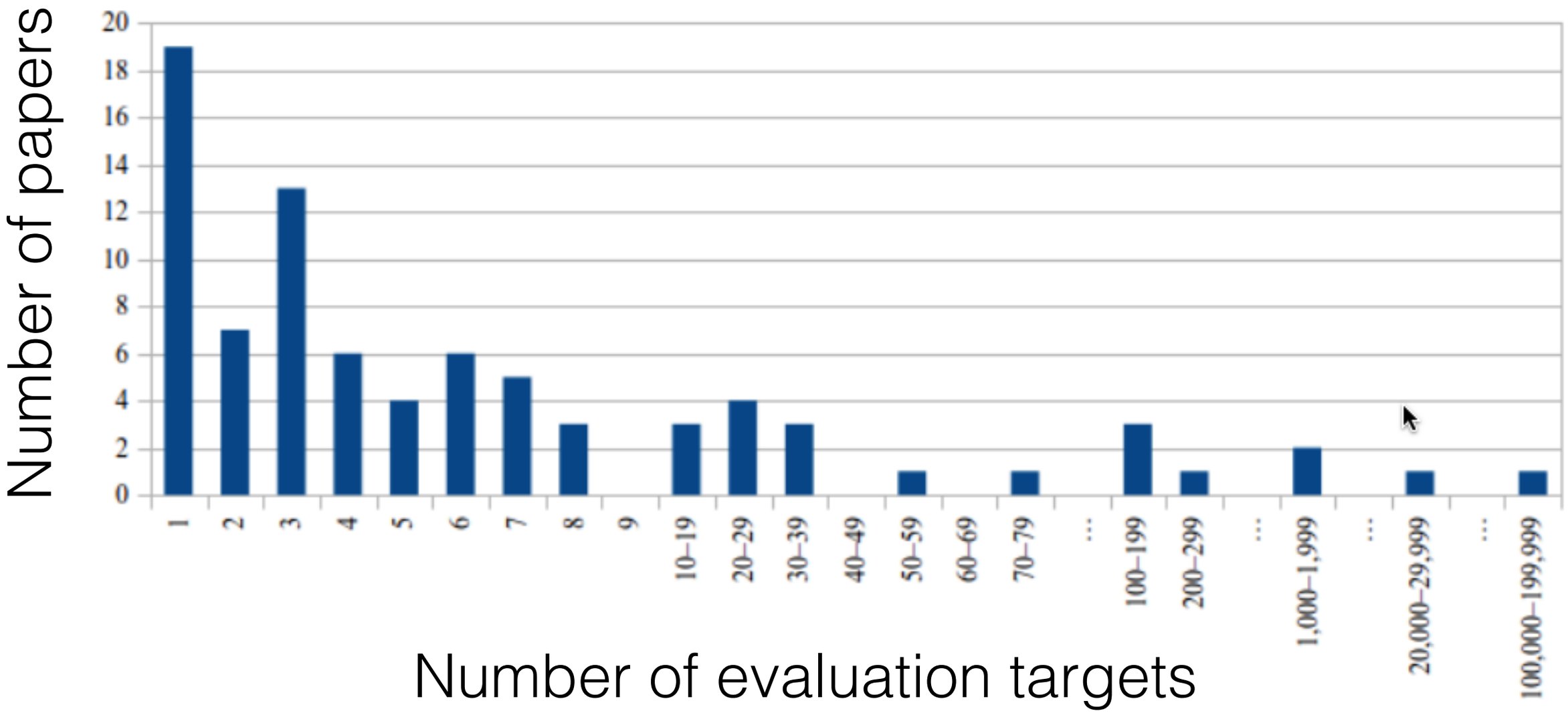
- Huge opportunity for researchers
- Each open source project is a potential evaluation target!

How many projects do paper authors use in their evaluation?

- **Experiment:** selected 114 papers from ICSE, FSE, ASE, MSR, ESEM (years 2012-2014)

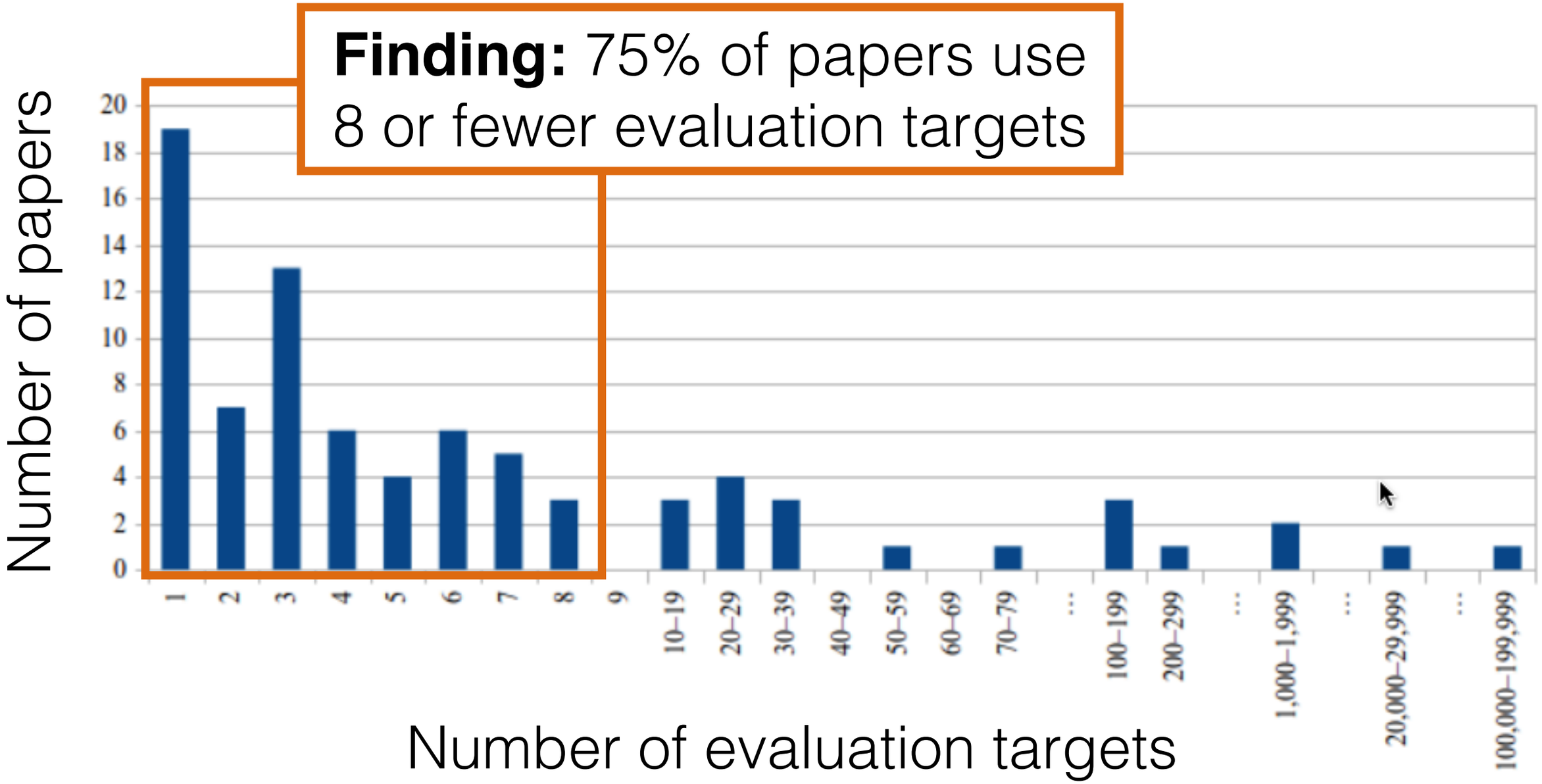
- Recorded number of targets that the authors claim to evaluate

How many projects do paper authors use in their evaluation?

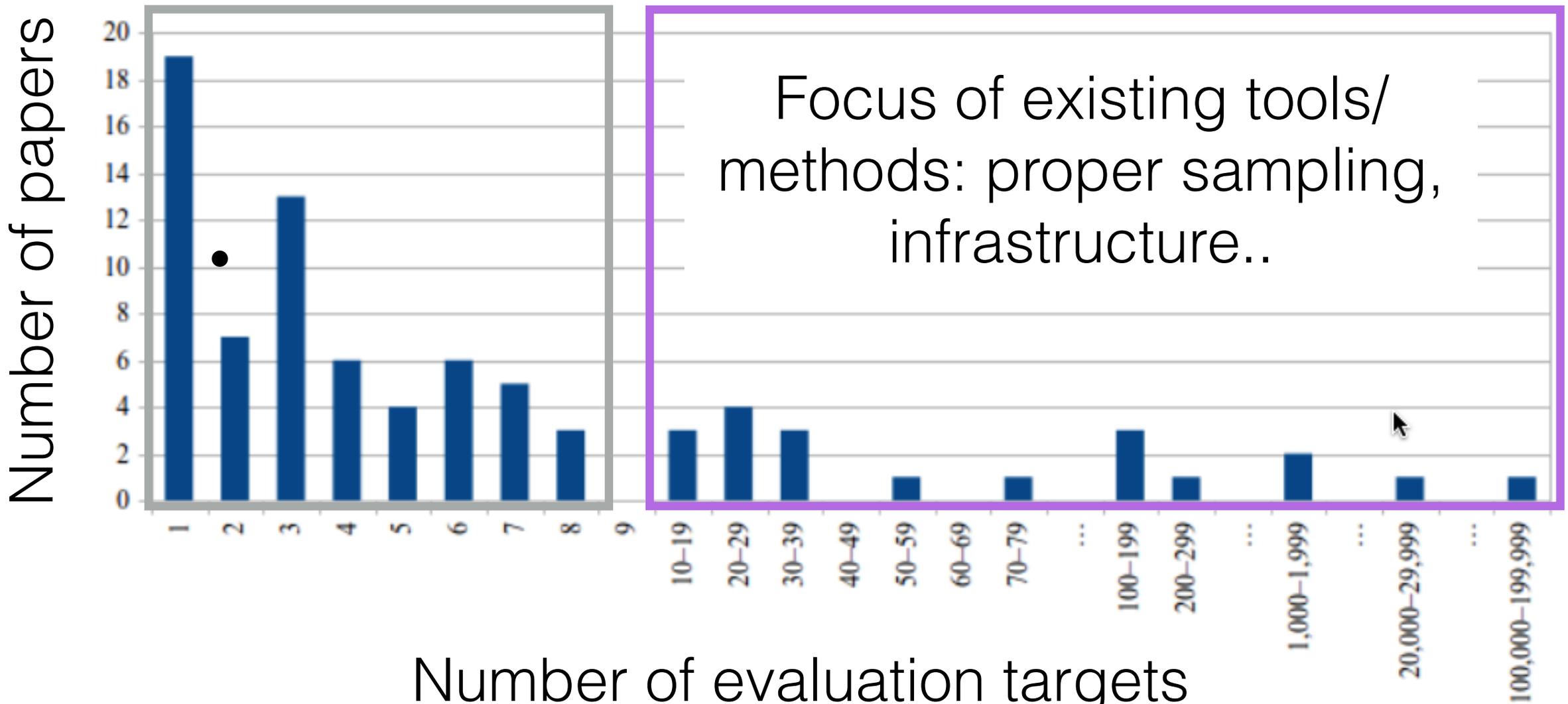


How many projects do paper authors use in their evaluation?

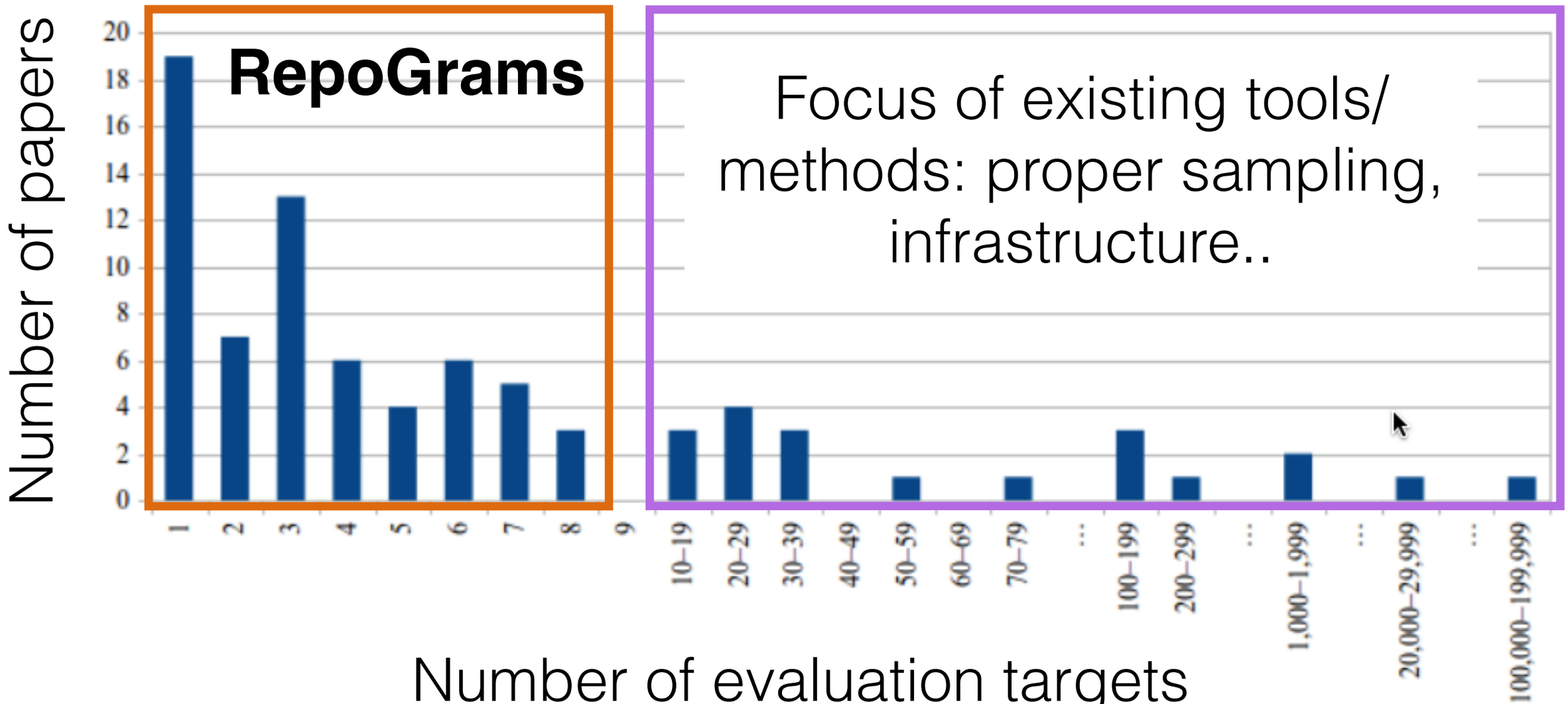
Finding: 75% of papers use 8 or fewer evaluation targets



Existing tools focus on supporting scalable analysis



Existing tools focus on supporting scalable analysis



RepoGrams:
Qualitative repository analysis

Presents data in a way that can be observed but not measured

RepoGrams:
Qualitative repository analysis

Presents data in a way that can be observed but not measured

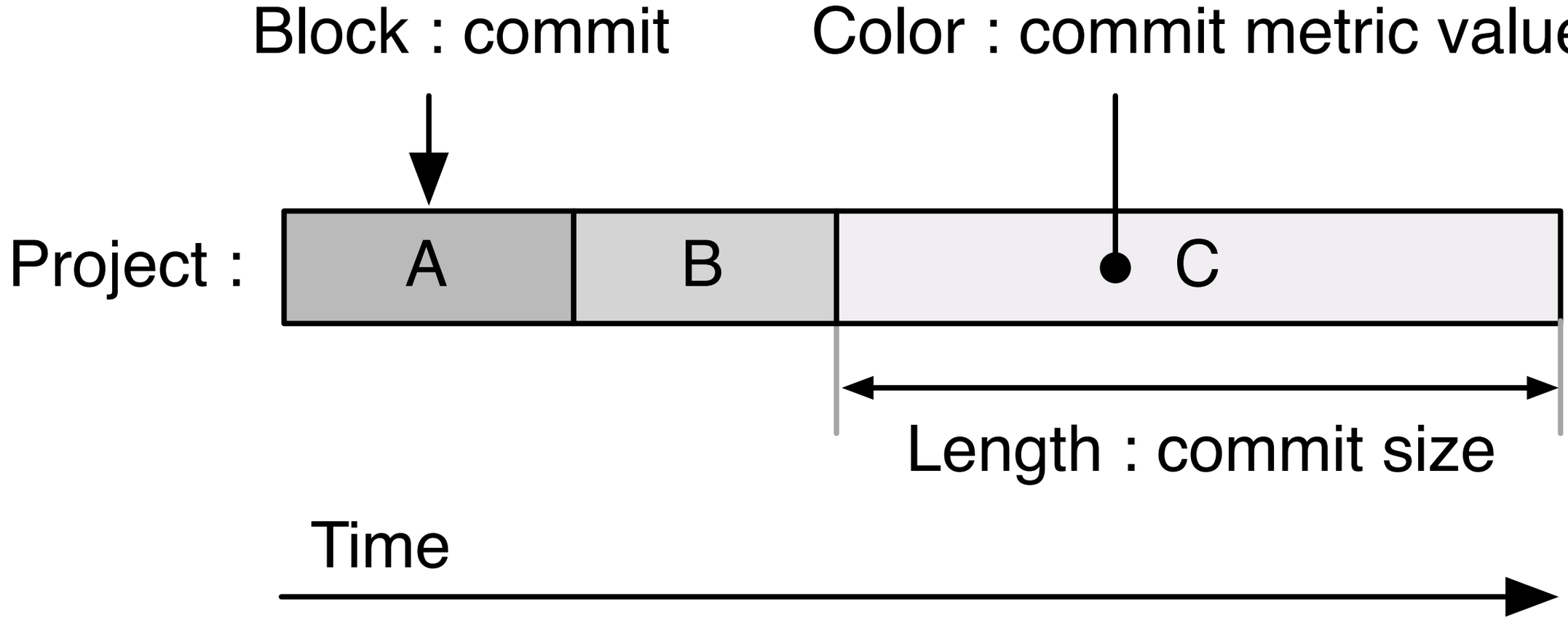
- Goal is not to provide an answer, but to surface relevant information

- Help the user think critically/contrast relevant features of a (small number of) projects

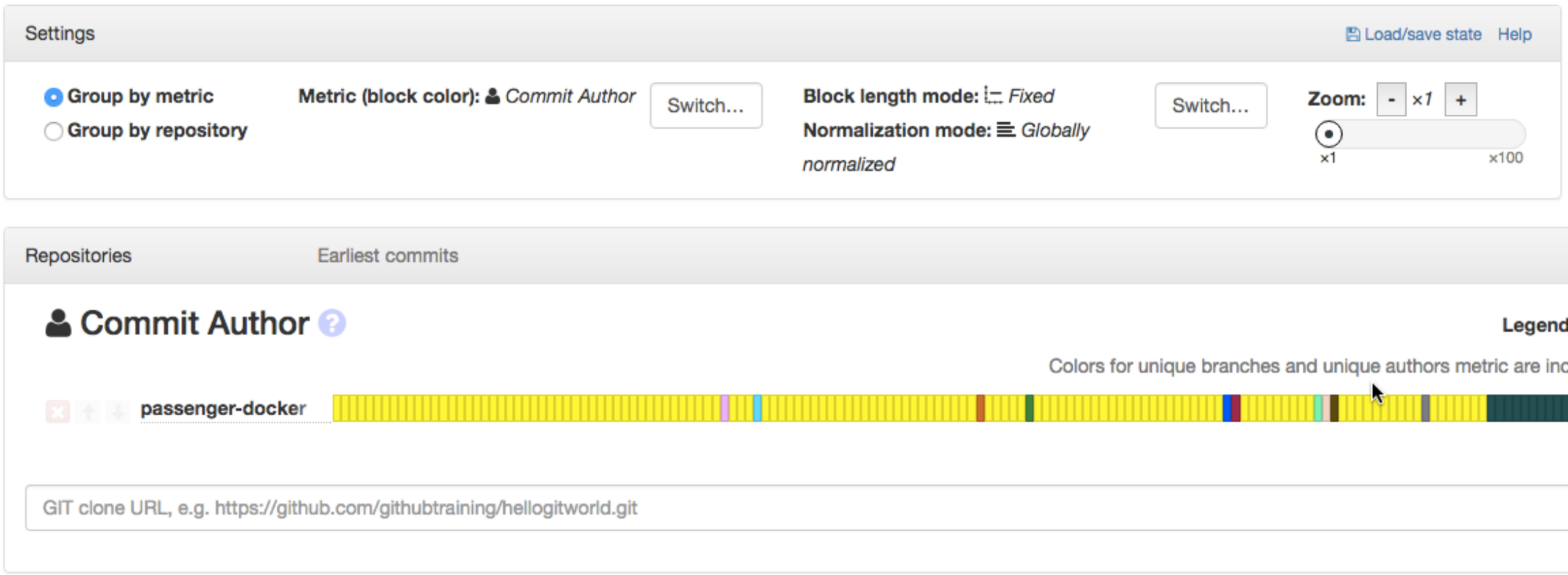
- Support curation of a small number of project ($\leq$ 8)

Visualization: a natural fit for qualitative analysis & nuance

Core abstraction in RepoGrams: Repository “footprint”



Demo: the basics

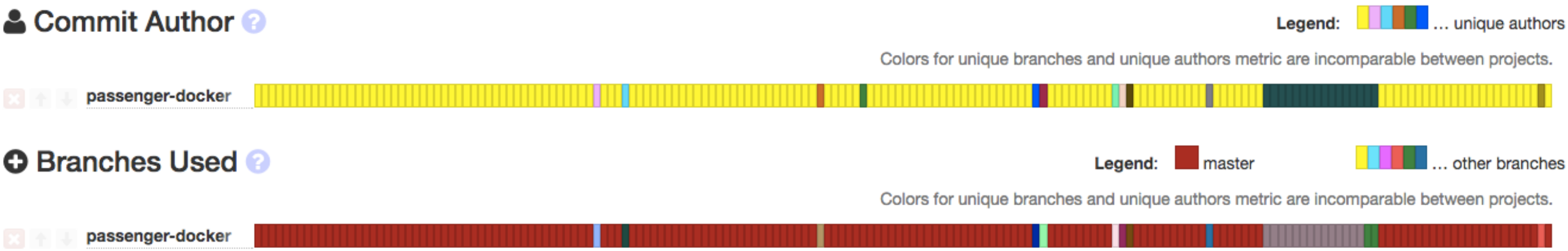


RepoGrams

Commit author metric: one unique color per author

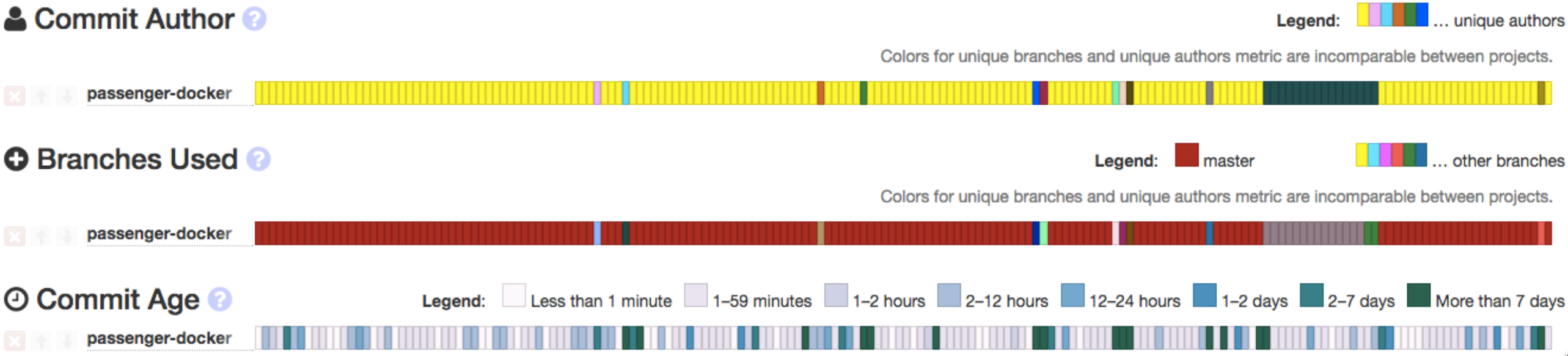
Constant commit block width

Demo: comparing two metrics



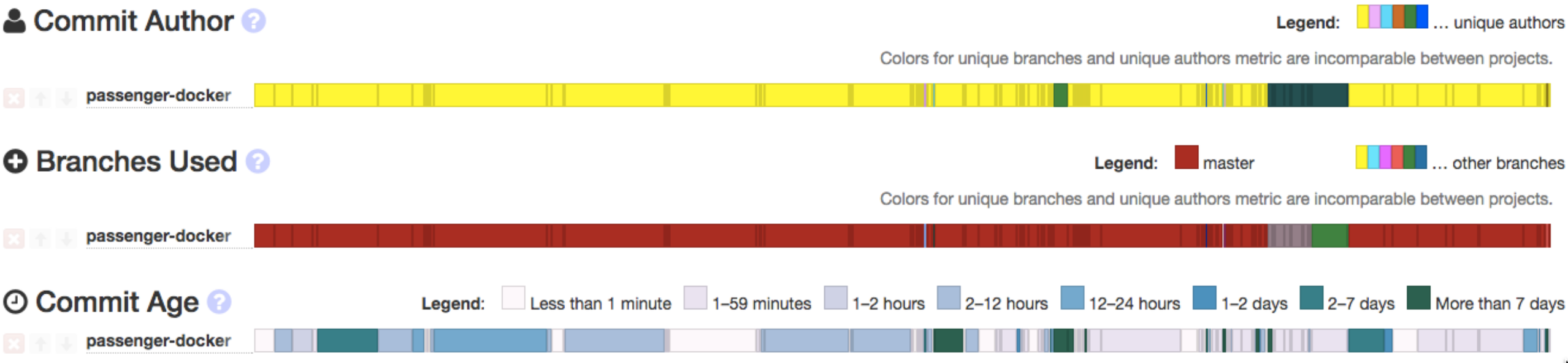
Branches used metric: one unique color per branch; master is always **red**

Demo: we can represent many things with a footprint



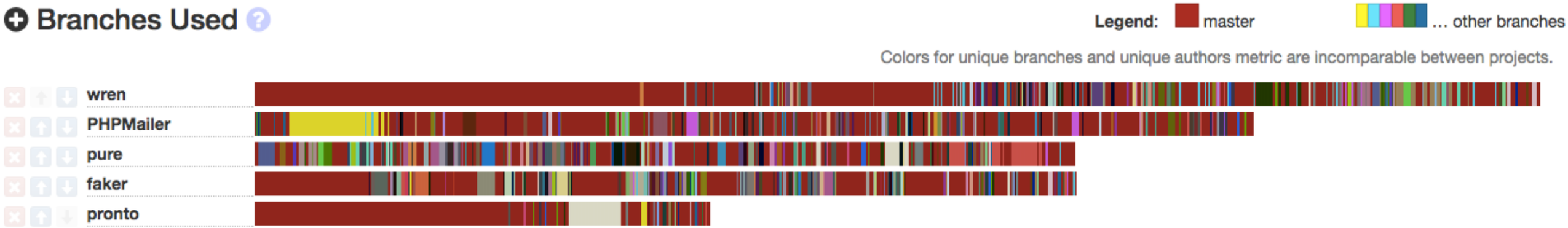
Commit age metric: elapsed time between commit and its parent

Demo: block width can denote magnitude of change



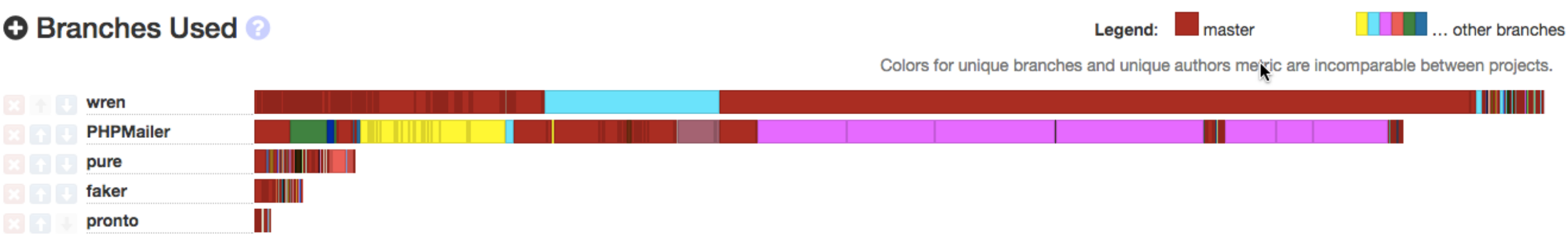
Block width: linear in the
LOC changed in commit

Demo: multiple projects



- *wren* has more commits than any other projects
- *wren, faker, pronto*, use master initially
- All projects eventually use a diversity of branches

Demo: multiple projects



- *wren* and *PHPMailer* have much larger commits
- PHPMailer has huge commits in the **purple** and **yellow** branches

Evaluation questions

RQ1: Can SE researchers use RepoGrams to understand and compare characteristics of a project's source repository?

RQ2: Will SE researchers consider using RepoGrams to select evaluation targets for experiments and case studies?

RQ3: How much effort is required to add metrics to RepoGrams?

Methodology

RQ1: Can SE researchers use RepoGrams to understand and compare characteristics of a project's source repository?

RQ2: Will SE researchers consider using RepoGrams to select evaluation targets for experiments and case studies?

- 14 authors from MSR'14
- Tasks using RepoGrams
- Semi-struct. interviews

RQ3: How much effort is required to add metrics to RepoGrams?

- 2 developers
- Each implemented 3 metrics

Evaluation highlights

RQ1: Can SE researchers use RepoGrams to understand and compare characteristics of a project's source repository?

- Successfully used RepoGrams for complex tasks
- Tools is of immediate use

RQ2: Will SE researchers consider using RepoGrams to select evaluation targets for experiments and case studies?

- Researchers want custom metrics

RQ3: How much effort is required to add metrics to RepoGrams?

- Setup: 1.5 hours
- Metric: avg/max = 40/52 min
- < 40 LOC total

Related work

- **Helping researchers with the selection process**
  - *Tools/Datasets:* GHTorrent, Boa, MetricMiner
  - *Methods:* “Diversity in software engineering research”, FSE13
- **Visualization**
  - *Tools:* CVSgrab, ConcernLines, Fractal Figures, Chronos, RelVis, Chronia, Evolution radar

- Lots of data, many potential evaluation targets!
- But, proper **project selection is complex**
  - Researcher must be highly aware of the features of the project that may influence the study results

- **RepoGrams: supports qualitative analysis of software repositories**
  - Presents data in a way that can be observed but not measured

Try our public deployment! **http://repograms.net**