



IDENTIFYING DEVELOPER PERSONAS COULD CREATE BETTER RESEARCH SOFTWARE (RS) DEVELOPMENT

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

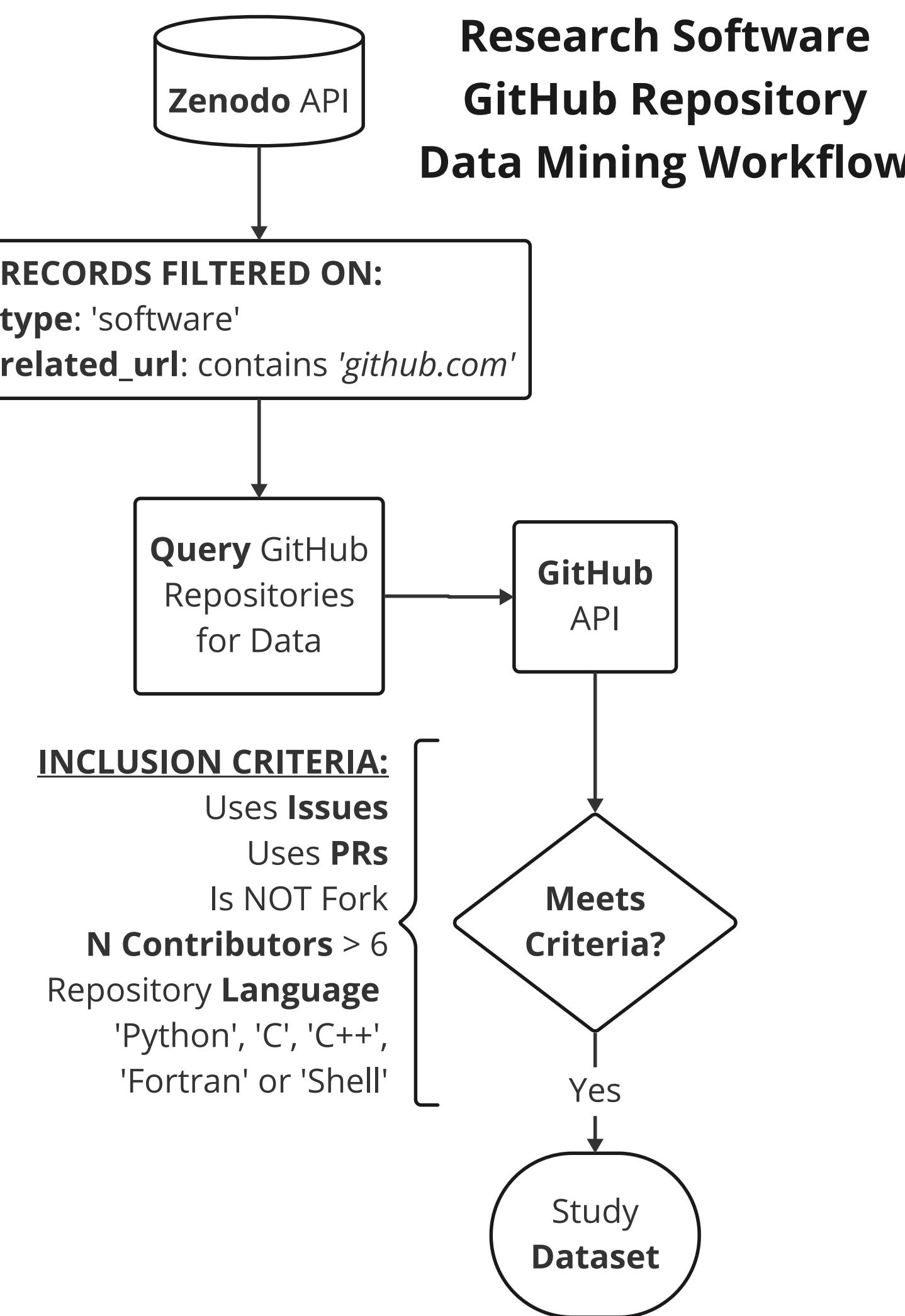
Felicity 'Flic' Anderson[†], Dr. Julien Sintt[†] & Prof. Neil Chue Hong[†] [†]EPCC, University of Edinburgh

0: Identifying the Habits of Highly Effective Developers

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

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

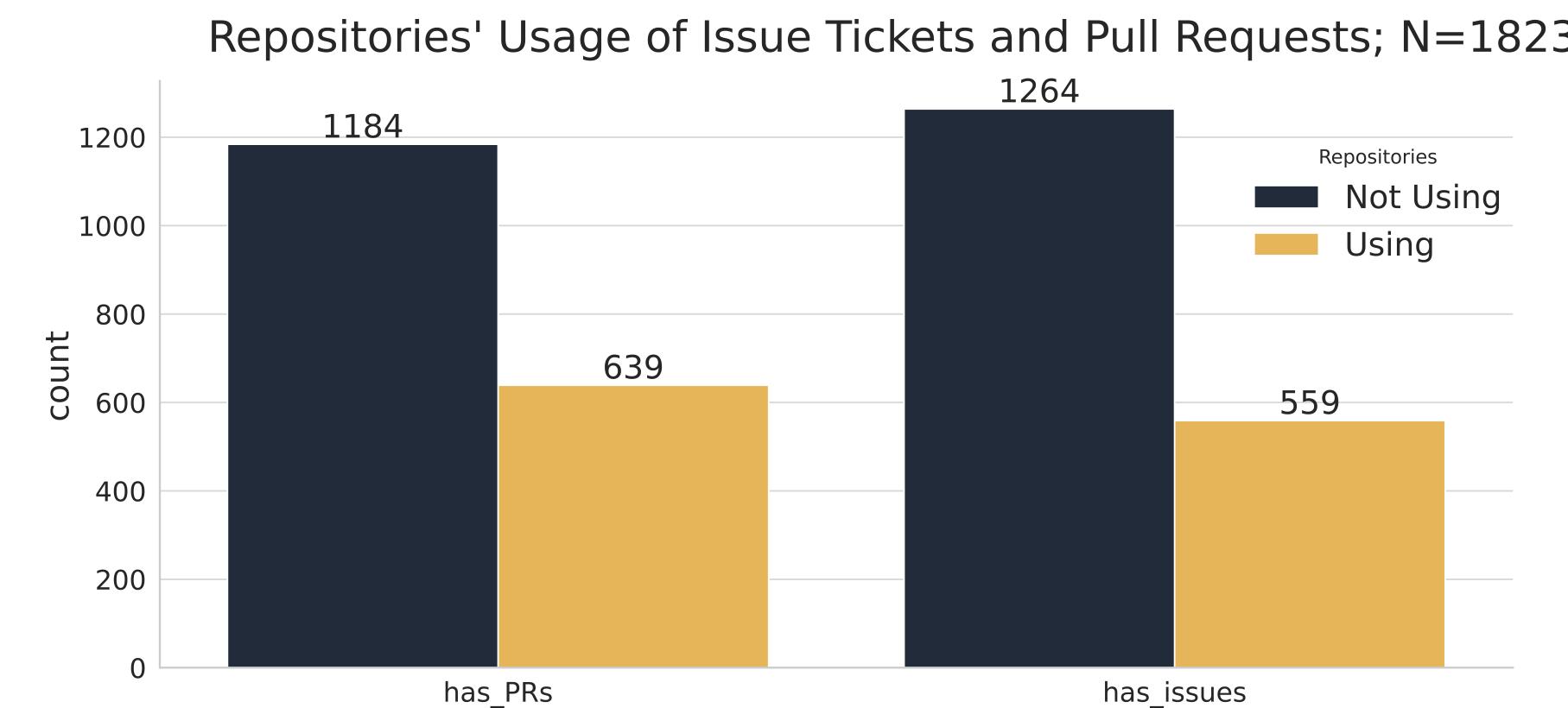
1: Mining RS Repository Data To Understand Dev/Repo Interactions



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

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

Zenodo is a multidisciplinary open repository for depositing research outputs and minting DOIs.



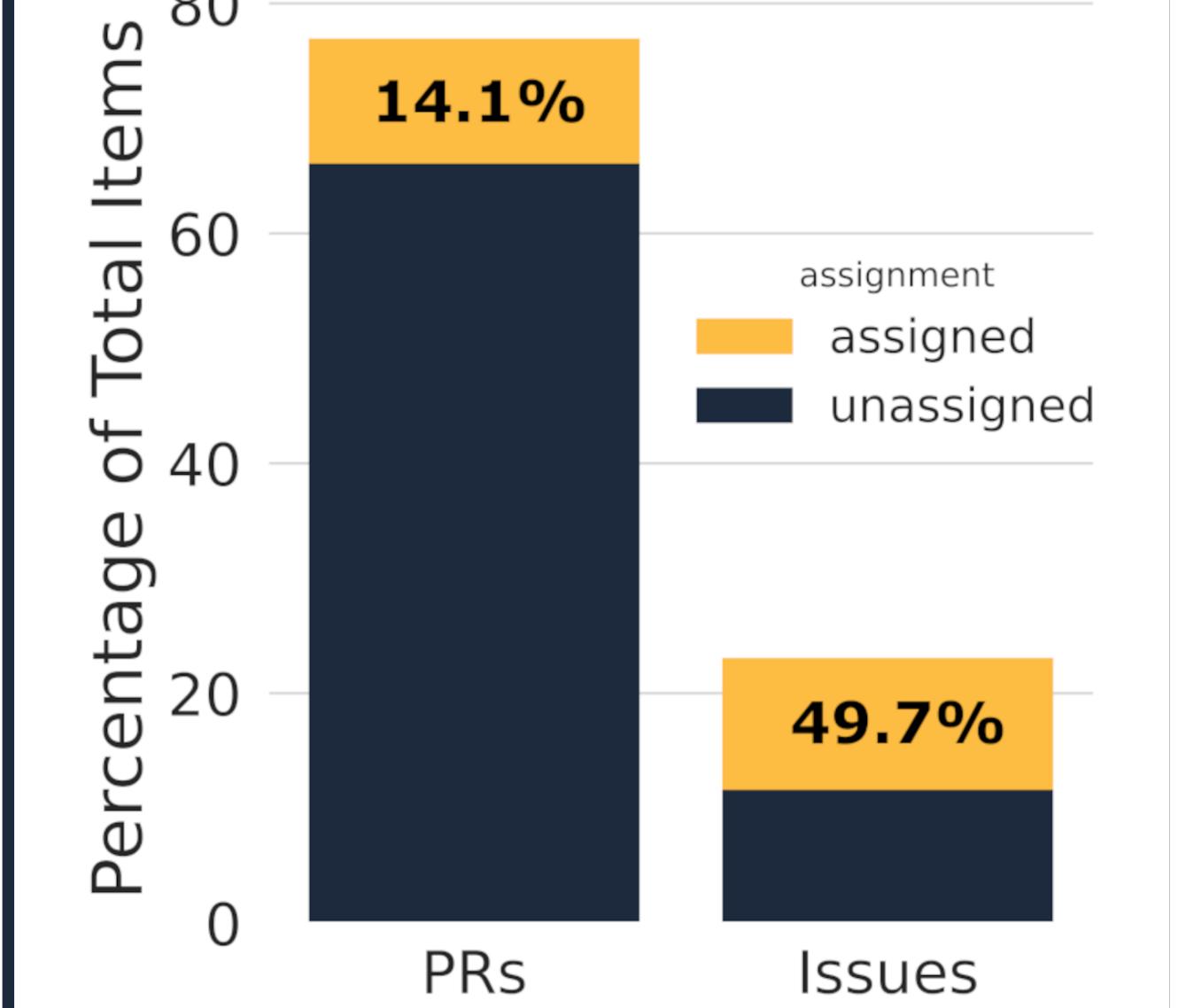
Of 1823 repositories identified from Zenodo, only 31% use Issue Tickets or Pull Requests (35%).

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

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

2: More Responsibility = More Action?

Assignment of Issues & PRs; N=12642



Category	N Devs	Devs (%)
both	48	18.5
issues_only	58	22.4
PRs_only	5	0.02
neither	148	57.1

Assignment is uncommon even within repositories using Issue Tickets and Pull Requests.

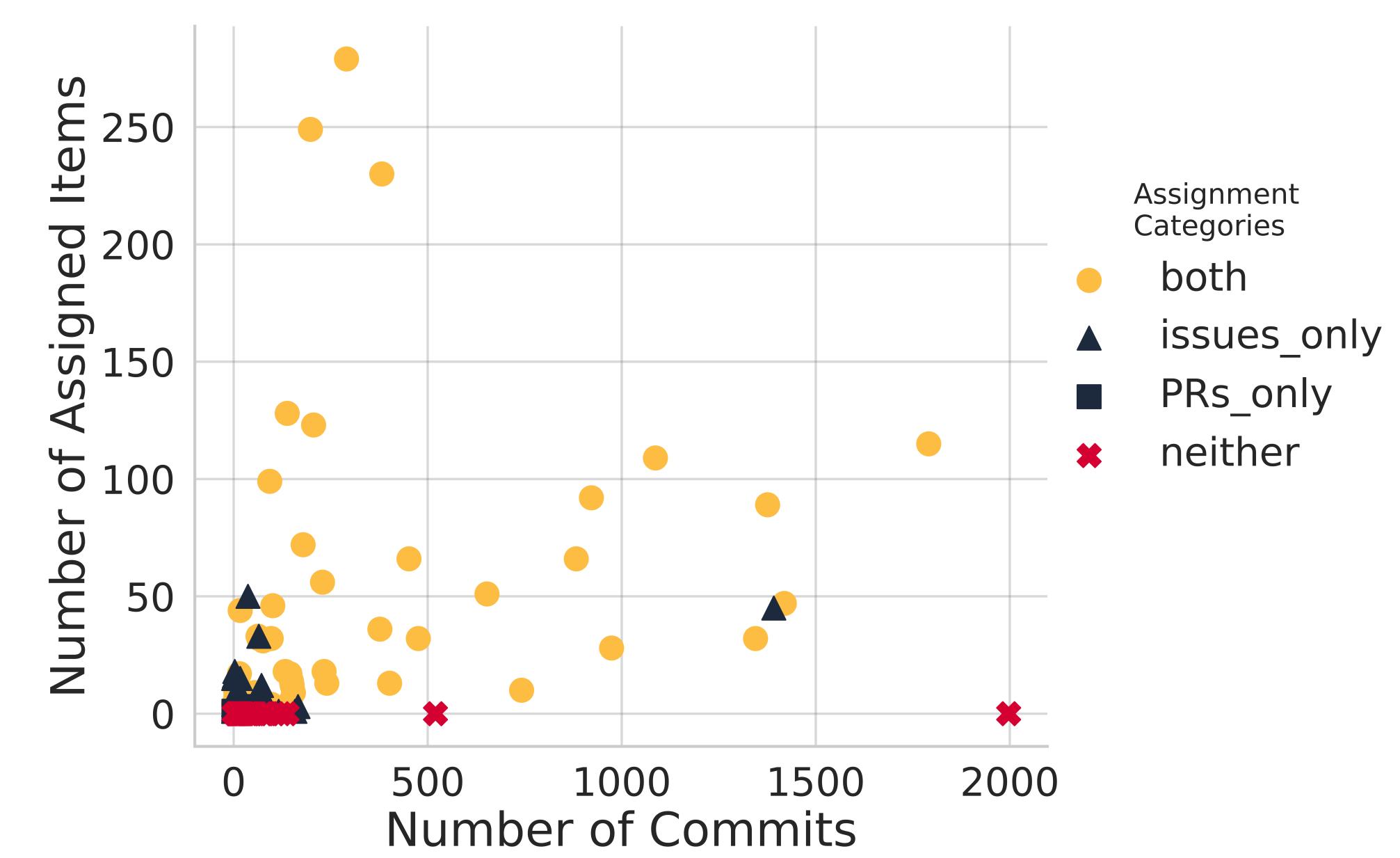
Commit numbers **positively correlate** with number of items assigned.

Developers assigned to 'both' item types are a minority but have **~10% higher commit numbers** than other assignment groups.

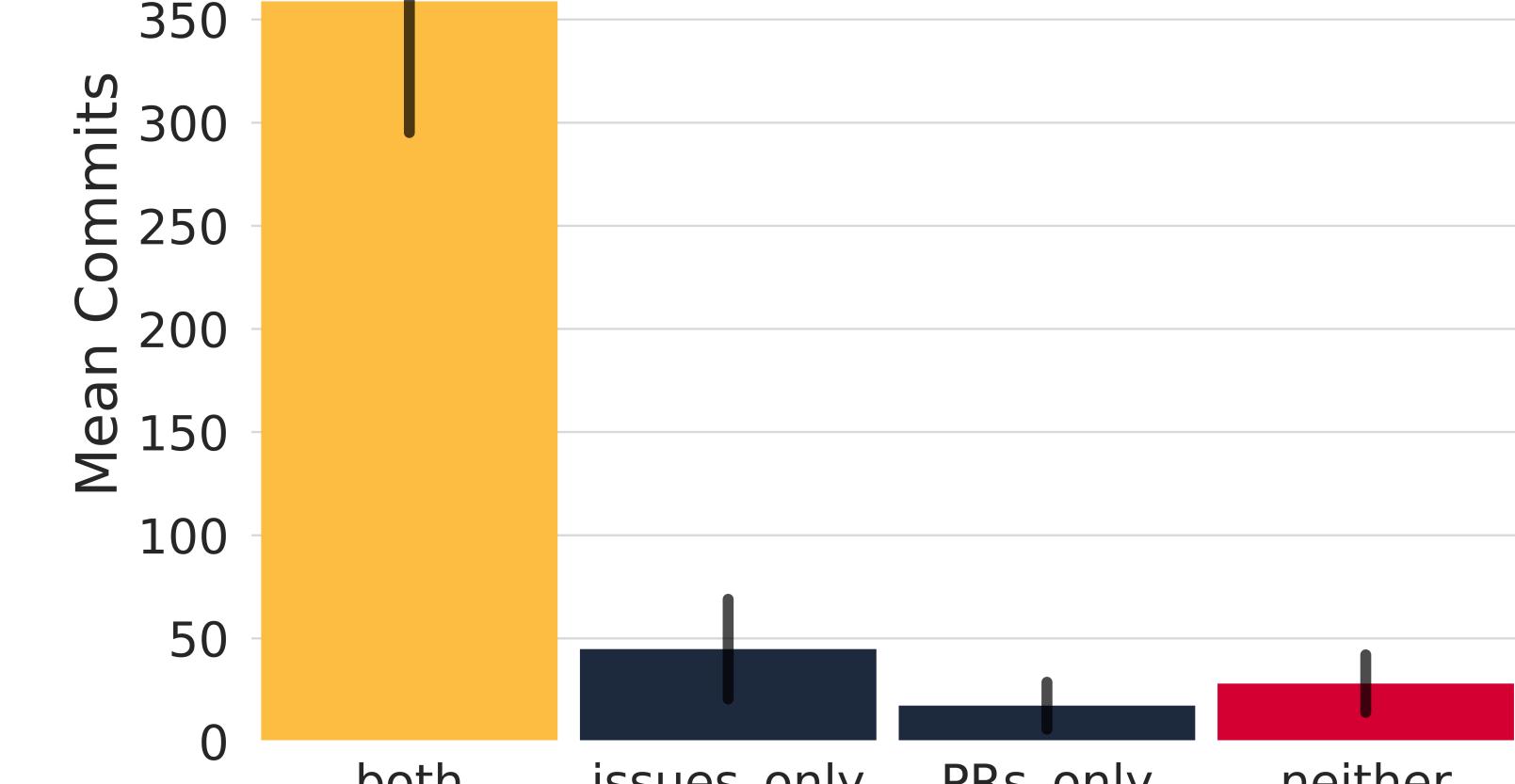
Do 'both'-type developers split into **two trends**?

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

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



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



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*)