

# Caracterizando a atividade de code review no GitHub

Gabriel Faria, João Victor Salim, Lucas  
Garcia, Maisa Pires e Miguel Vieira



# HIPÓTESES

---

## H01

PRs menores  
(menos arquivos e  
linhas modificadas)  
têm maior  
probabilidade de  
serem merged.

## H02

PRs com  
descrições mais  
detalhadas têm  
maior probabilidade  
de serem merged.

## H03

PRs que demoram  
mais tempo para  
serem analisados  
têm menor  
probabilidade de  
serem merged.

# HIPÓTESES

---

## H01

PRs com mais interações (comentários e participantes) têm maior probabilidade de serem merged.

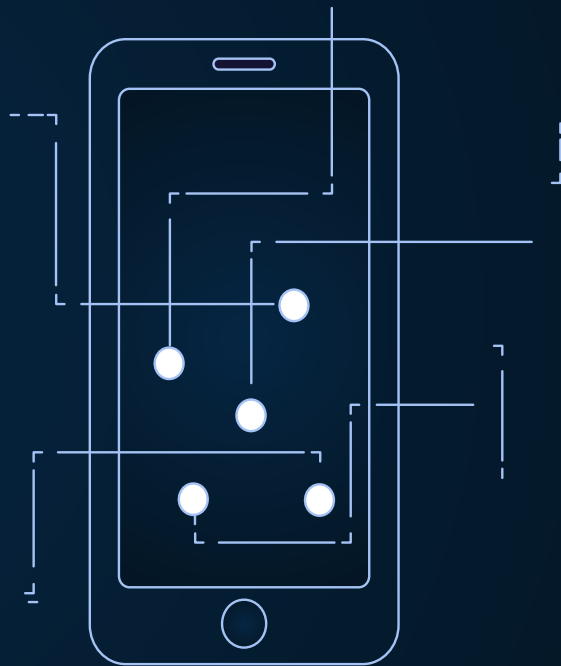
## H03

PRs maiores requerem mais revisões.

# TECNOLOGIAS

---

- **Linguagens:**
  - a. Python 3.10+ (análise e geração de gráficos)
- **Bibliotecas**
  - a. requests, pandas, numpy, matplotlib, seaborn, scipy
- **Análise estática**
  - a. Teste de correlação de Spearman
- **API:**
  - a. GitHub GraphQL API (consulta de repositórios Java por popularidade)



# METODOLOGIA

---

- **Coleta de dados**

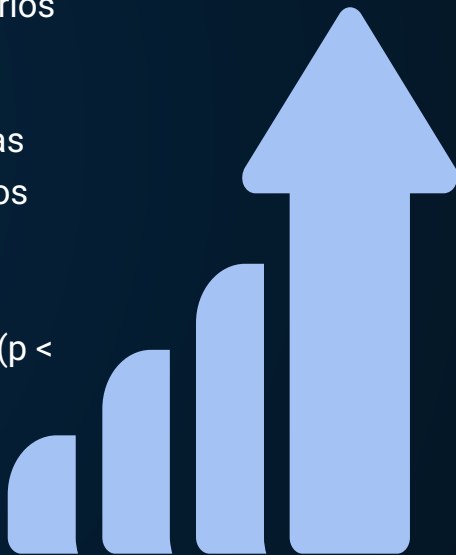
- Foram coletados 500 PRs de 200 repositórios populares utilizando a GitHub REST API. Critério: amostra balanceada por repositório (2–3 PRs), repositórios ordenados por estrelas.

- **Consolidação**

- Os dados foram consolidados em `data/pull_requests_500.csv` com métricas derivadas (tamanho, tempo, interações). Resultados e figuras foram gerados em `results_500pr/`, com resumo em `results_500pr/analysis_results.json`.

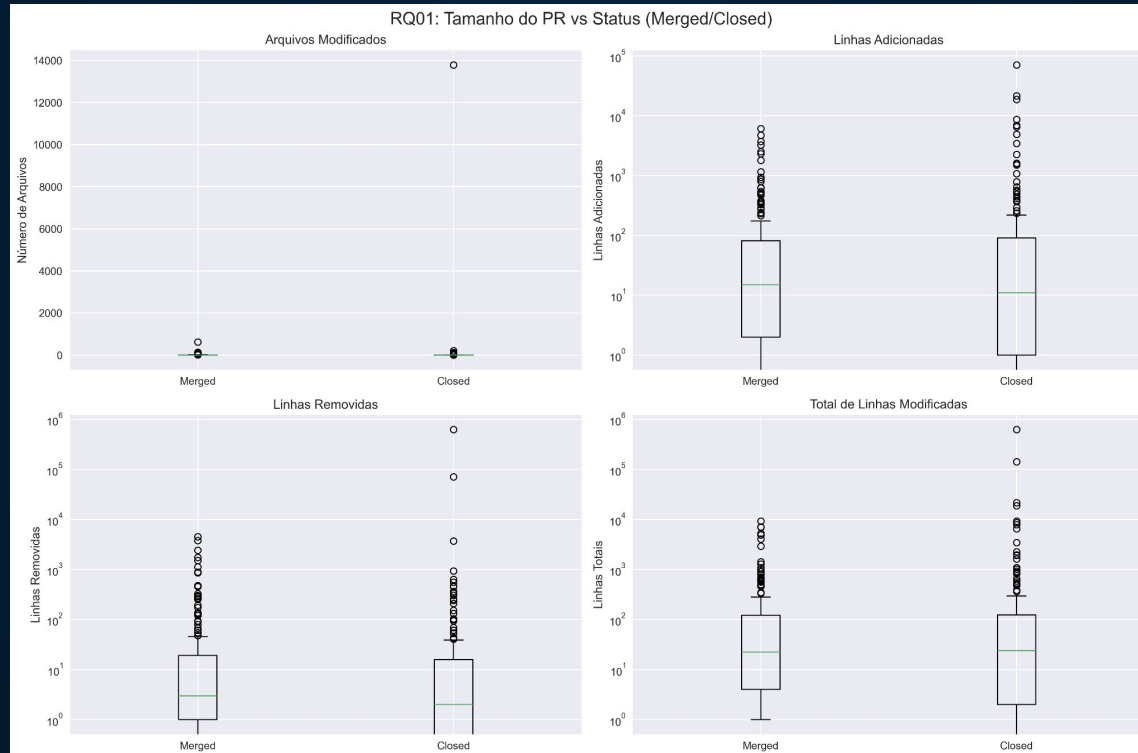
- **Análise**

- Foi realizada estatística descritiva (medianas) e correlações de Spearman ( $p < 0,05$ ). Foram gerados gráficos por RQ e uma matriz de correlação das principais variáveis.



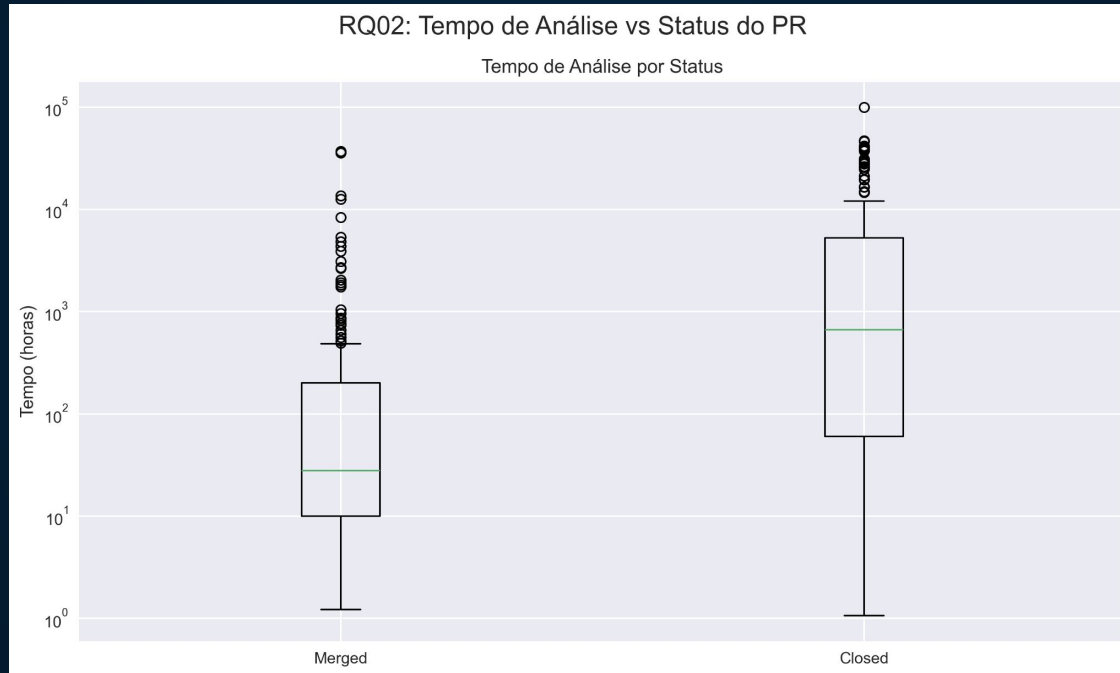
# RESULTADO 1

Qual a relação entre o tamanho dos PRs e o feedback final das revisões?



# RESULTADO 2

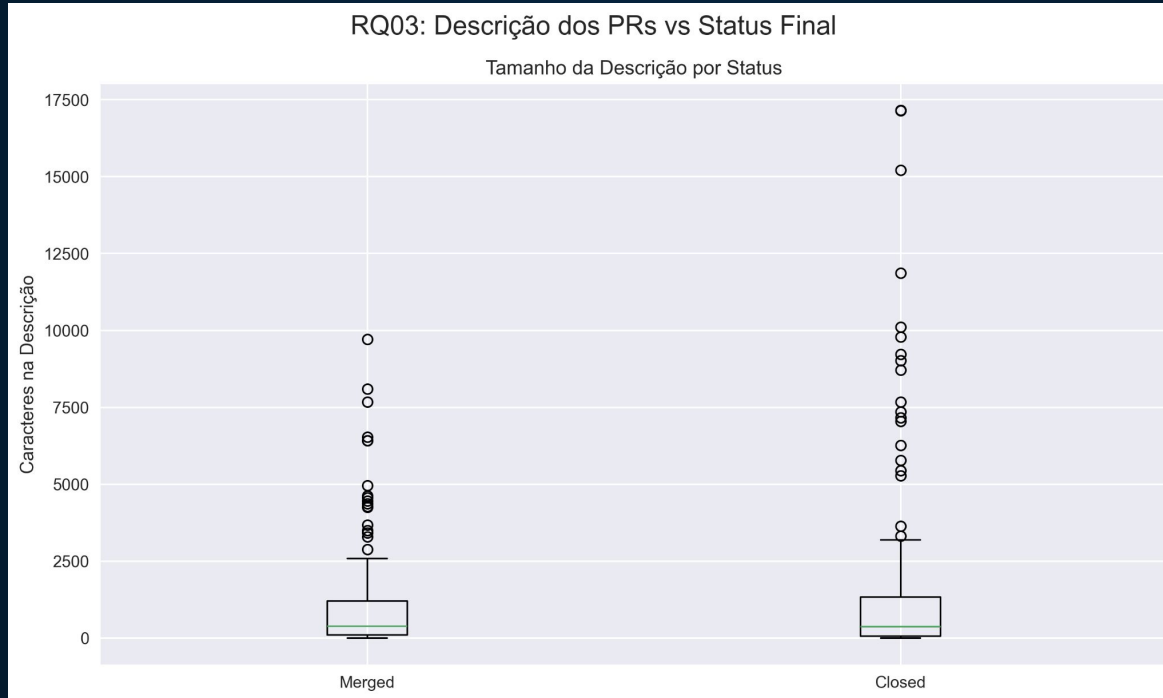
Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?





# RESULTADO 3

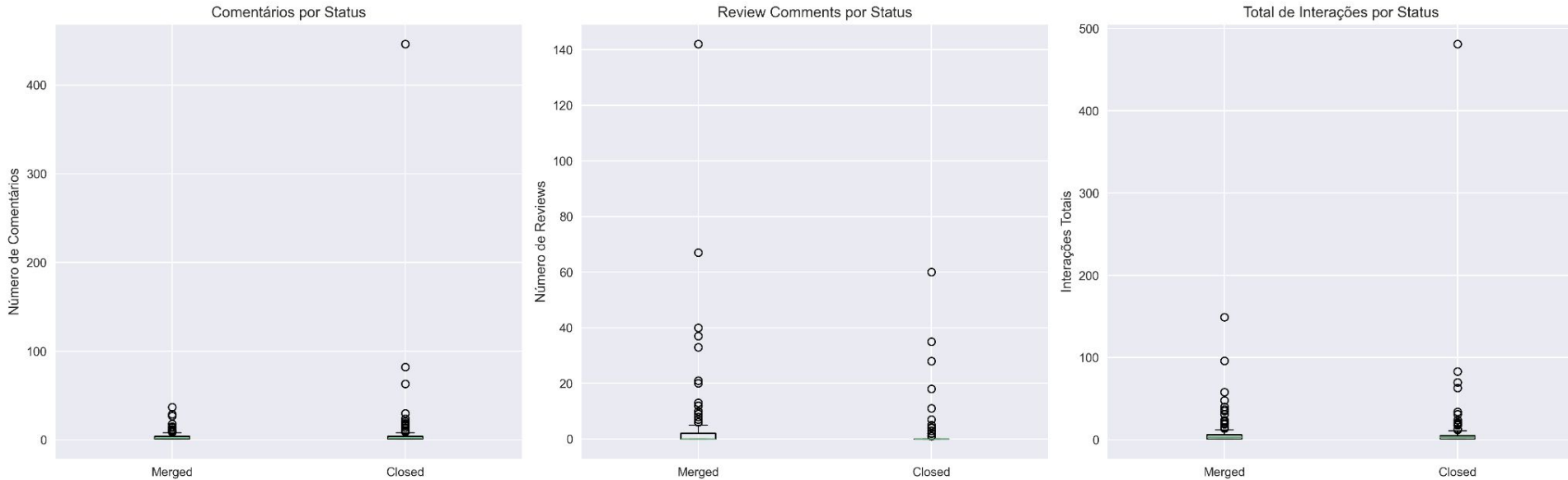
Qual a relação entre a descrição dos PRs e o feedback final das revisões?



# RESULTADO 4

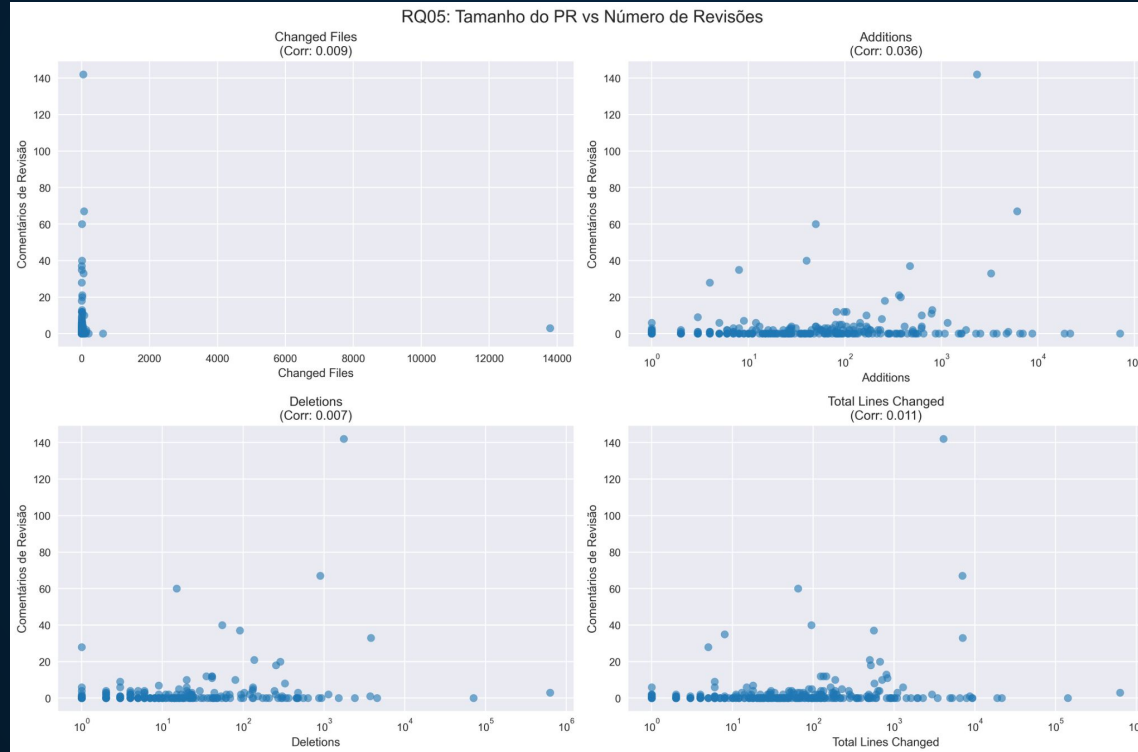
Qual a relação entre as interações nos PRs e o feedback final das revisões?

RQ04: Interações nos PRs vs Status Final



# RESULTADO 5

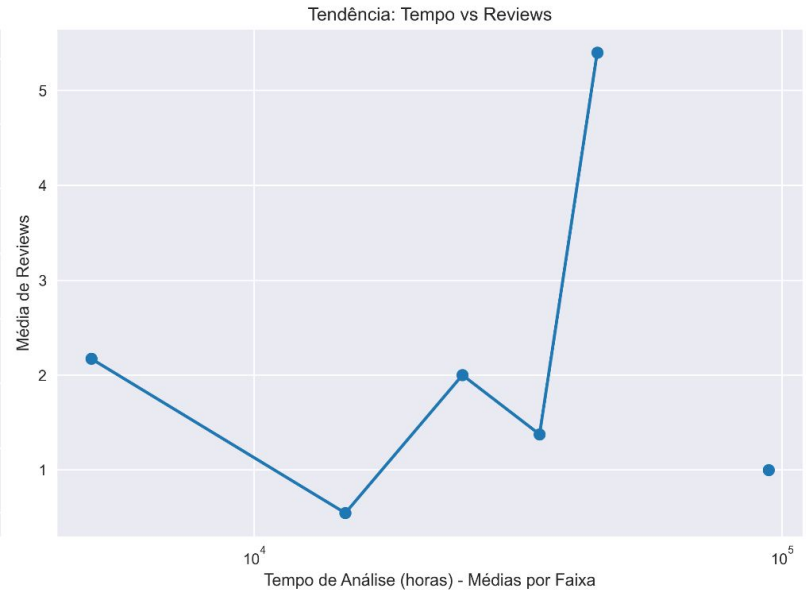
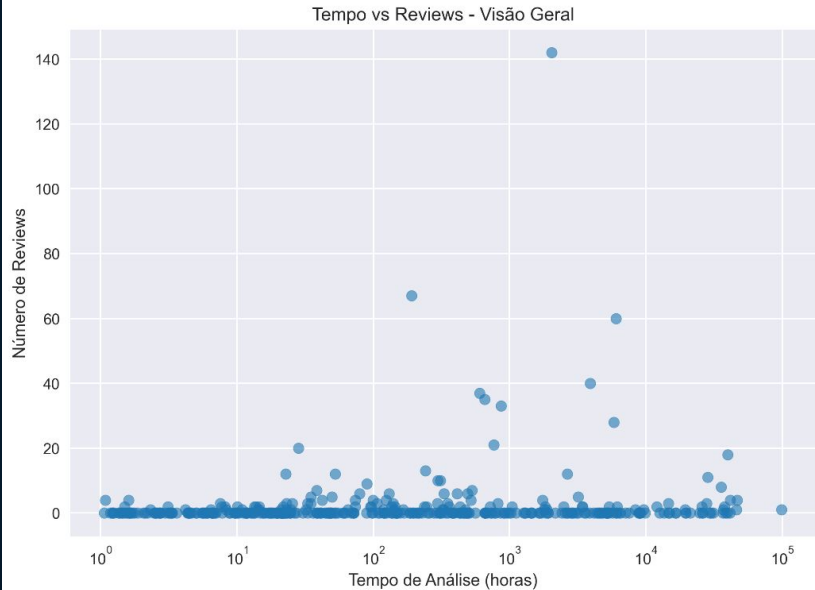
Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?



# RESULTADO 6

Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?

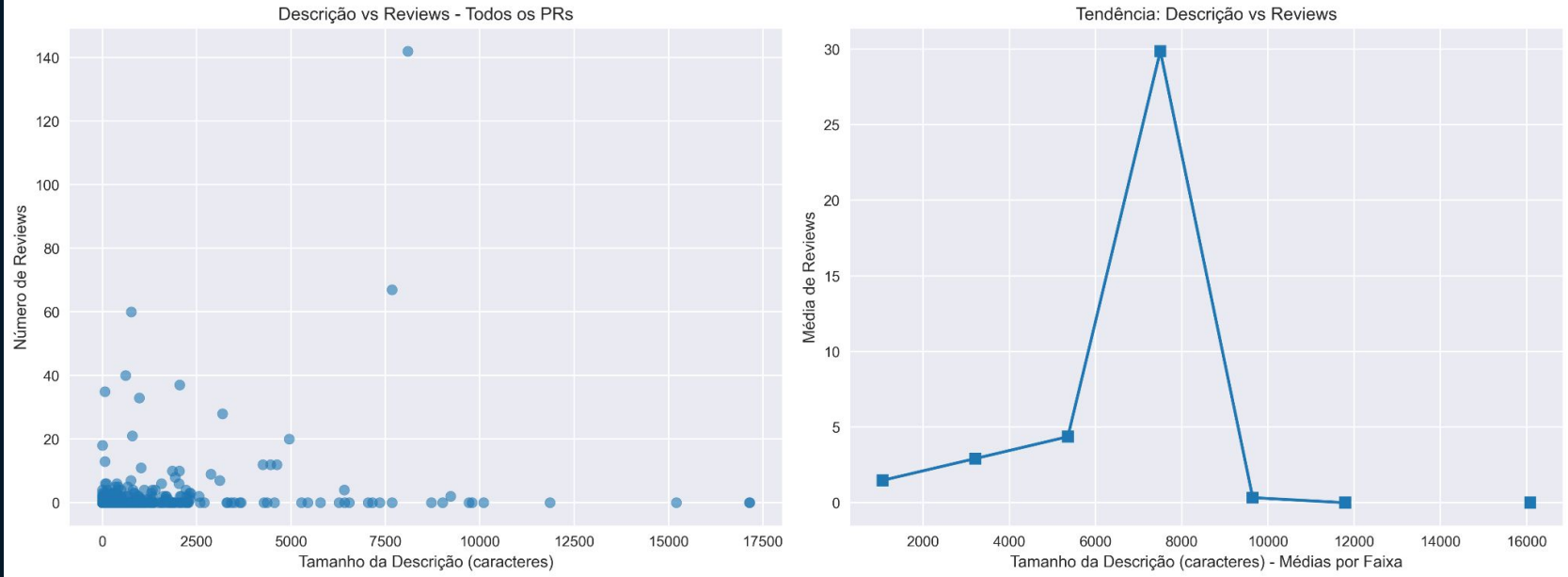
RQ06: Tempo de Análise vs Número de Revisões



# RESULTADO 7

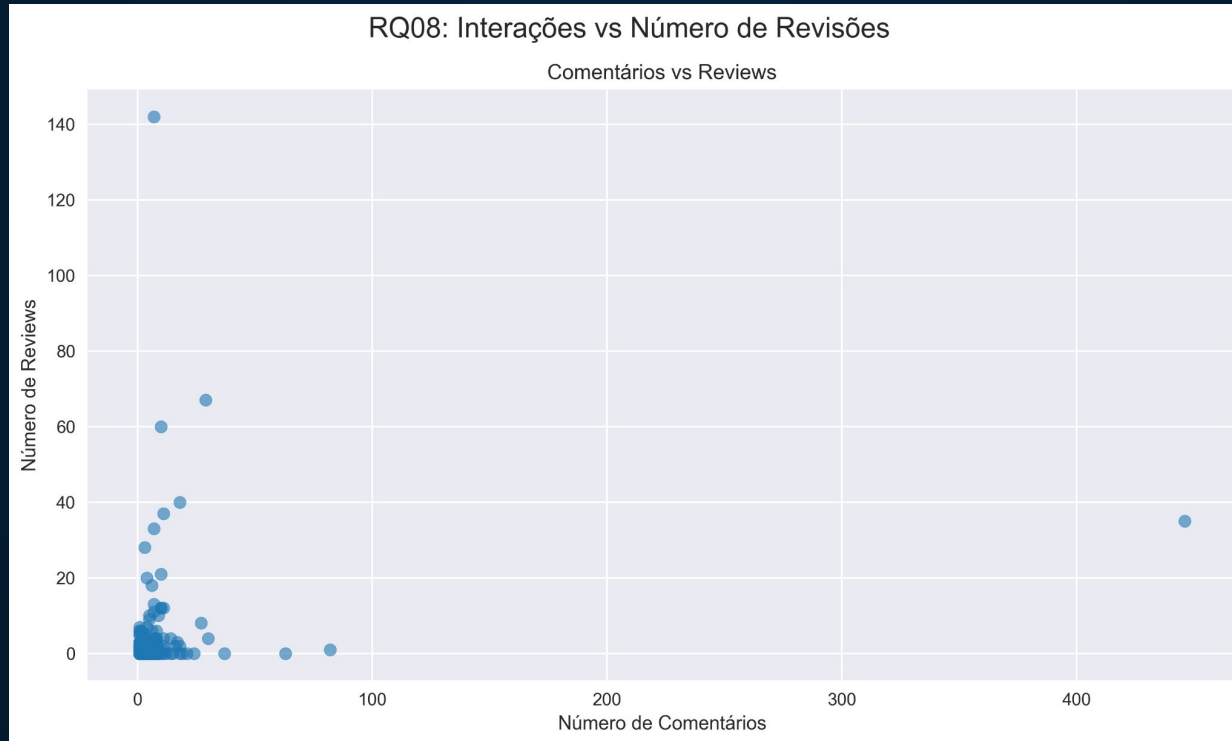
Qual a relação entre a descrição dos PRs e o número de revisões realizadas?

RQ07: Tamanho da Descrição vs Número de Revisões



# RESULTADO 8

Qual a relação entre as interações nos PRs e o número de revisões realizadas?



# Correlação Geral



# DISCUSSÃO

---

Tempo é crítico para o merge: correlação negativa forte entre tempo de análise e status. PRs que ficam muito tempo abertos tendem a fechar sem merge.

Tamanho influencia revisões: PRs maiores recebem mais reviews alinhado à IH05.

Interações: comments gerais se correlacionam com reviews mas não alteraram significativamente o status. Participantes ficou constante e não foi informativo.

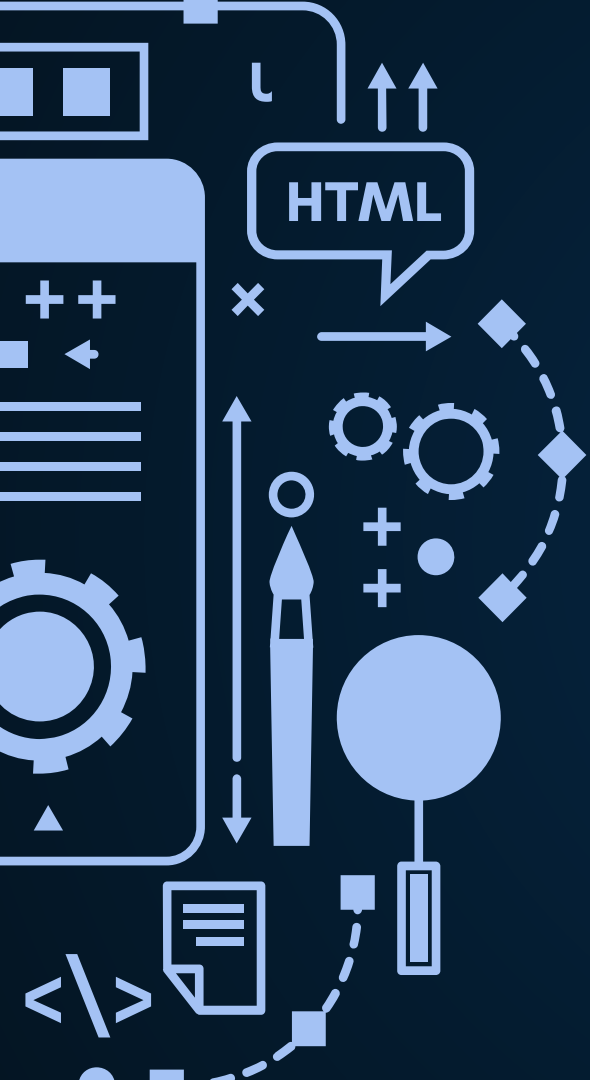
- IH01: NÃO CONFIRMADA — efeitos pequenos; leve tendência oposta em arquivos modificados.
- IH02 NÃO CONFIRMADA — sem efeito significativo em status.
- IH03: CONFIRMADA —  $r \approx -0,409$ .
- IH04 PARCIALMENTE — depende do tipo de interação (comments se relacionam apenas a reviews).
- IH05 CONFIRMADA —  $r \approx 0,22-0,26$ .



# CONCLUSÃO

Na amostra de 500 PRs (381 válidos), concluímos:

- O tempo de análise se relaciona fortemente com o resultado ( $r \approx -0,409$ ): PRs com longas filas têm menor chance de merge.
- PRs maiores demandam mais revisões ( $r \approx 0,22-0,26$ ), mas tamanho não determinou o status de forma forte.
- Comentários gerais aumentam a atividade de review ( $r \approx 0,222$ ), sem evidência clara de impacto no merge; participantes não foi informativo.
- Descrições maiores não mostraram ganho significativo de probabilidade de merge.



# Obrigado!

Alguma Pergunta?

# Bibliografía

- GitHub REST API: <https://docs.github.com/en/rest>
- Pandas: <https://pandas.pydata.org/>
- Matplotlib: <https://matplotlib.org/>
- Seaborn: <https://seaborn.pydata.org/>
- SciPy: <https://scipy.org/>