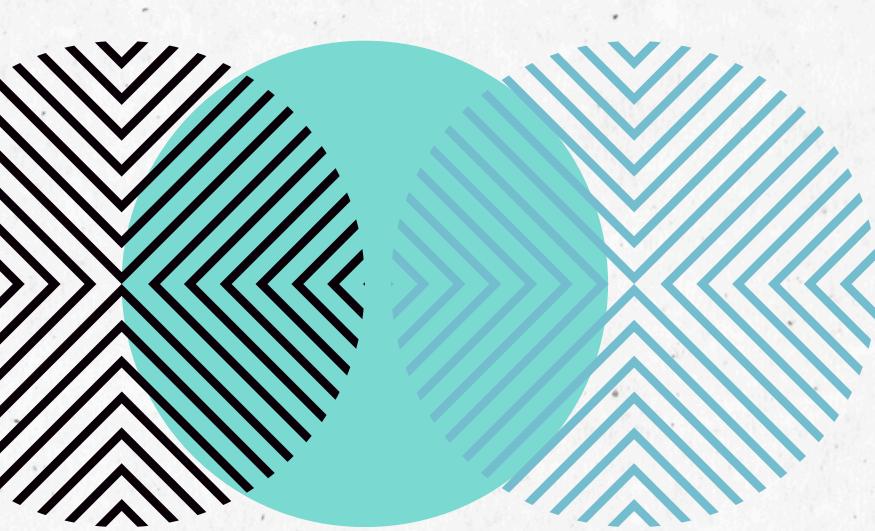


Laboratório 1

Análise das principais características de
repositórios populares open-source

Pedro Henrique Moreira
Lucca
Jhonata Dias



Objetivos

- Responder às questões de pesquisa sobre características de sistemas populares no GitHub
- Identificar padrões que levam ao sucesso dos repositórios open-source



Questões de Pesquisa

RQ01: Sistemas populares são maduros/antigos?

RQ02: Recebem muita contribuição?

RQ03: Lançam releases com frequência?

RQ04: São atualizados com frequência?

RQ05: Usam linguagens mais populares?

RQ06: Possuem alto percentual de issues fechadas?

RQ07: Linguagens populares atraem mais contribuições? (bônus)

Hipóteses Informais

H01: Idade média 7-8 anos

H02: Muitos PR aceitos

H03: Lançam versões frequentemente

H04: Commits regulares e manutenção ativa

H05: Uso de JS, Python e Java

H06: Alto percentual de issues fechadas

H07: Linguagens populares atraem mais contribuições

metodologia

Coleta: API GraphQL do GitHub

Linguagem: Python 3.12

Bibliotecas: Requests, OS, python-dotenv

Processo:

- Coleta dados
- Sumarização por medianas
- Análises estatísticas + gráficos

Resultados Gerais

Métricas coletadas:

- Idade dos repositórios
- Pull Requests aceitos
- Releases
- Dias desde última atualização
- Linguagens mais usadas
- Percentual de issues fechadas
- Comparação por linguagens populares

RQ01: Idade

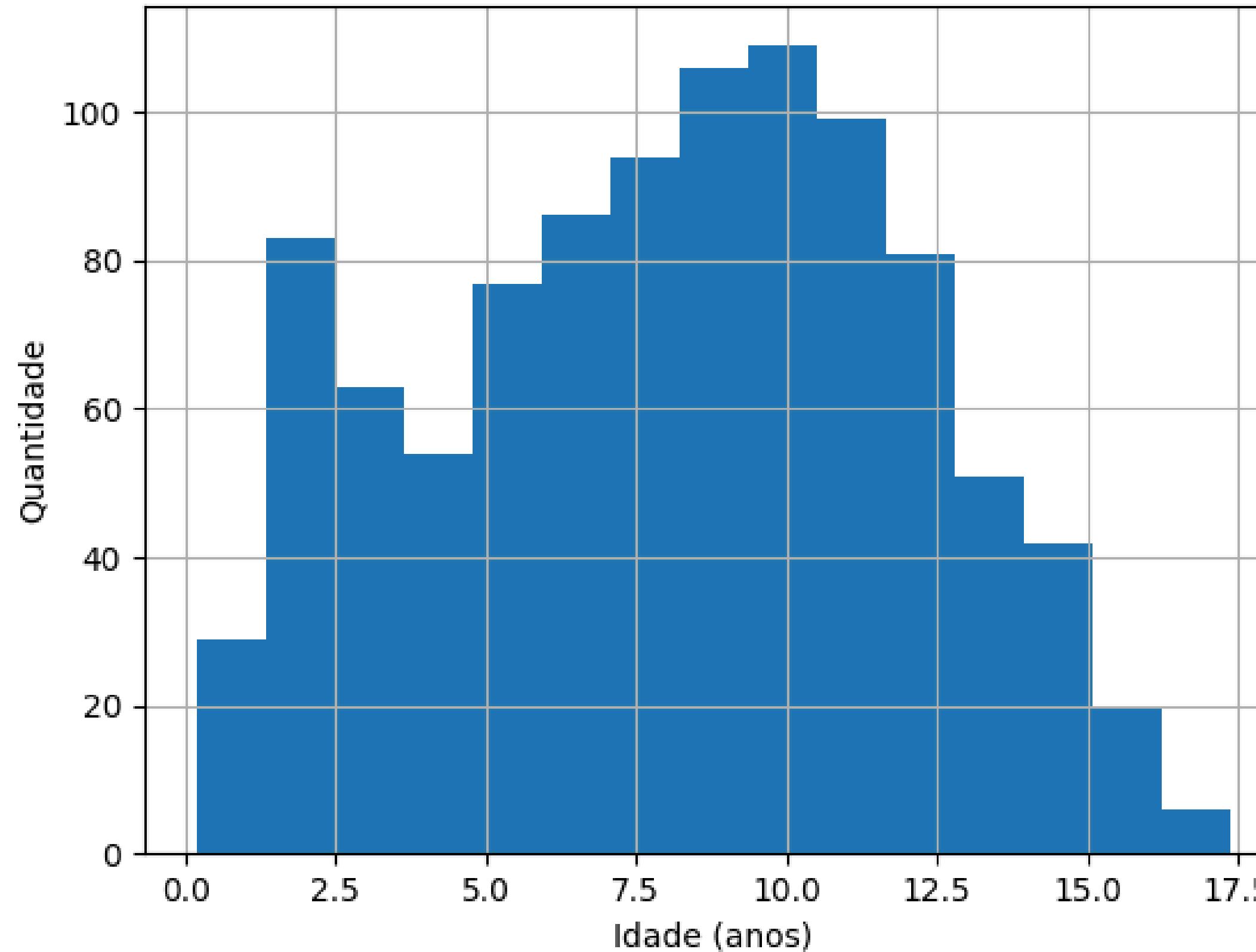
Mediana: 8.37 anos

Média: 8.09 anos

Moda: 9 anos

Conclusão: repositórios populares são relativamente maduros

Distribuição da Idade dos Repositórios (anos)



RQ02: Contribuições

PRs aceitos:

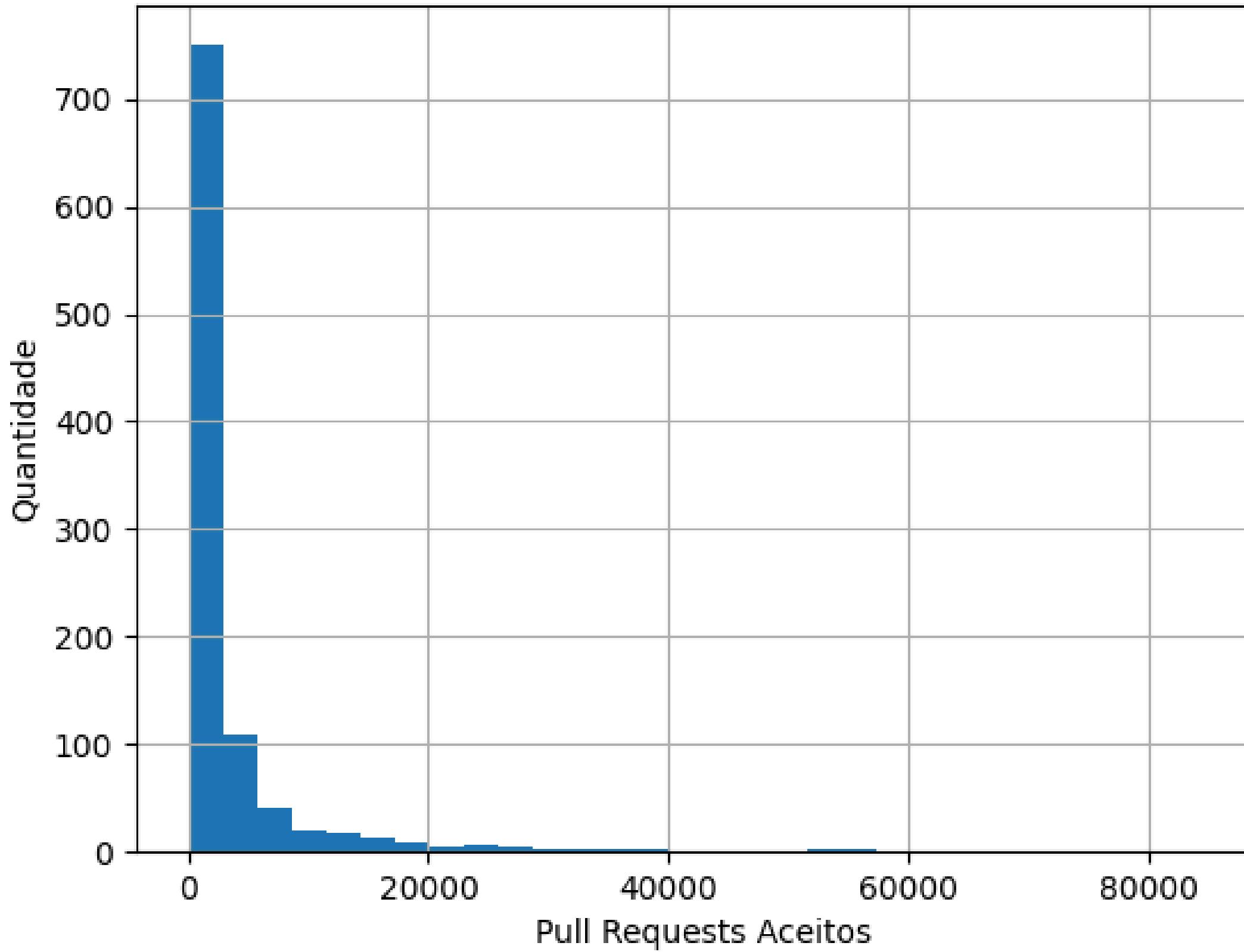
Mediana: 702

Média: 3587 (influenciada por outliers)

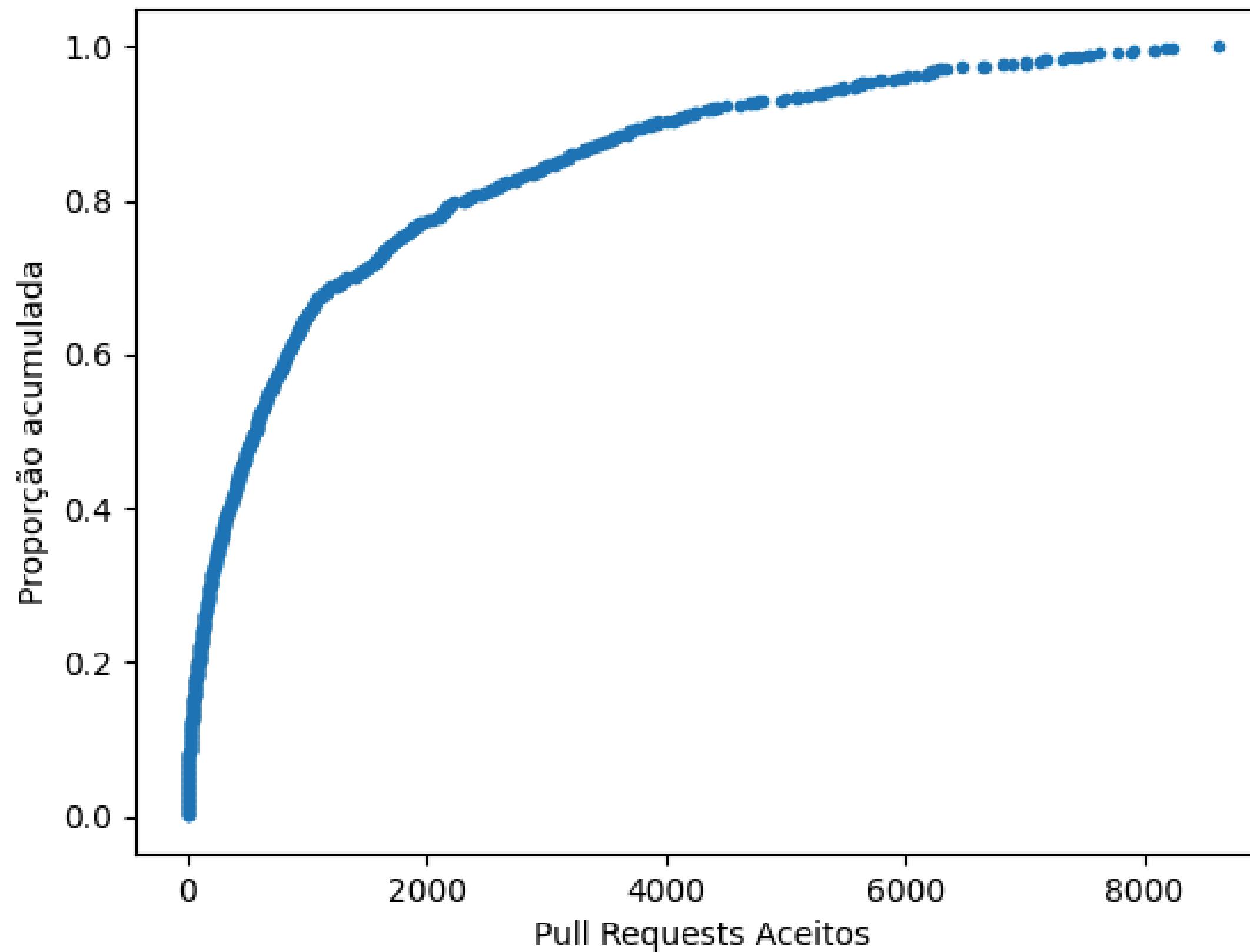
Moda: 1

Conclusão: maioria dos repositórios não recebe tanta contribuição

Distribuição Pull Requests Aceitos



ECDF de Pull Requests Aceitos (até o 90º percentil)



RQ03: Releases

Mediana: 35

Média: 109

- Muitos projetos não usam releases formalmente
- Hipótese de lançamentos frequentes foi refutada

RQ04: Atualizações

Dias desde última atualização:

Mediana: 7 dias

Média: 7.01 dias

Moda: 7

Conclusão: atualizações constantes → Hipótese confirmada

RQ05: Linguagens

Top linguagens:

- Python (189)
- TypeScript (156)
- JavaScript (129)
- Go (73)
- Java (50)

RQ06: Issues Fechadas

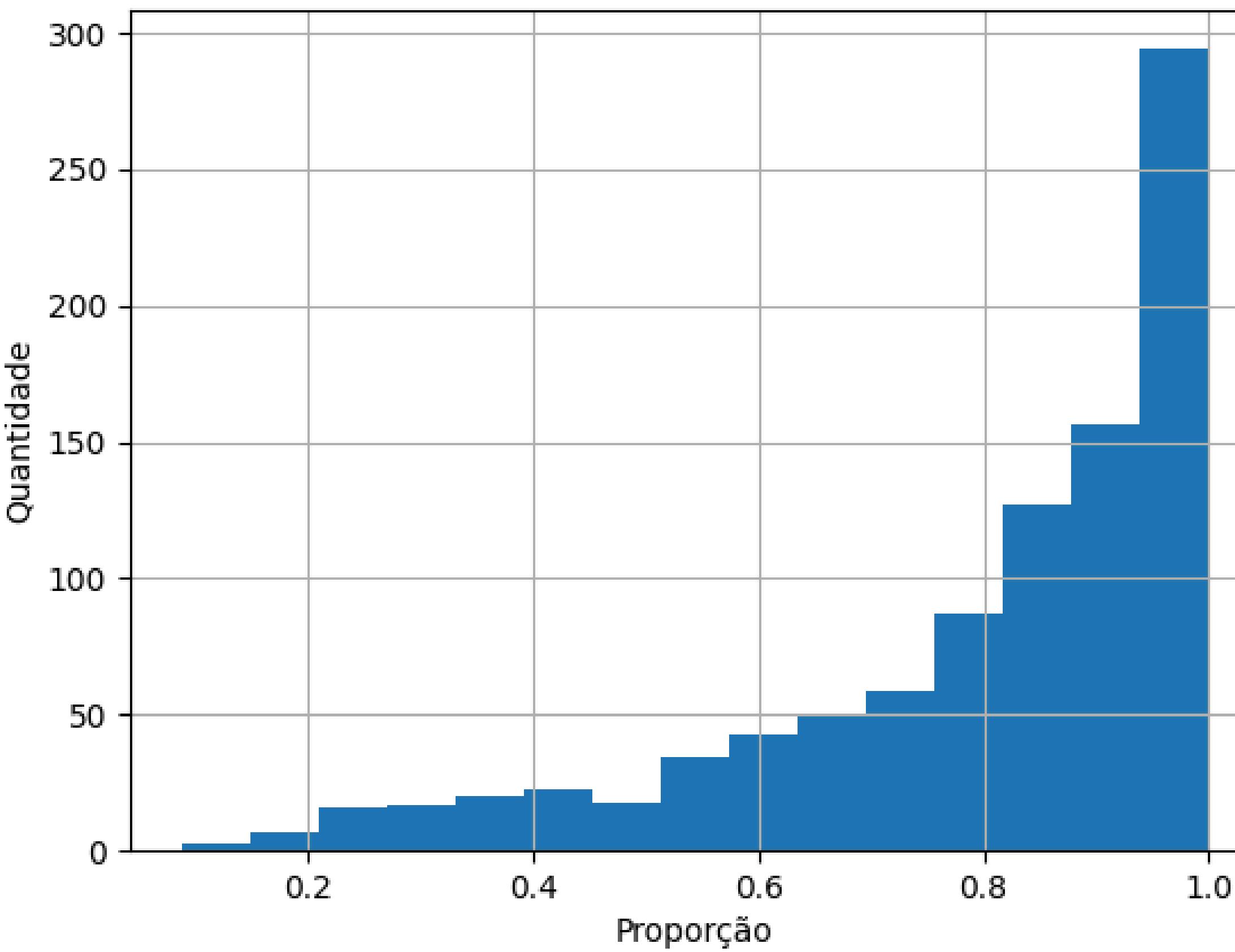
Mediana: 86.56%

Média: 79.95%

Moda: 100%

Conclusão: repositórios mantêm alta taxa de resolução

Distribuição da Razão de Issues Fechadas



RQ07: Linguagens Populares

PRs aceitos (mediana):

- **Populares:** 927
- **Outras:** 415

Releases:

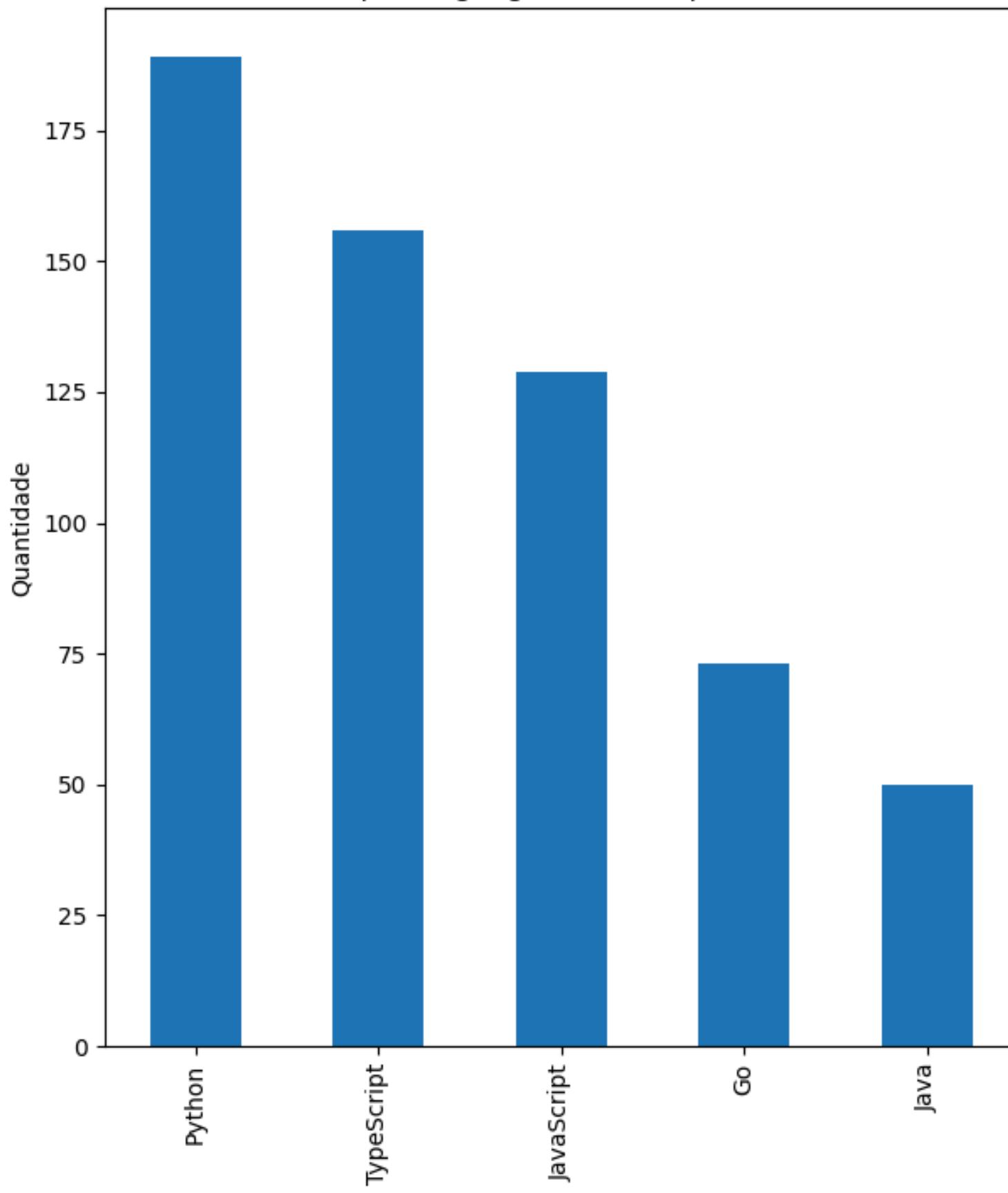
- **Populares:** 62
- **Outras:** 3

Dias desde última atualização:

- **Populares:** 7.0
- **Outras:** 7.0

Conclusão: linguagens populares atraem mais contribuições

Top 5 Linguagens Mais Populares



Conclusões

- Projetos populares são maduros
- Atualizados com frequência
- Usam linguagens amplamente conhecidas
- Issues geralmente bem resolvidas
- Contribuições concentradas em poucos projetos
- Linguagem influencia engajamento da comunidade

OBRIGADO!

