

Pontifícia Universidade Católica de Minas Gerais

Curso: Engenharia de Software

Matéria: Laboratório de Experimentação de Software

Características de Repositórios Populares

Relatório Técnico

Autores:

Andre Hyodo
Gustavo Pereira

August 27, 2025

1 Introdução e Hipóteses

Neste estudo, analisamos repositórios populares do GitHub para responder questões relacionadas à contribuição externa, número de releases e frequência de atualização, especialmente em relação às linguagens de programação utilizadas. Hipotetizamos que repositórios escritos em linguagens populares tendem a receber mais contribuições externas, lançar mais releases e serem atualizados com maior frequência do que repositórios em outras linguagens.

2 Metodologia

- Coletamos dados dos 1000 repositórios mais populares do GitHub (com mais de 10.000 estrelas) usando a API REST e GraphQL. - Para cada repositório, extraímos: idade, PRs aceitos, total de releases, tempo desde a última atualização, linguagem primária e percentual de issues fechadas. - Os dados foram processados e sumarizados por linguagem, com foco nas linguagens populares (Python, JavaScript, TypeScript, Java, C, C++, PHP, Shell, C, Go). - Para cada métrica, calculamos valores medianos, além de contagem de repositórios por linguagem.

3 Resultados

Contagem por Categoria (Linguagem de Programação)

- Python: 187 repositórios
- JavaScript: 130
- TypeScript: 156
- Java: 50
- C#: 12
- C++: 48
- PHP: 7
- Shell: 19
- C: 25
- Go: 73
- Outras linguagens: 293

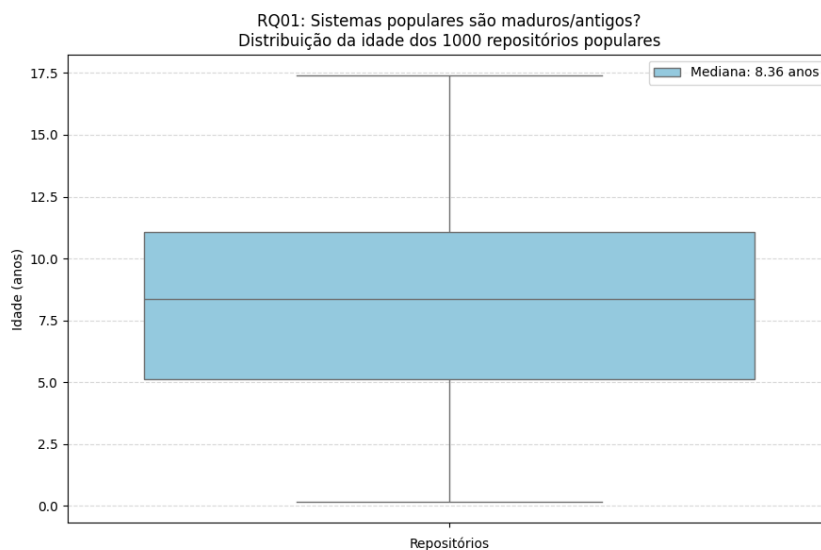
Valores Medianos por Linguagem Popular

Table 1: Valores Medianos por Linguagem Popular

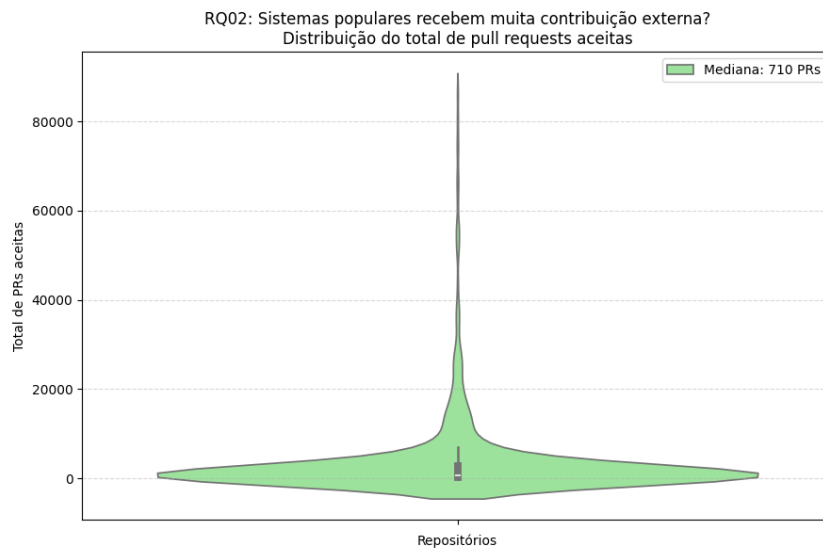
Linguagem	PRs Aceitos	Releases	Última Atualização
Python	548	22	1.0 horas
JavaScript	533.5	33	3.0 horas
TypeScript	2103.5	146.5	1.0 horas
Java	645	42.5	3.0 horas
C#	2780.5	100	0.5 horas
C++	932	58	1.0 horas
PHP	2944	346	0.0 horas
Shell	437	2	4.0 horas
C	113	32	0.0 horas
Go	1677	124	2.0 horas

Valores Medianos Gerais

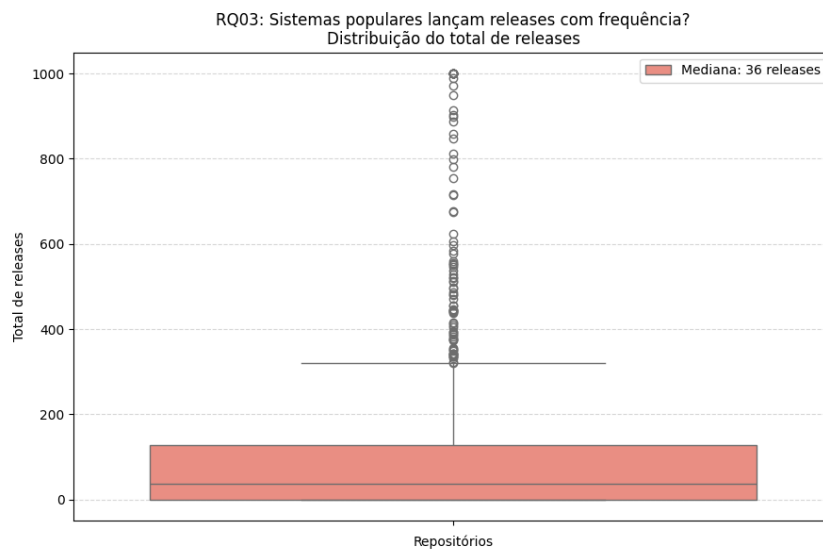
- Mediana de PRs aceitos (todas as linguagens populares): varia de 113 (C) a 2944 (PHP). - Mediana de releases: varia de 2 (Shell) a 346 (PHP). - Mediana do tempo desde a última atualização: a maioria das linguagens populares tem valores baixos (0 a 4 horas), indicando alta frequência de atualização.



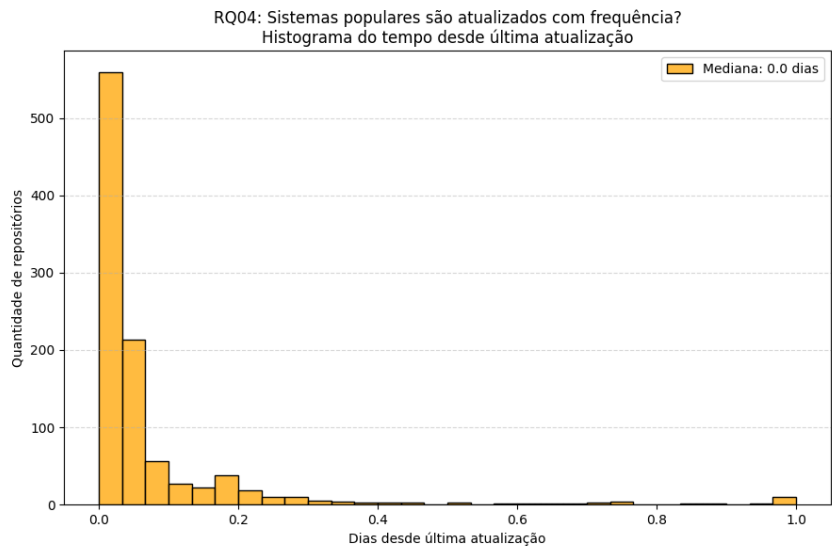
Distribuição da idade dos repositórios (boxplot)



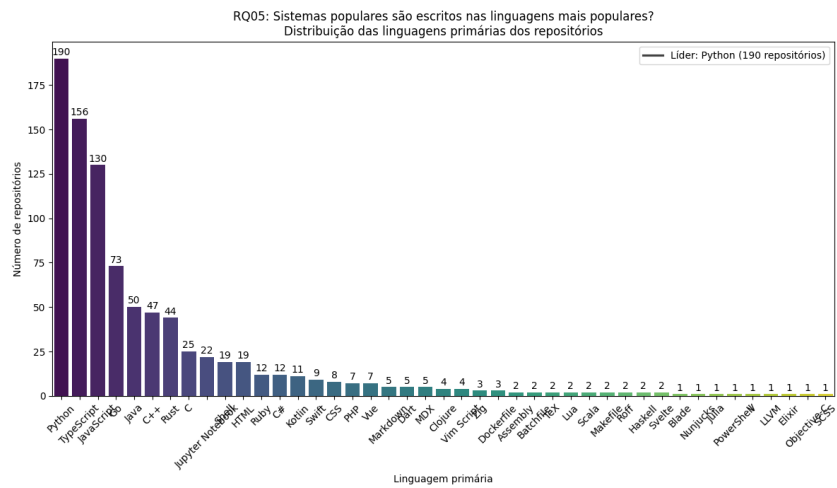
Distribuição do total de PRs aceitas (violinplot)



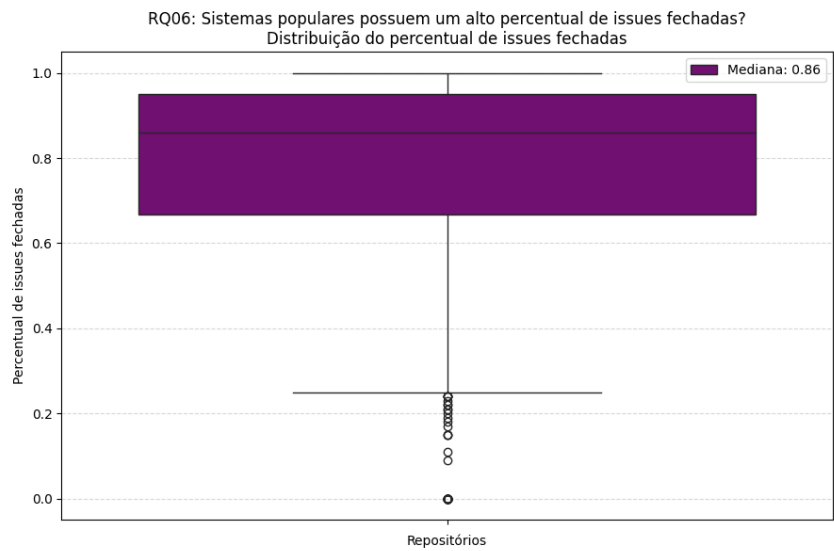
Distribuição do total de releases (boxplot)



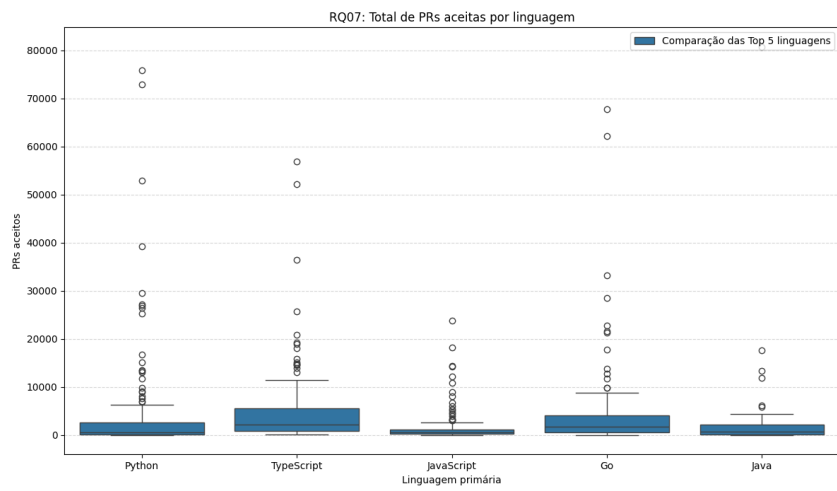
Histograma do tempo desde última atualização



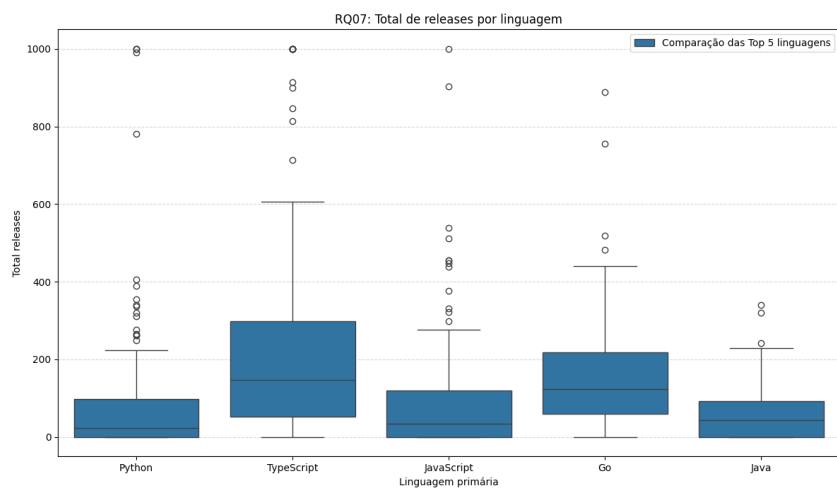
Distribuição das linguagens primárias (barplot)



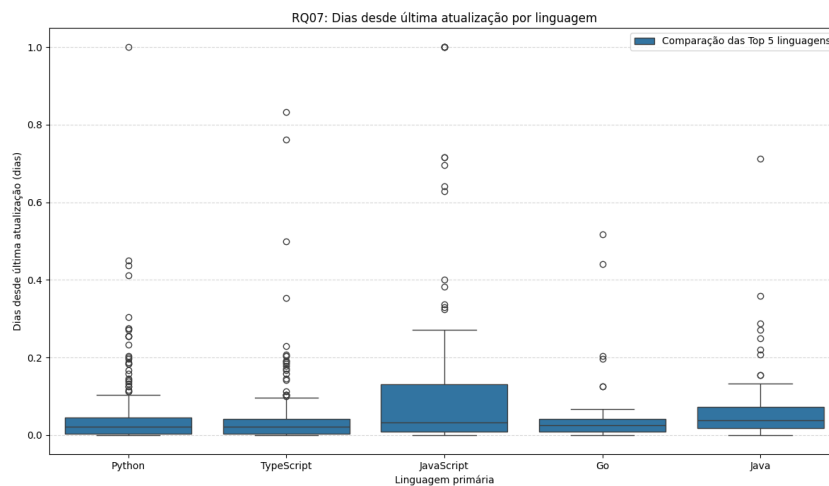
Percentual de issues fechadas (boxplot)



Total de PRs aceitos por linguagem (mediana)



Total de releases por linguagem (mediana)



Tempo desde última atualização (dias) por linguagem (mediana)

4 Discussão

Hipóteses e Expectativas

Esperávamos que repositórios em linguagens populares apresentassem:

- Mais contribuições externas (PRs aceitos)
- Mais releases
- Atualizações mais frequentes

Análise dos Resultados

- As linguagens populares realmente concentram a maior parte dos repositórios analisados. - PHP, TypeScript, C# e Go apresentam valores medianos de PRs aceitos e releases bastante elevados, sugerindo que grandes projetos nessas linguagens recebem muitas contribuições e lançam releases com frequência. - O tempo mediano desde a última atualização é baixo para quase todas as linguagens populares, indicando que esses projetos são mantidos ativamente. - Linguagens menos populares ou "outras" tendem a ter menos repositórios e, em geral, valores medianos menores para PRs aceitos e releases.

Considerações

- Os resultados confirmam parcialmente as hipóteses: linguagens populares concentram projetos com mais atividade e manutenção. - Entretanto, há variação significativa entre as linguagens populares (ex: Shell e C têm valores medianos bem menores). - A amostra pode ser enviesada por projetos muito grandes em certas linguagens (ex: PHP com poucos, mas enormes projetos). - O tempo desde a última atualização é baixo para quase todos, pois os repositórios populares tendem a ser muito ativos.

5 Conclusão

O estudo das principais características de sistemas open-source populares no GitHub revela um padrão de maturidade, colaboração intensa, entrega contínua e manutenção eficiente. A escolha da linguagem é um fator determinante para o engajamento e sustentabilidade dos projetos, reforçando a importância de ecossistemas ativos para o sucesso do software livre.

Referências

- GitHub Octoverse 2024. Disponível em: <https://github.blog/news-insights/octoverse/octoverse-2024>