**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**

**INSTITUTO DE CIÊNCIAS EXATAS E INFORMÁTICA  
UNIDADE EDUCACIONAL PRAÇA DA LIBERDADE**

**Bacharelado em Engenharia de Software**

**Camila Silva Romagnoli**

**Leonardo Augusto Moreira Lourenço**

**Octávio Tabai Ribeiro Lage**

**Trabalho Prático 3**

Belo Horizonte

2022

**Camila Silva Romagnoli**

**Leonardo Augusto Moreira Lourenço**

**Octávio Tabai Ribeiro Lage**

**Trabalho Prático 3**

Trabalho apresentado na disciplina Laboratório de Experimentação de Software.

Professor: José Laerte Pires Xavier

Belo Horizonte

2022

# 1 INTRODUÇÃO

Essa atividade foi proposta na disciplina de Laboratório de Experimentação de Software. O relatório apresentado tem como objetivo de caracterizar a atividade de code review em repositórios no GitHub, utilizando métricas de tamanho, tempo de análise, descrição e interação.

## 1.1 Hipóteses

**RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**

**Hipótese (Quantidade de arquivos x Quantidade de PRs aceitos):** Os pull requests com muitos arquivos possuem menos chances de serem aceitos.

**RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**

**Hipótese (Tempo de análise x Quantidade de PRs aceitos):** Os pull requests com maior tempo de análise possuem menos chances de serem aceitos.

**RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?**

**Hipótese (Número de caracteres x Status dos PRs):** Os pull requests com maior número de caracteres na descrição possuem mais chances de serem aceitos.

**RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?**

**Hipótese (Número de comentários x Status das PRs):** Os pull requests com mais interações possuem menos chances de serem aceitos.

**RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

**Hipótese (Número de arquivos x Número de revisões):** Quanto maior o tamanho do PR, maior o número de revisões realizadas.

**RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

**Hipótese (Número de revisões x Tempo de análise [Dias]):** Quanto maior o tempo de análise, maior o número de revisões realizadas.

**RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

**Hipótese (Quantidade de caracteres x Número de revisões):** Quanto maior o número de caracteres na descrição, maior a quantidade de revisões realizadas do PR.

**RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

**Hipótese (Número de revisões x Número de comentários):** Quanto maior a quantidade de comentários, maior o número de revisões realizadas na PR.

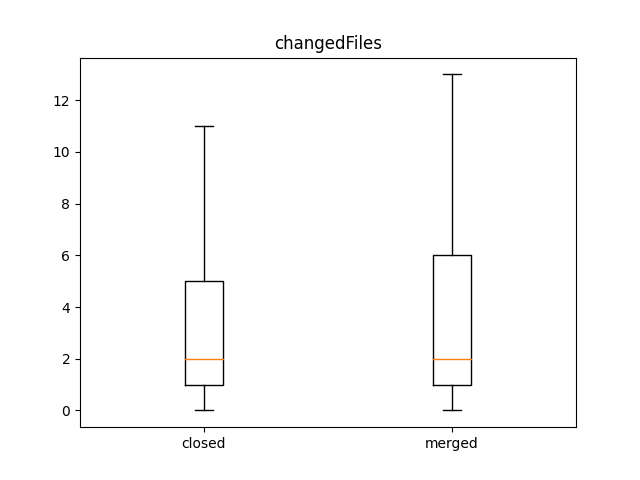
# 2 METODOLOGIA

A metodologia utilizada para a análise das atividades de Code Review consiste na coleta de dados quantitativos e análises estatísticas. Esta abordagem, ocorre a partir da migração de PRs submetidos a repositórios populares do GitHub, que possuam pelo menos cem PRs com status MERGED ou CLOSED. Além disso, a fim de analisar os pull requests que tenham passado pelo processo de code review, serão analisados apenas os PRs que possuem os status citados anteriormente e que possuam pelo menos uma revisão. Além disso, foram filtrados os PRs que tenham sido revisados em pelo menos uma hora, ou seja, o cálculo entre a data de criação do PR e de seu encerramento (merge/closed) deve ser maior que uma hora.

Para a coleta de dados, foi elaborado e executado um script em Python que realiza uma consulta em GraphQL dos 100 repositórios mais populares do GitHub e exporta os dados em um arquivo csv.

Para a análise estatística sobre o Feedback Final das Revisões foram gerados gráficos de caixa levando em consideração as medianas dos dados obtidos. Para a análise estatística sobre o Número de Revisões foram gerados gráficos de correlação e calculados o coeficiente de Spearman entre os aspectos da questão de pesquisa.

# 3 RESULTADOS

**RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?**

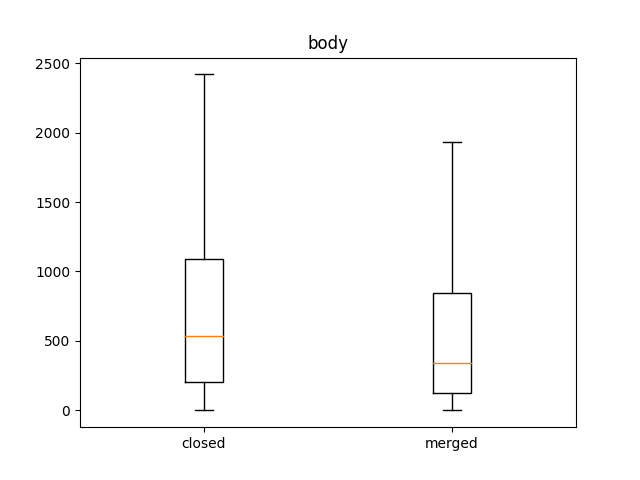
De acordo com os dados obtidos, pôde-se observar que a quantidade de arquivos modificados em ambos os status (fechados e aceitos) são semelhantes, pois a mediana é igual e seus limites e distribuições são próximos. Dessa forma, foi refutada a hipótese inicial de que os pull requests com menos arquivos possuem mais chances de serem aceitos, pois não foi observada essa relação.

**RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?**

****

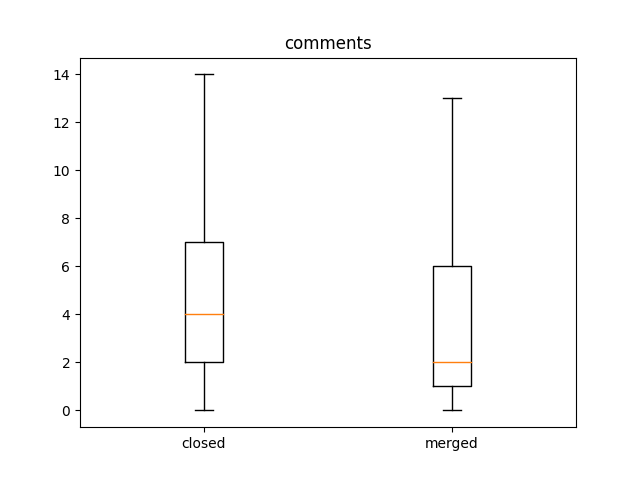
O tempo de análise entre pull requests fechados e aceitos têm medianas semelhantes. No entanto, nos pull requests fechados o intervalo de variações das distribuições é maior. Dessa forma, podemos confirmar a hipótese inicial de que pull requests que possuem maior tempo de análise possuem menos chances de serem aceitos.

**RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?**

****

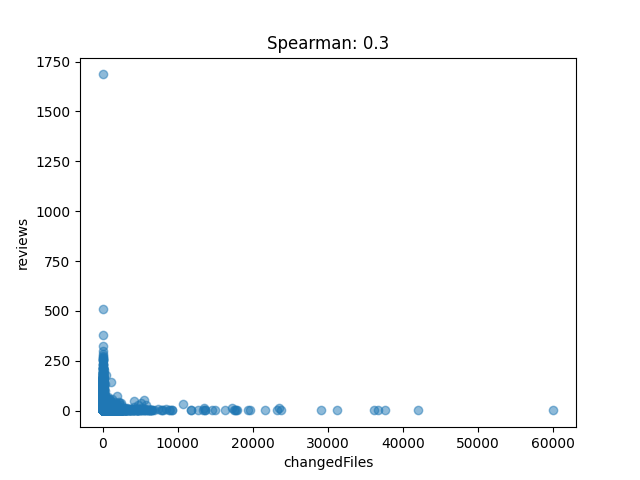
Os pull requests fechados e aceitos têm distribuições semelhantes e a mediana é próxima no quesito quantidade de caracteres da descrição, portanto esse aspecto é indiferente no feedback da revisão. Dessa forma, podemos refutar a hipótese inicial que pull requests com mais caracteres na descrição possuem mais chances de serem aceitos.

**RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?**



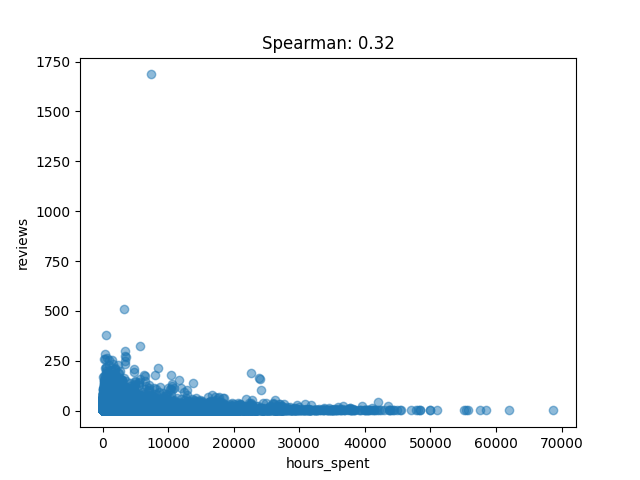
A quantidade de interações de pull requests fechados e aceitos possuem variação de distribuição semelhantes. No entanto, a mediana de interações em pull requests aceitos é menor e existe uma distribuição mais simétrica em pull requests fechados. Dessa forma, podemos ver que pull requests aceitos tem uma quantidade de comentários significativa e não muito distinta dos pull requests fechados. Dessa forma, podemos refutar a hipótese inicial de que pull requests com muitas interações tem mais interações têm menos chances de serem aceitos.

**RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?**

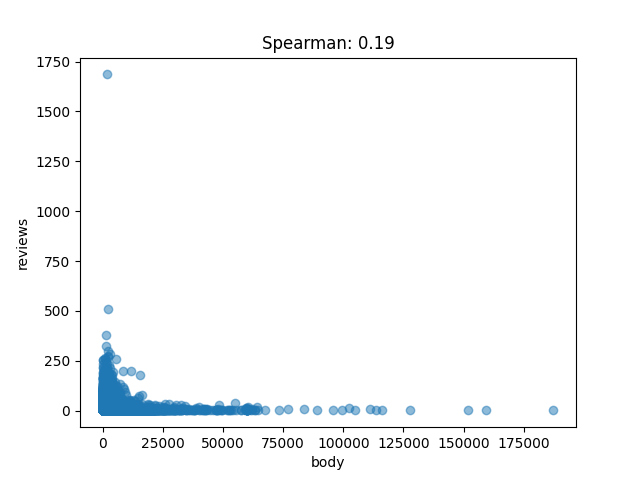
****

O tamanho do pull request e o número de revisões apresentou uma correlação positiva fraca (0.3) entre os dados. Portanto, apesar de existir uma pequena relação entre os dois aspectos, não é possível concluir que a quantidade de arquivos influencia na revisão. Dessa forma, podemos refutar a hipótese inicial de que quanto maior o PR, maior o número de revisões realizadas.

**RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?**

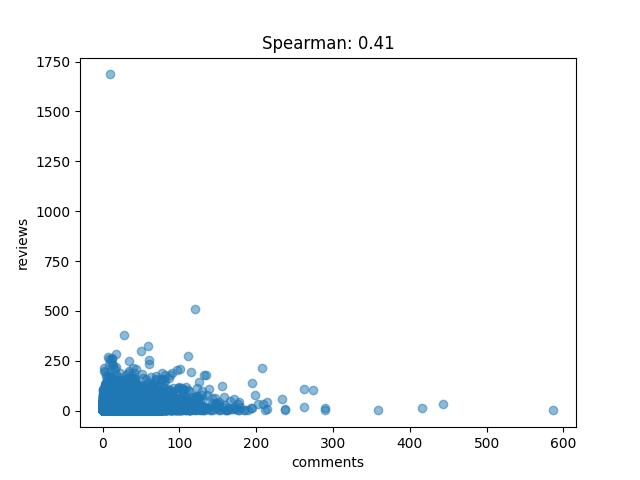
****

Foi possível obter uma pequena relação entre os dois aspectos. O coeficiente de Spearman calculado mostra que existe uma correlação positiva (0.32) entre os dados. Desse modo, podemos confirmar a hipótese inicial de que quanto maior o tempo de análise, maior o número de revisões realizadas.

**RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?**

A descrição de pull requests e o número de revisões apresentou uma correlação positiva fraca (0.19) entre os dados. Portanto, apesar de existir uma pequena relação entre os dois aspectos, não é possível concluir que a quantidade de caracteres na descrição influencia na revisão. Dessa forma, podemos refutar a hipótese inicial de que quanto maior a descrição do PR, maior o número de revisões realizadas.

**RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?**

****

A quantidade de interações nos pull requests e o número de revisões apresentou uma correlação positiva considerável (0.41) entre os dados. Portanto, existe uma correlação entre os aspectos, pois os pull requests com menos comentários estão concentrados no eixo que apresenta menor número de revisões. Dessa forma, podemos confirmar a hipótese inicial de que pull requests com mais interações possuem mais revisões.