

Do programmers follow better practices when others are watching?

Mengchen (Sam) Yong, Professor Bogdan Vasilescu, Carnegie Mellon University

Background

- People behave better in public.
- Programmers are often required to write clean code and good comments in the industry.
- Many open source projects also have contributor guidelines.

Research Question

Do programmers behave *differently* in varying open source settings?

- In highly visible, large open source projects vs. personal projects,
- More and better code comments?
- Detailed and structured commit messages?

And, how do we measure these behavioral differences and compare and interpret them?

Method

We selected 376 open source developers on GitHub. For each of them we selected repositories where they have contributed at least 60 commits.

Method

These include repositories they own and some large community projects. We focused on repositories where the primary source code is *Python*.

We aggregated several statistics for a developer's commits in their selected repositories, in particular

- ✓ *number of lines of comments in each commit*
- ✓ *added comments to total code change ratio in each commit*
- ✓ *commit message word count*

We compared these statistics across the selected repositories for a developer, using the *number of stars* as an indicator of repository visibility, and generated plots to understand the aggregated data.

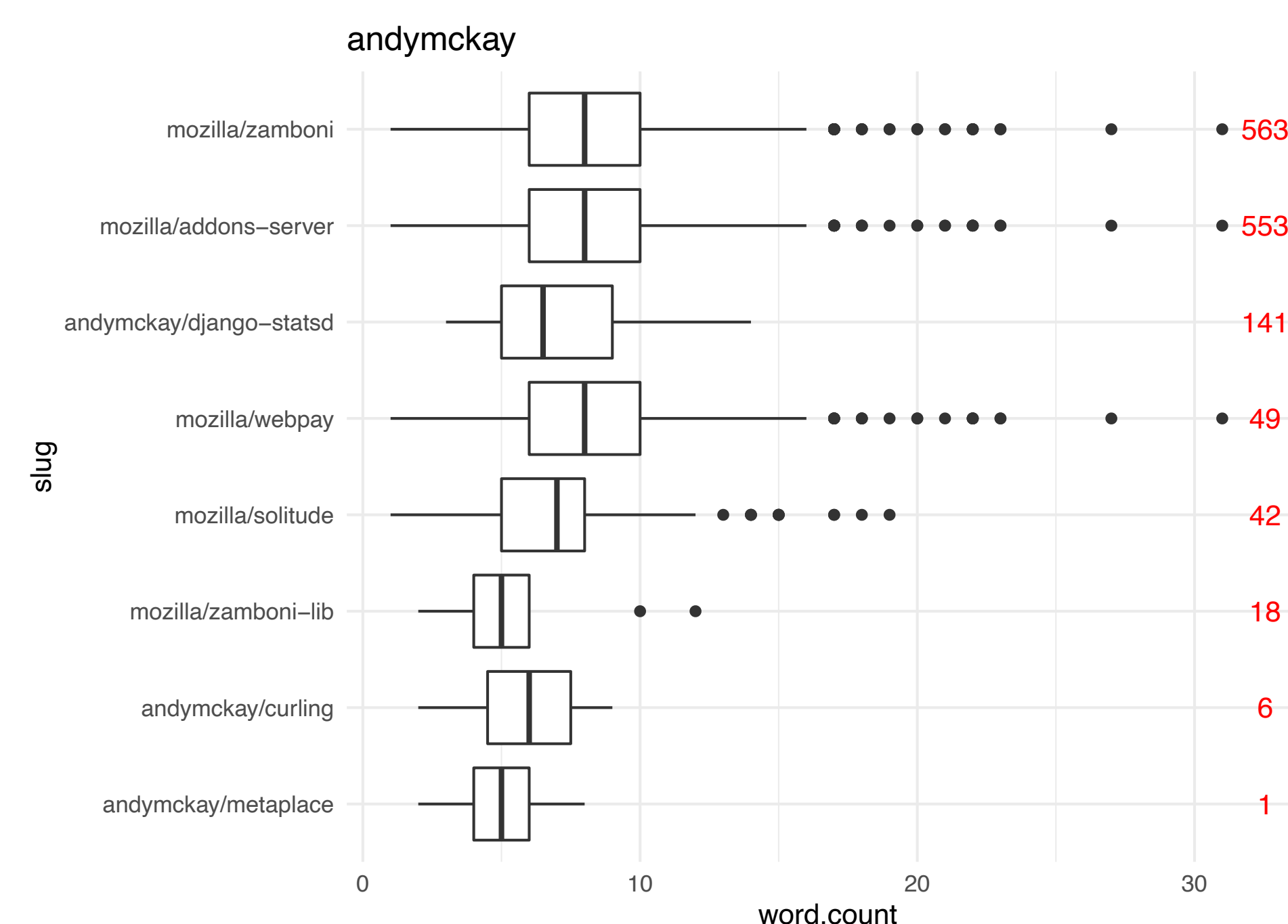


figure 1 user andymckay's commit message word count, sorted by repository star count

Results

Unfortunately, we *did not* find noticeable and consistent trends. The plots for some developers show interesting pattern (e.g. figure 1), but they are quite rare. Most other data sets look random (e.g. figure 2).

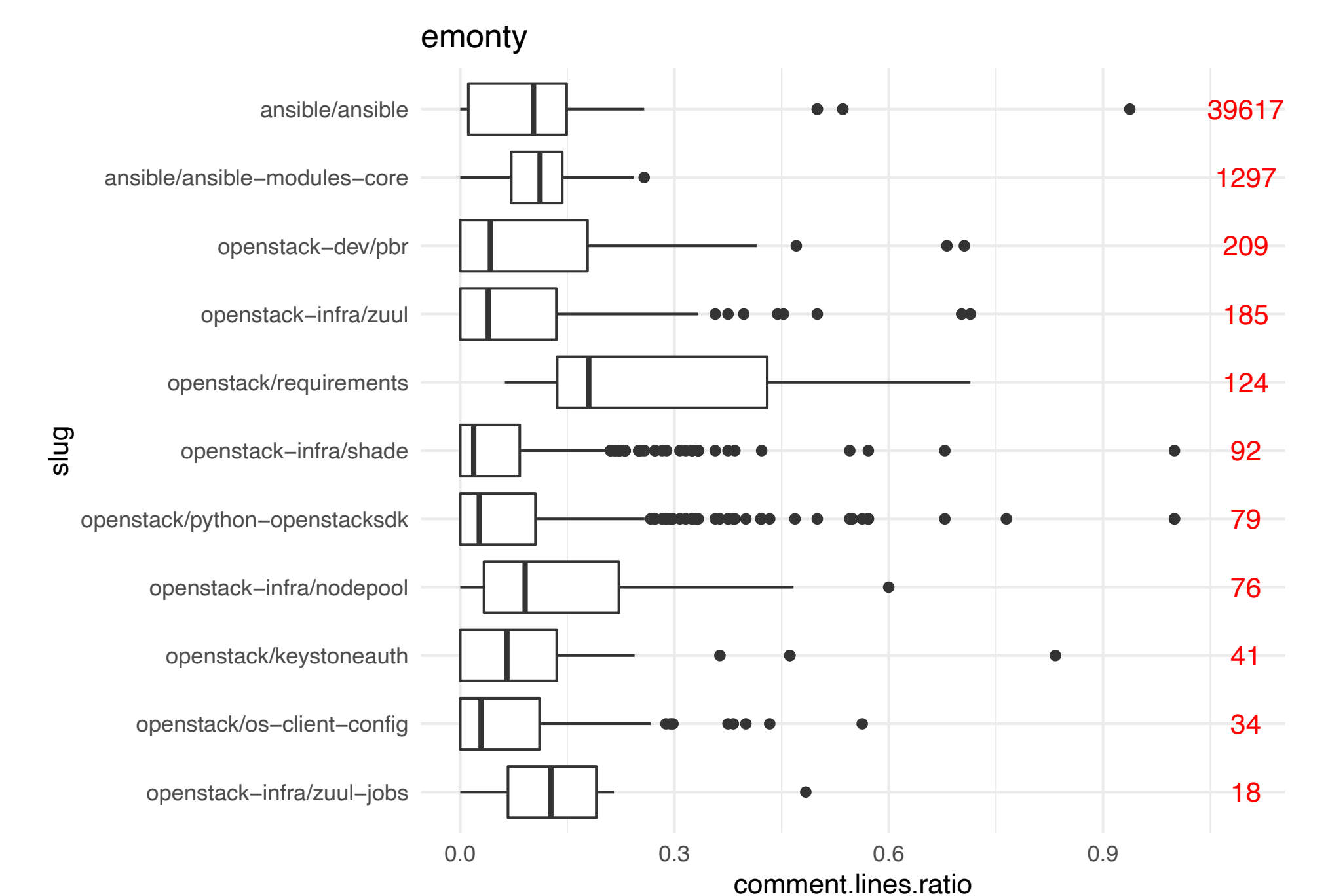


figure 2 user emonty's added comments to total code change ratio, sorted by repository star count

Future Work

Although we do not see a clear relation from the data we collected, the question remains a very complicated one. We are considering other variables at play, e.g.

- Repository visibility changing over time and how it affects developer behavior
- Developer's education and industry background and how it affects their behavior consistency across settings