

A Large Scale Study of Programming Languages and Code Quality in Github

Summary

The authors wanted to look programming languages' effect on code quality. The measurement is through looking at bugs. The authors looked at a very large dataset from Github by looking at most popular repositories (by number of stars) from Github across 17 languages. By using statistical methods, they were able to conclude that properties of programming language do have a small effect on code quality. But because of the number of variables and the small effect, nothing significant can be concluded. However, it is clear by the results that some type of errors occur more in some programming languages than others. And the domain of the project does not show relation with code quality.

Things I would like to see discussed:

- What are some other factors that contribute to code quality.
- The result from this paper seems to be different from a small controlled experiment. What could cause these differences.
- The data came only from the numbers and the numbers directly relates to the project chosen. Are there any larger scale analysis on this topic?