

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS E INFORMÁTICA
UNIDADE EDUCACIONAL PRAÇA DA LIBERDADE
Bacharelado em Engenharia de Software

Belle Nerissa Aguiar Elizeu
Felipe Caldas Liduario
Letícia Amanda Franco Gonçalves

RELATÓRIO: CARACTERÍSTICAS DO PROCESSO DE
DESENVOLVIMENTO DE REPOSITÓRIOS POPULARES

Belo Horizonte
2023

Belle Nerissa Aguiar Elizeu
Felipe Caldas Liduario
Letícia Amanda Franco Gonçalves

**RELATÓRIO: CARACTERÍSTICAS DO PROCESSO DE
DESENVOLVIMENTO DE REPOSITÓRIOS POPULARES**

Relatório apresentado o Laboratório 2 - entrega
final, da disciplina Laboratório de
Experimentação de Software.

Professor: José Laerte Pires Xavier

Belo Horizonte

2023

1 INTRODUÇÃO

O relatório apresentado visa analisar características do processo de desenvolvimento de 1000 repositórios do Github mais populares na linguagem java. Essa atividade foi proposta na disciplina de Laboratório de Experimentação de Software, Laboratório 2 - Sprint final.

2 QUESTÕES DE PESQUISA E HIPÓTESES

RQ01. Qual a relação entre a popularidade dos repositórios e as suas características de qualidade?

Métrica: Número de estrelas

Hipótese: Repositórios populares possuem alto engajamento da sua comunidade, ou seja, possui maior tendência de manutenção e melhora na usabilidade, tornando-o um sistema com maior qualidade. Podemos deduzir que o conteúdo contido nele foi de grande qualidade e utilidade para os usuários do GitHub. Portanto:

Hipótese (Estrelas x CBO): Considera-se que quanto mais popular é o repositório, menor é o CBO.

Hipótese (Estrelas x DIT): Considera-se que quanto mais popular é o repositório, menor é o DIT.

Hipótese (Estrelas x LCOM): Considera-se que quanto mais popular é o repositório, menor é o LCOM.

RQ02. Qual a relação entre a maturidade dos repositórios e as suas características de qualidade?

Métrica: Linhas de código (LOC) e linhas de comentários

Hipótese: Repositórios maduros possuem mais tempo de desenvolvimento e recebem uma quantidade maior de atualizações e contribuições, o que acarreta

numa maior qualidade ao passar dos anos. Considerando, que um sistema maduro possui em média 7 anos de duração, espera-se que a qualidade do produto acompanhe a sua evolução e siga certos aspectos de qualidade. Portanto:

Hipótese (Idade x CBO): Considera-se que quanto maior a maturidade do repositório, menor é o CBO.

Hipótese (Estrelas x DIT): Considera-se que quanto maior a maturidade do repositório, maior é o DIT.

Hipótese (Estrelas x LCOM): Considera-se que quanto maior a maturidade do repositório, maior é o LCOM.

RQ03. Qual a relação entre a atividade dos repositórios e as suas características de qualidade?

Métrica: Número de releases

Hipótese: Repositórios que geralmente possuem um grande volume de atividades tendem a lançar funcionalidades e melhorias com mais frequência, contribuindo diretamente na qualidade do sistema e os índices de qualidade devem refletir essa melhoria. Portanto:

Hipótese (Atividade x CBO): Espera-se que quanto maior a atividade, menor é o CBO.

Hipótese (Atividade x DIT): Espera-se que quanto maior a atividade, menor é o DIT.

Hipótese (Atividade x LCOM): Espera-se que quanto maior a atividade, menor é o LCOM.

RQ04. Qual a relação entre o tamanho dos repositórios e as suas características de qualidade?

Métrica: Idade (em anos) de cada repositório coletado

Hipótese: Considerando que quanto maior o tamanho de um repositório, maior é a

sua durabilidade, mesmo que o tamanho não tenha uma relação direta com a qualidade do repositório, já que um código menor pode ser escrito melhor do que um maior. Entretanto, códigos com um grande tamanho e com bons comentários tendem a ser mais fáceis de realizar manutenções e consequentemente considera-se que é um sistema de qualidade. Portanto:

Hipótese 1 (Tamanho x CBO): Considera-se que quanto menor o tamanho do repositório, menor é o CBO.

Hipótese 2 (Tamanho x DIT): Considera-se que quanto menor o tamanho do repositório, menor é o DIT.

Hipótese 3 (Tamanho x LCOM): Considera-se que quanto menor o tamanho do repositório, menor é o LCOM.

2.1 Repositórios com valores nulos retornados pela ferramenta ck:

- react-native-camera/react-native-camera
- unofficial-openjdk/openjdk

3 METODOLOGIA

Para examinar as suposições, foi criado um código em Python que executa uma pesquisa em GraphQL dos 1000 repositórios mais populares escritos em Java. Esses repositórios foram copiados e avaliados utilizando a ferramenta CK. Posteriormente, um relatório de análise foi produzido contendo a mediana das métricas de qualidade CBO, DIT e LCOM de cada um desses repositórios. Com base nas medianas, foram gerados gráficos Scatter Plot para verificar cada uma das hipóteses introduzidas na **seção 2**. A coleta dos dados ocorreu no dia 17/03/2022. O script em código-fonte e o arquivo csv estão disponíveis no repositório: <https://github.com/BelleNerissa/LAB6s2>.

Para responder às questões de pesquisa, o grupo utilizou uma metodologia onde analisamos a métrica de cada questão e utilizando uma lógica relacionado a métrica em questão com pontos que fazem do software um produto de qualidade como, por exemplo, a usabilidade. Foi utilizado da dedução para assumir alguns fatores e assim

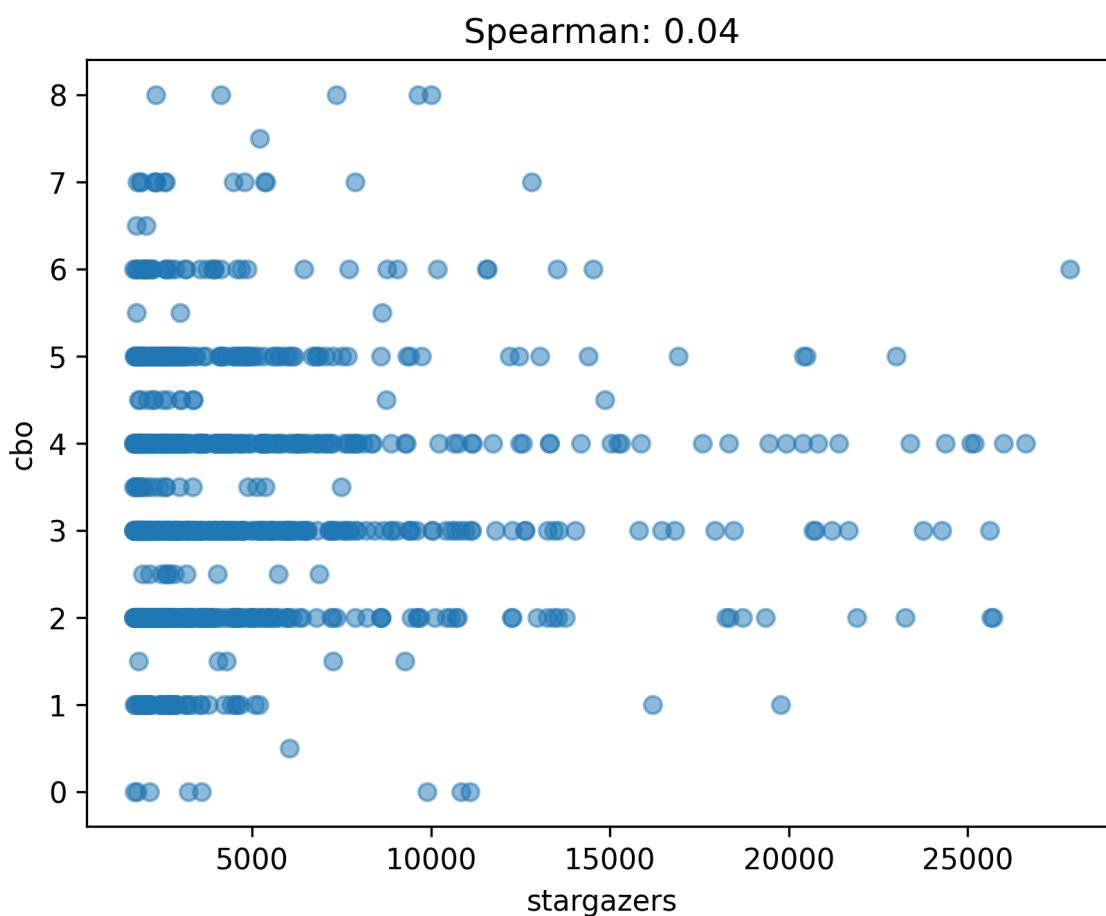
obter o resultado das hipóteses iniciais levantadas para as perguntas de pesquisa.

4 RESULTADOS OBTIDOS

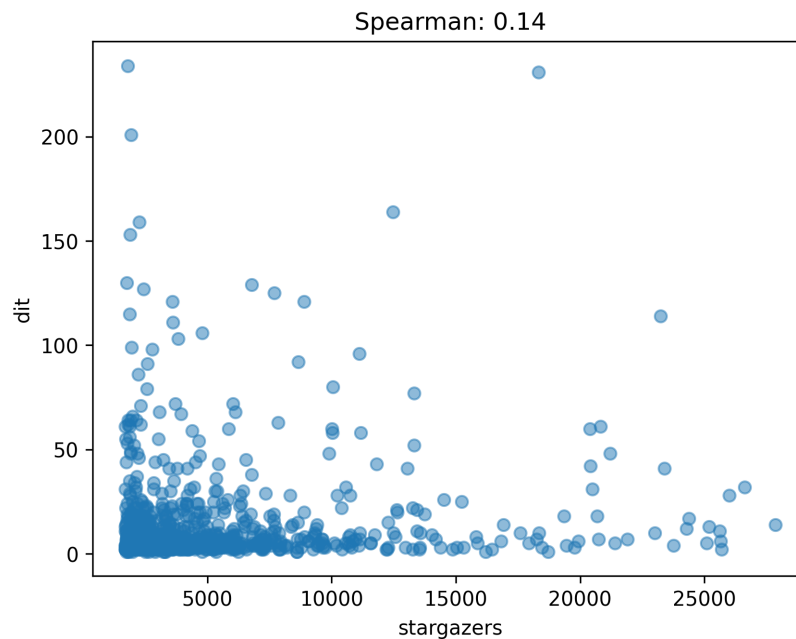
Esta seção tem o objetivo de apresentar a apuração dos resultados executados no projeto e análise geral, de modo a comparar com as hipóteses levantadas.

RQ1. Popularidade:

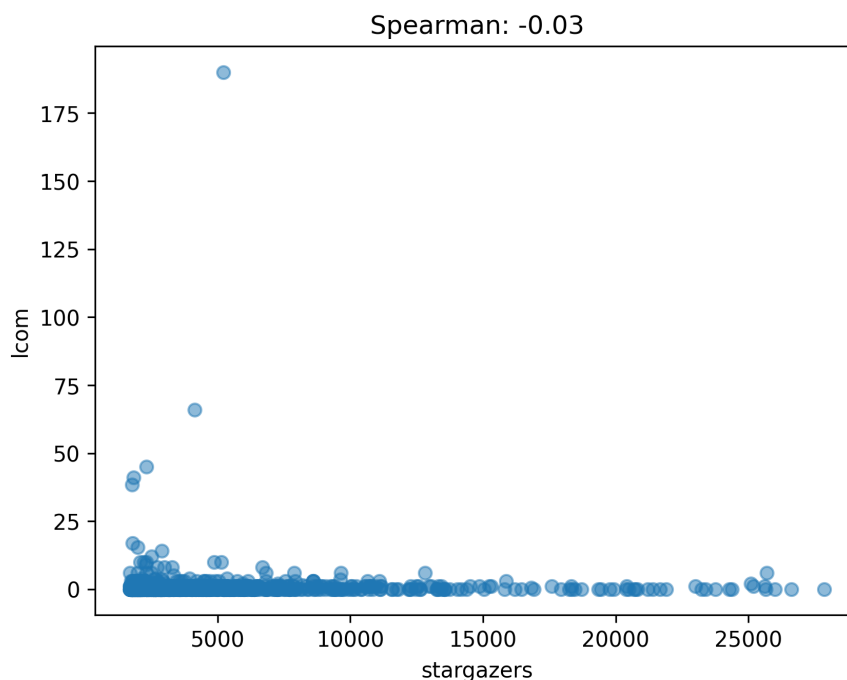
Estrelas x CBO: Observa-se que existe uma pequena relação do coeficiente de Spearman (0.04), entre os aspectos analisados. Entretanto, não podemos assumir que existe uma correlação direta e confirmar que quanto maior a quantidade de estrelas menor é o CBO. Então a hipótese inicial foi refutada.



Estrelas x DIT: Podemos observar que possui relação do coeficiente de Spearman (0.14), boa parte dos repositórios possui populares possuem um DIT baixo, ou seja, existe uma correlação de que quanto mais popular um repositório, menor é o DIT. Portanto, a hipótese inicial foi confirmada.

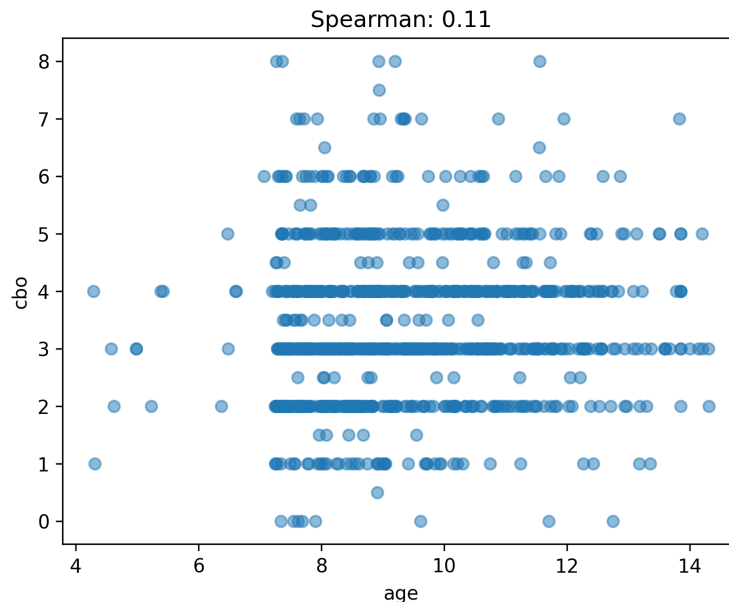


Estrelas x LCOM: O resultado obtido não permite obter uma relação significativa entre os dois pontos, já que o coeficiente de Spearman apresenta uma correlação negativa muito fraca (-0.03). Portanto, não podemos assumir que existe uma correlação direta e confirmar que quanto mais popular é o repositório, menor é o LCOM. Então a hipótese inicial foi refutada.

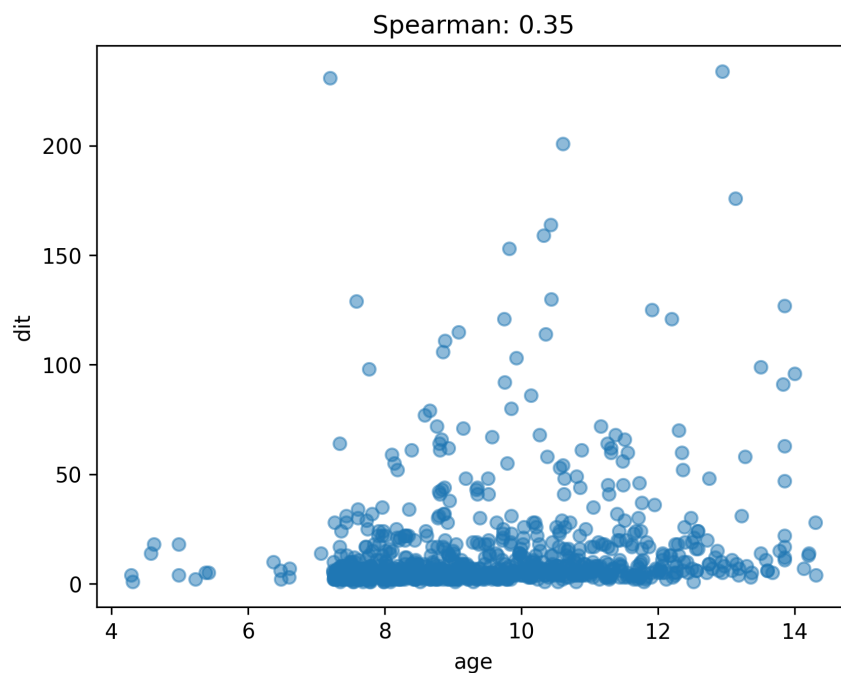


RQ2. Maturidade:

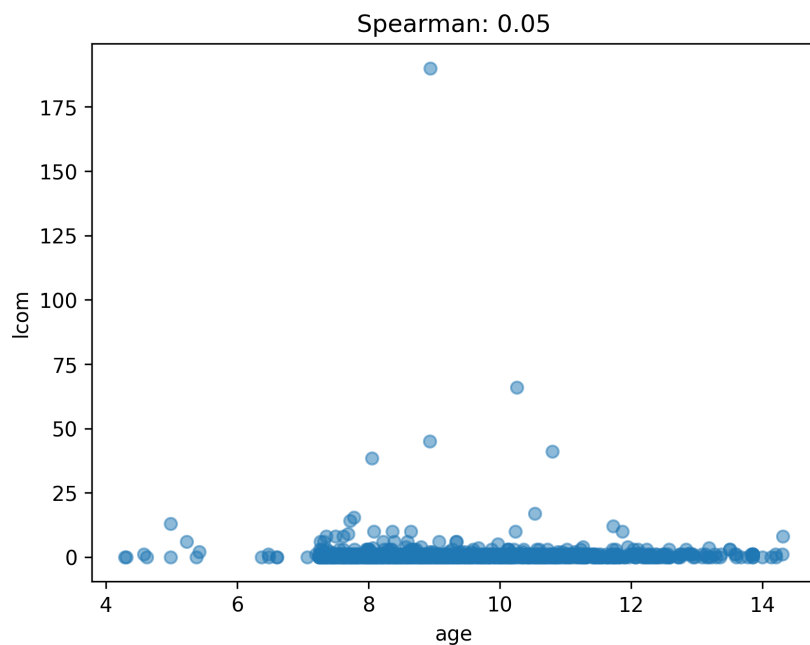
Maturidade x CBO: Observa-se no resultado obtido que repositórios com maior a maturidade não necessariamente terá um CBO menor. O coeficiente de Spearman apresenta uma correlação de (0.11) entre os dados. Dessa forma, a hipótese inicial foi refutada.



Maturidade x DIT: O coeficiente de Spearman apresenta uma correlação de (0.35) entre os dados. Observa-se no resultado obtido que grande parte dos repositórios com maior a maturidade possui menor DIT, então a hipótese inicial foi confirmada. Porém, não necessariamente em todos os casos terá um DIT menor.

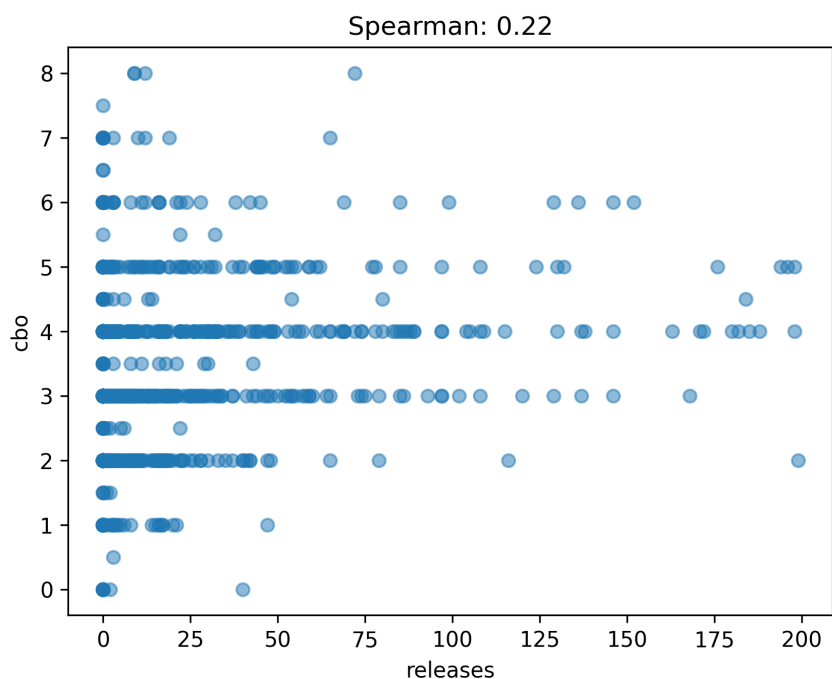


Maturidade x LCOM: O resultado obtido apresenta uniformemente o LCOM baixo, apresentando pequena relação do coeficiente de Spearman (0.05), o que indica que pode não existir uma correlação direta entre quanto maior a maturidade do repositório, menor ser o LCOM.

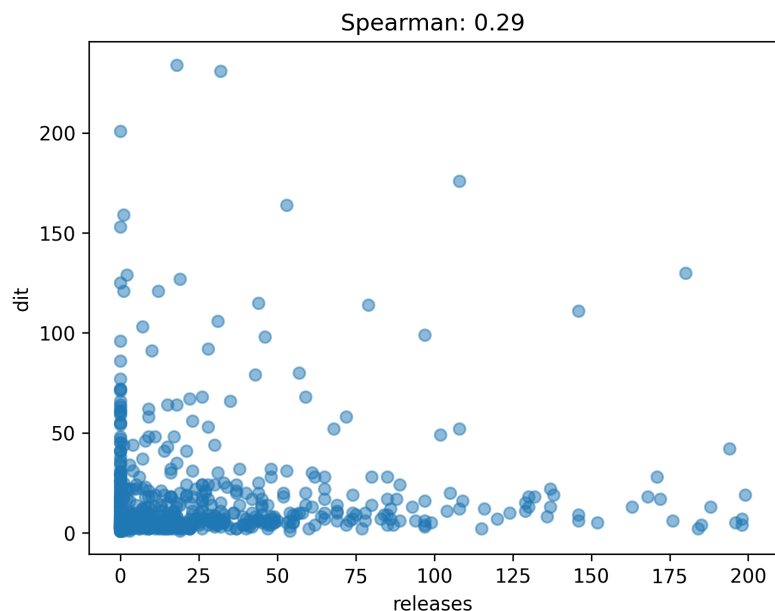


RQ3. Atividade:

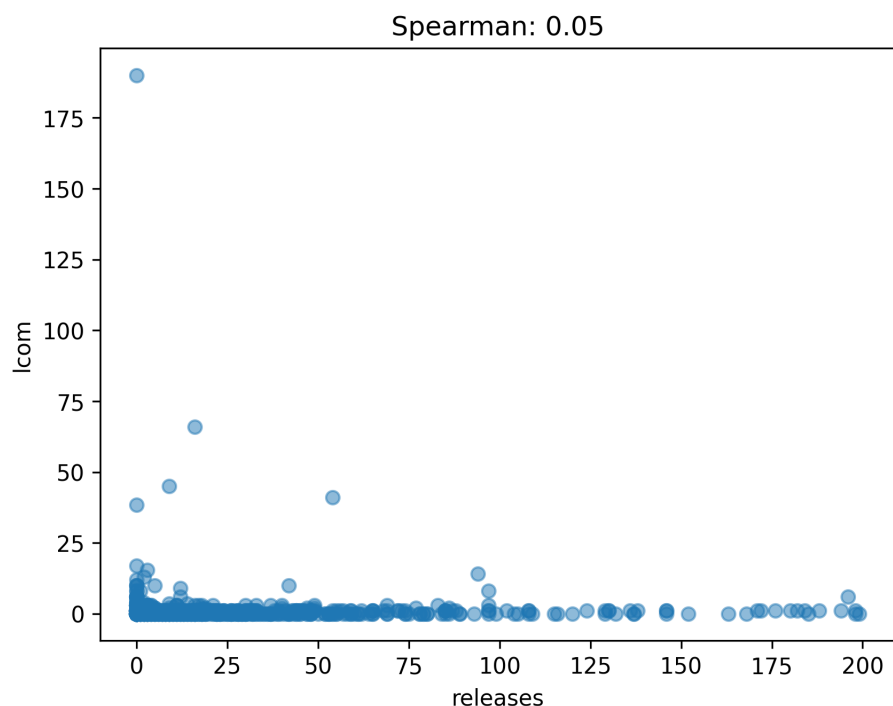
Atividade x CBO: O resultado obtido não deixa claro uma correlação entre os dois aspectos. Desta forma, podemos negar que quanto maior a atividade, menor o CBO. O coeficiente de Spearman apresenta uma correlação positiva de (0.22) entre os dados. Então a hipótese inicial foi refutada.



Atividade x DIT: O coeficiente de Spearman apresenta uma correlação positiva de (0.29). É possível identificar uma relação entre os dois aspectos apresentados, quanto maior a atividade menor é o DIT. Portanto, a hipótese inicial foi confirmada. Entretanto, observa-se que mesmo em casos onde ocorre pouca atividade é possível identificar um DIT baixo.

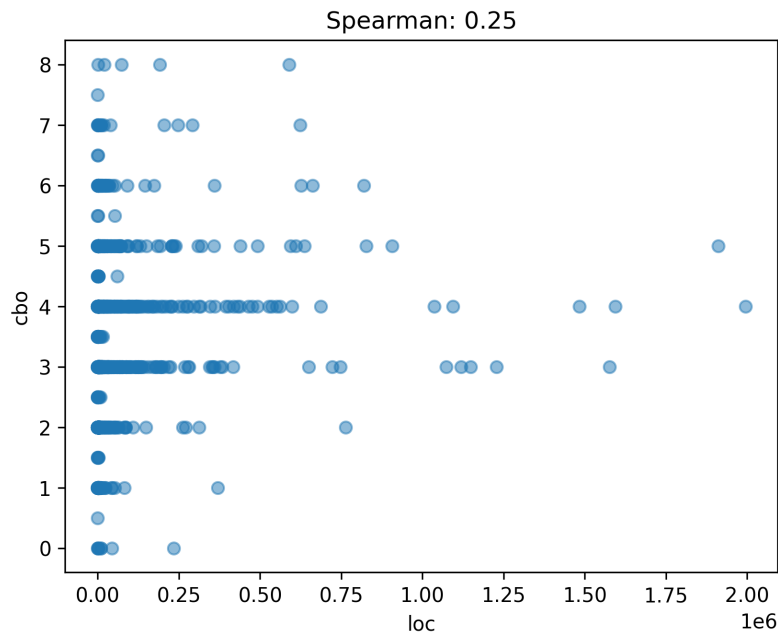


Atividade x LCOM: O resultado obtido não permite uma relação significativa entre a quantidade de atividades obtendo menor LCOM, já que o gráfico apresentado abaixo segue praticamente uniforme. O coeficiente de Spearman apresenta baixa correlação positiva de (0.05) entre os dados.

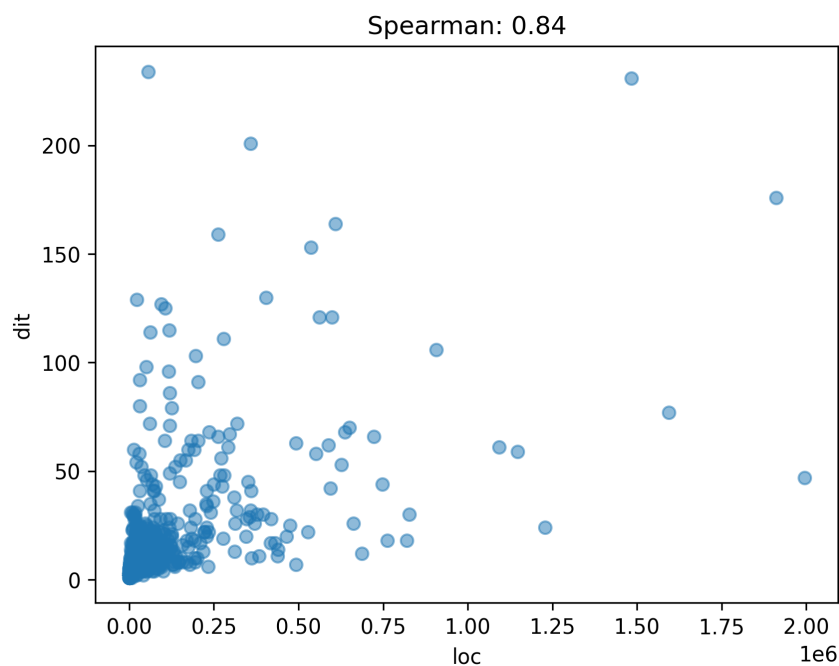


RQ4. Tamanho:

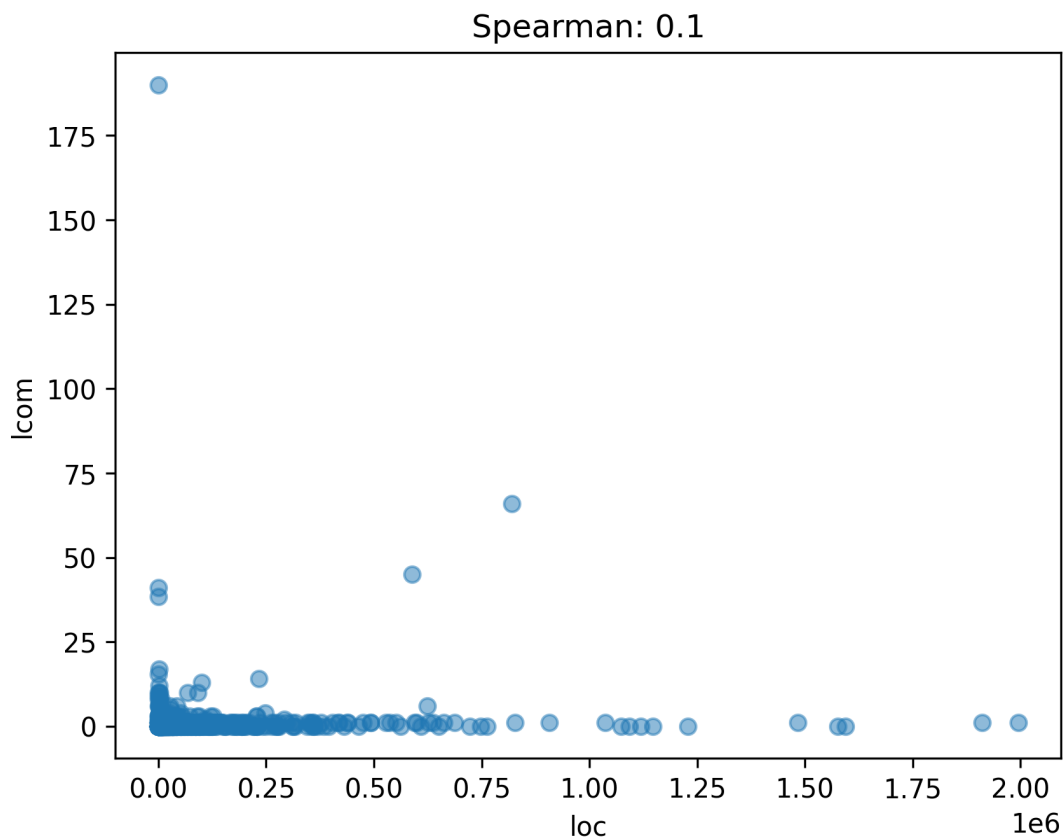
Tamanho x CBO: O coeficiente de Spearman apresenta correlação positiva de (0.25) entre os dados. O resultado obtido apresenta relação entre os dois aspectos. De certa forma, indica que quanto maior o tamanho o CBO fica na média. Portanto, não podemos confirmar a hipótese.



Atividade x DIT: No resultado obtido, é possível identificar que existe uma relação entre os aspectos. Portanto, a análise indica que quanto menor a atividade também é menor o DIT. O coeficiente de Spearman apresenta correlação positiva de (0.84). Portanto, a hipótese inicial foi refutada.



Atividade x LCOM: O coeficiente de Spearman apresenta correlação pouca positiva de (0.1). O resultado obtido não permite uma relação significativa entre a quantidade de tamanho obtendo menor LCOM, já que o gráfico apresentado abaixo segue praticamente de forma uniforme. Portanto, não podemos confirmar a hipótese.



5 CONCLUSÃO

As bases de dados investigadas geralmente não apresentaram resultados robustos, dificultando a avaliação das hipóteses içadas. Percebe-se que a diversidade existente entre elas, tanto em relação à implementação quanto à qualidade. Embora seja desconcertante diversas perguntas levantadas sem obter respostas claras e objetivas, a busca leva a criar possíveis padrões, representando essa situação a um desafio para aprimorar em futuras minerações.