



IDENTIFYING DEVELOPER PERSONAS COULD CREATE BETTER RESEARCH SOFTWARE DEVELOPMENT

Are YOU an RS-10X? Mining GitHub to Describe Developer/Repository Interaction Types

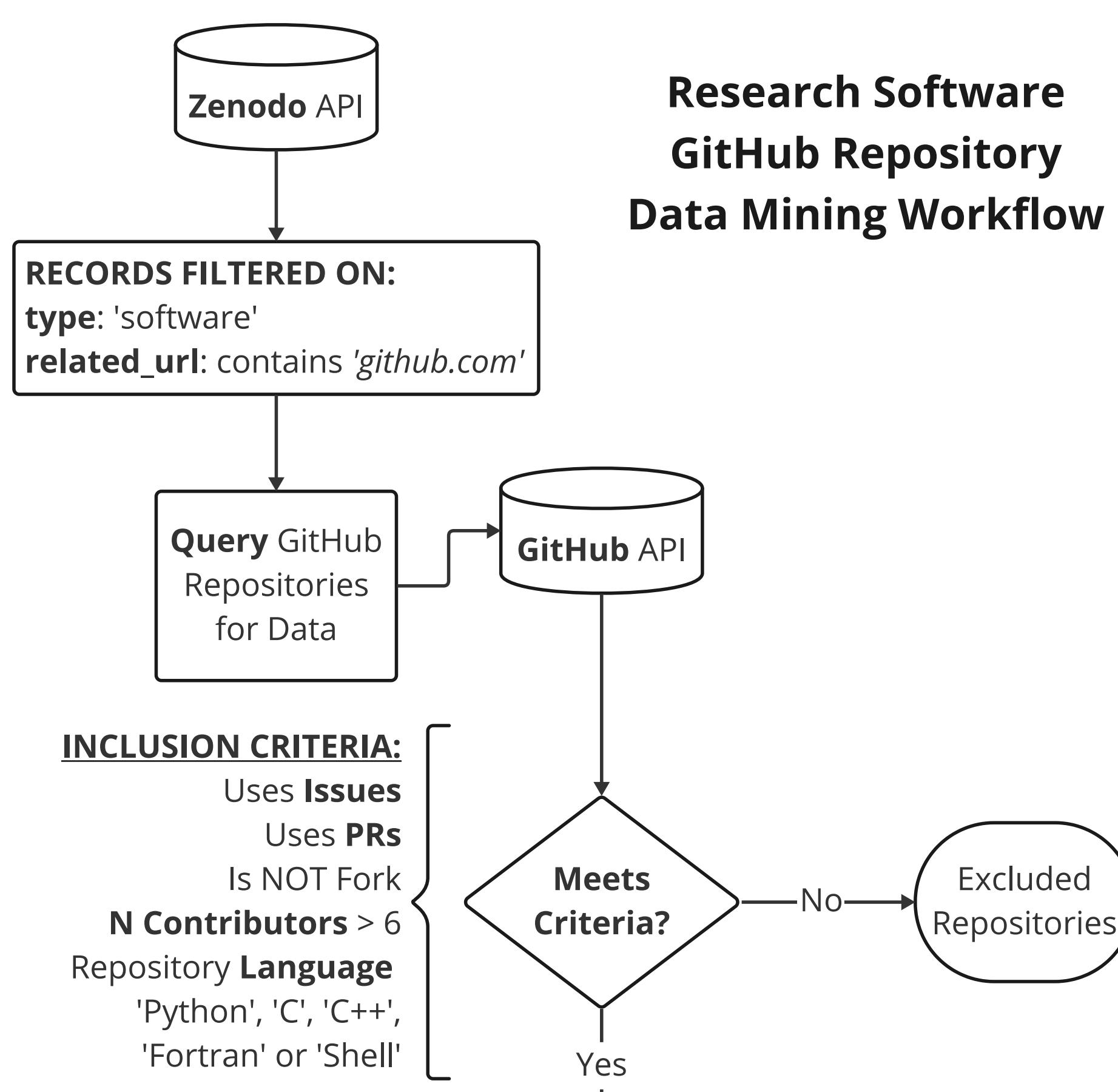
Felicity 'Flic' Anderson[†], Dr. Julien Sindt[†] & Prof. Neil Chue Hong[†] [†]EPCC, University of Edinburgh

0: Identifying the Habits of Highly Effective Developers

This pilot study attempts to identify 'superstar developers' within 10 larger RS repositories by exploring assignment and contributions data.

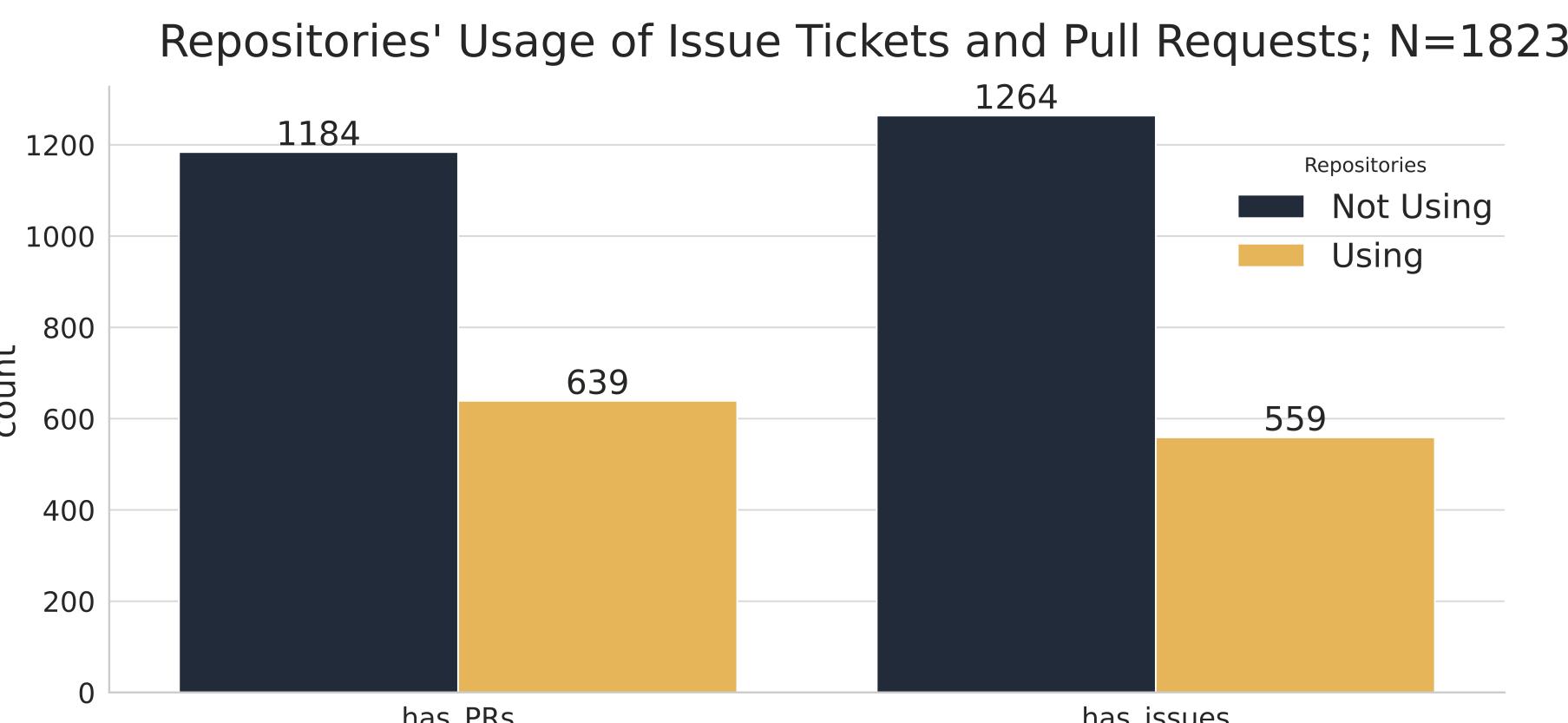
- Assignment to GitHub features **Issue Tickets and Pull Requests** (PRs) correlate with commit contributions.
- Individually limited metrics can be combined to give better picture of development responsibilities or activity.
- Can this approach locate distinct clusters of behaviours and define and describe RS developer personas? Can they predict effectiveness?**

1: Mining RS Repository Data To Understand Dev/Repo Interactions



Research Software (RS) is any software used to generate research, across academic fields.

Public GitHub projects with a DOI (often required for publication) across all fields have been used.



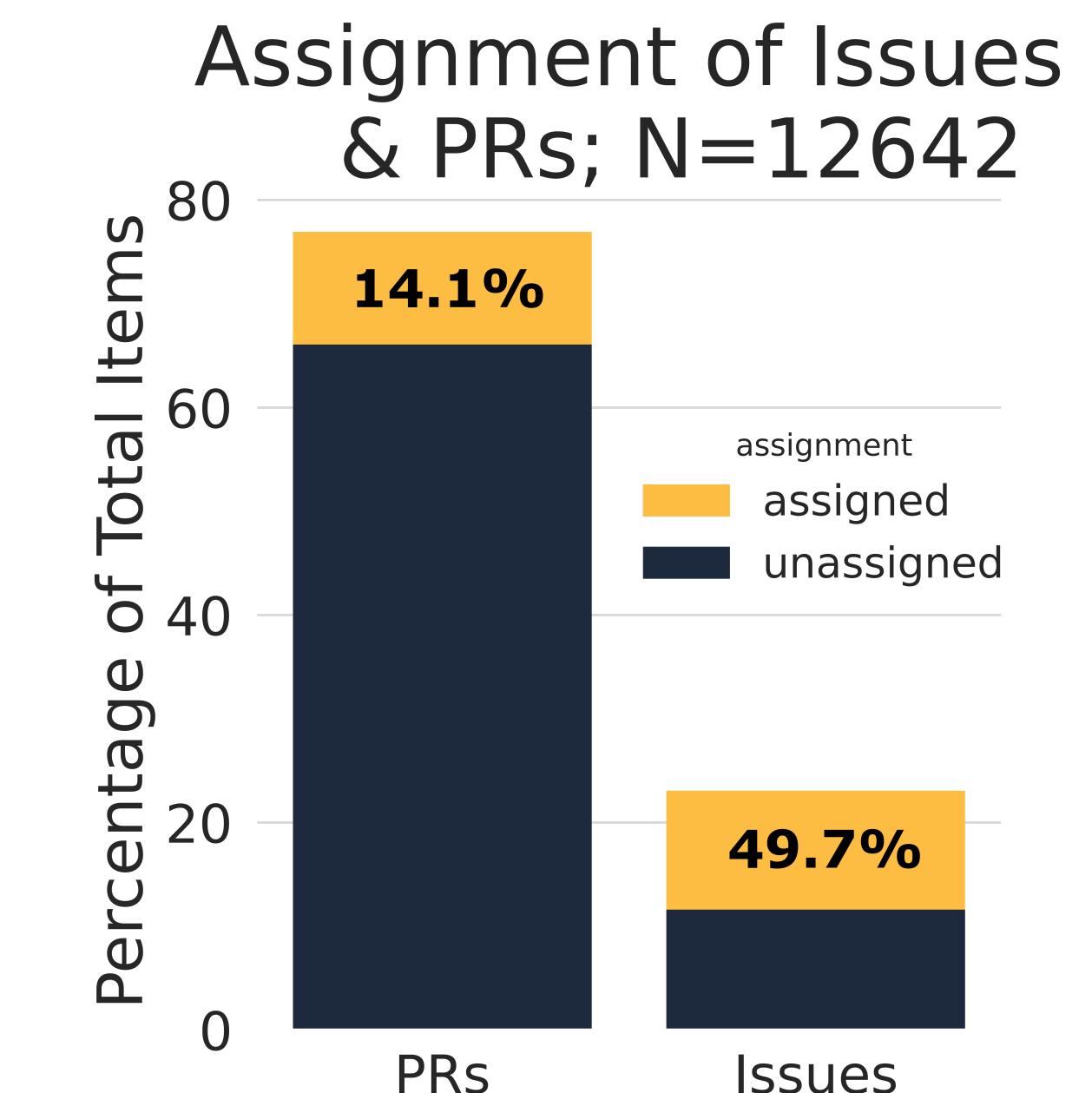
Of 1823 repositories identified from Zenodo, only 31% use Issue Tickets or Pull Requests (35%).

Assignment and Commit data were gathered from **10 larger repositories** to form a pilot study dataset.

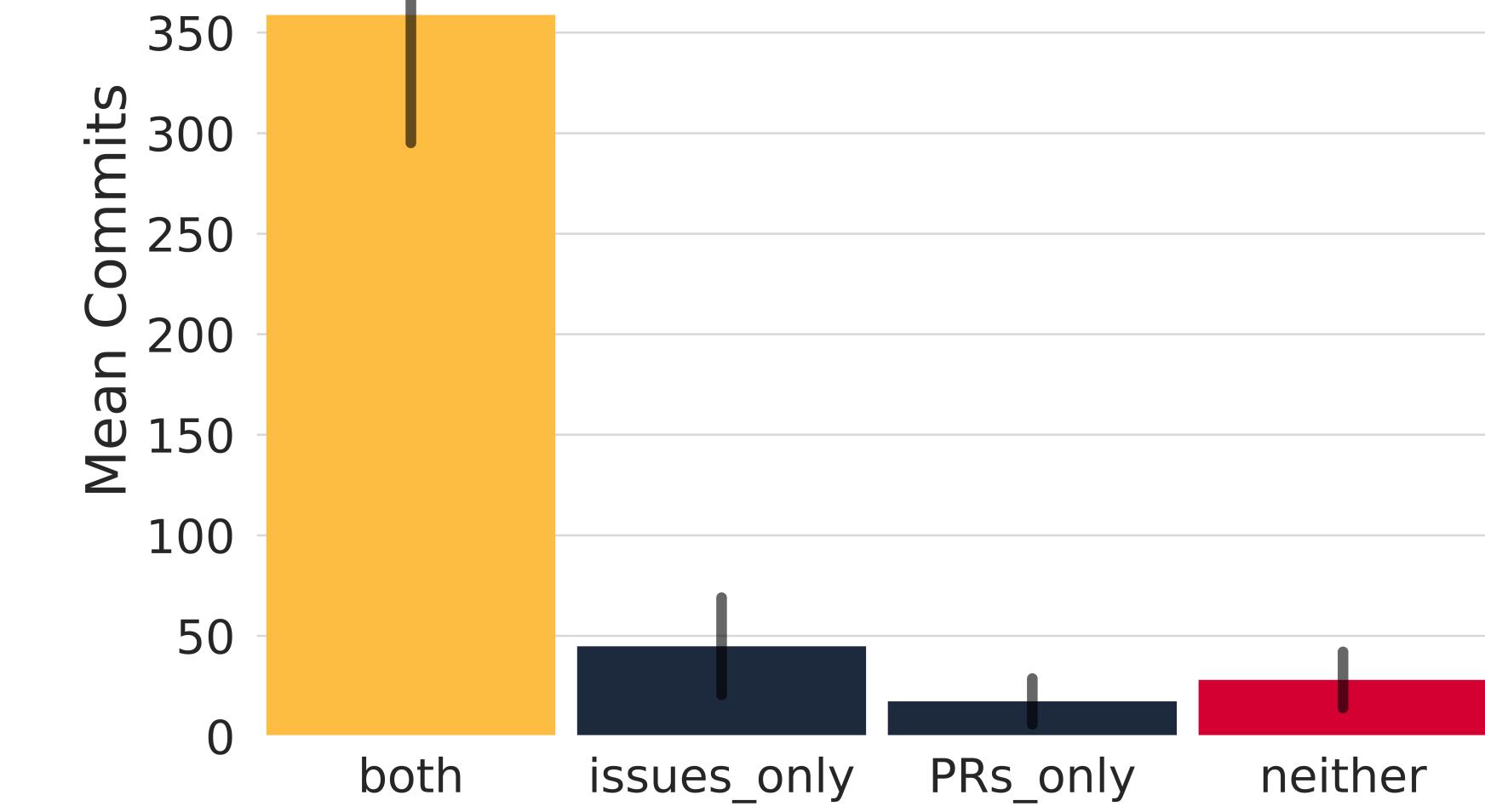
Developers were grouped by '**assignment categories**', defined as *being assigned to 1+ Issue and/or 1+ PR*

2: More Responsibility = More Action?

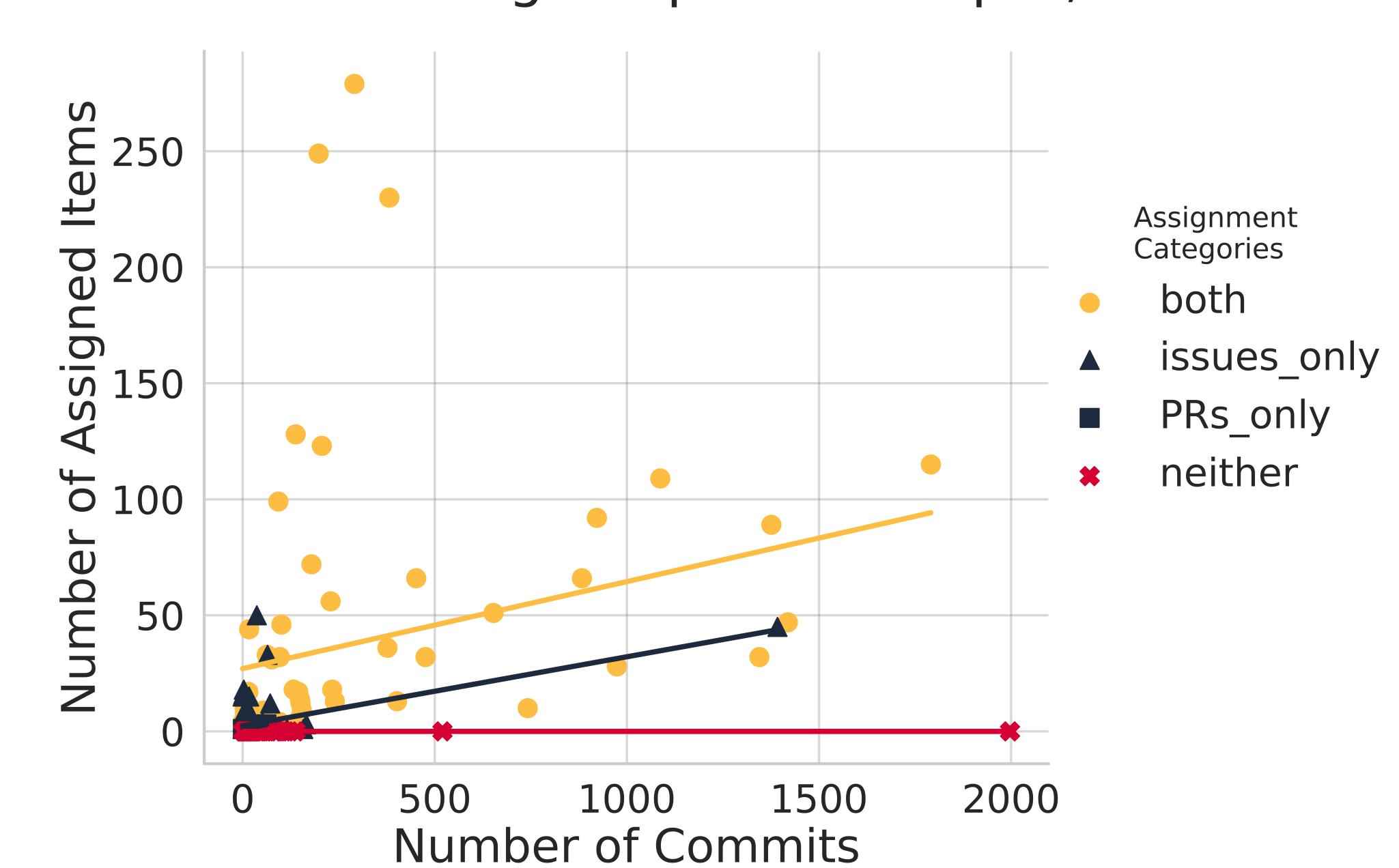
Assignment of Issues & PRs; N=12642



Average Commits per Developer by Assignment (+SE); N=259



Correlation Between Number of Commits & Total Items Assigned per Developer; N=259



Commit numbers positively correlate with number of items assigned. Developers assigned to 'BOTH' item types are a minority but have ~10% higher commit numbers than all other assignment groups.

Are 'highly interactive' developers potential 'superstars' compared to less-committed developers?

3: Using RS Developer Personas in Future Work

Ask Me About Limitations... Sample potentially skews towards 'best practice'; assignment vs activity; not all commits are equal; developer metrics problems; project comparison difficulties; capturing practice changes over time; correlation vs causation.

Future Work:

- Using **additional properties** and **clustering analysis** to fully describe Personas
- Linking Personas to technical **software quality metrics** via static analysis
- Distribution of Personas** within project teams
- Comparing developer Personas to software effectiveness** (*popularity, usage, citations*)