

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

**Arthur Bicalho Lana Corrêa Fernandes
João Victor Tadeu Chaves Frois
Lucas Gabriel Padrão Rezende**

Trabalho laboratório de Experimentação de Software

Belo Horizonte
2022

Arthur Bicalho Lana Corrêa Fernandes
João Victor Tadeu Chaves Frois
Lucas Gabriel Padrão Rezende

Trabalho laboratório de Experimentação de Software

Trabalho apresentado na disciplina Laboratório
de Experimentação de Software.

Professor: José Laerte Pires Xavier

Belo Horizonte
2022

1 INTRODUÇÃO

O relatório apresentado tem como objetivo analisar as características de qualidade de sistemas Java utilizando a ferramenta de métricas de software CK. Essa atividade foi proposta na disciplina de Laboratório de Experimentação de Software.

1.1 Hipóteses

RQ 01. Qual a relação entre a popularidade dos repositórios e as suas características de qualidade?

Hipótese 1 (Estrelas x CBO): Os repositórios mais populares acabam tendendo a serem mais atrativos para colaboradores externos, visto que, espera-se que a junção acabe sendo baixa para que os desenvolvedores possam atuar de forma eficaz no mesmo código. Dessa forma, se espera que quanto mais popular é o repositório, menor é o CBO.

Hipótese 2 (Estrelas x DIT): Repositórios populares são importantes e precisam ser menos complexos para que assim facilite na contribuição da comunidade. Dessa forma, espera-se que quanto mais popular o repositório, menor é o DIT.

Hipótese 3 (Estrelas x LCOM): Repositórios populares tem uma manutenção e reuso para assim facilitar a distribuição de esforços entre os desenvolvedores e assim melhorar a sua qualidade. Assim, quanto mais popular é o repositório, menor é o LCOM.

RQ 02. Qual a relação entre a maturidade dos repositórios e as suas características de qualidade ?

Hipótese 1 (Idade x CBO): Repositórios mais maduros acabam tendo classes menos acopladas, visto que, passaram por mais evoluções. Dessa forma, quanto maior a maturidade do repositório, menor é o CBO.

Hipótese 2 (Estrelas x DIT): Repositórios mais maduros acabam possuindo mais classes e são mais complexos. Dessa forma, quanto maior a maturidade do repositório, maior é o DIT.

Hipótese 3 (Estrelas x LCOM): Repositórios mais maduros são capazes de dificultar a refatoração conforme o seu crescimento, podendo ficar menos lógico. Dessa forma, quanto maior a maturidade do repositório, maior é o LCOM.

RQ 03. Qual a relação entre a atividade dos repositórios e as suas características de qualidade?

Hipótese 1 (Atividade x CBO): Repositórios que são mais ativos tendem a ser mais atualizados e por consequência disso tendem a ter o seu código melhorado nesse processo. Assim, os índices de qualidade acabam refletindo nesta melhoria. Portanto, acaba-se esperando que quanto maior atividade, menor é o CBO.

Hipótese 2 (Atividade x DIT): Repositórios que são mais ativos tendem a ser mais atualizados e por consequência disso tendem a ter o seu código melhorado nesse processo. Assim, os índices de qualidade acabam refletindo nesta melhoria. Portanto, acaba-se esperando que quanto maior atividade, menor é o DIT.

Hipótese 3 (Atividade x LCOM): Repositórios que são mais ativos tendem a ser mais atualizados e por consequência disso tendem a ter o seu código melhorado nesse processo. Assim, os índices de qualidade acabam refletindo nesta melhoria. Portanto, acaba-se esperando que quanto maior atividade, menor é o LCOM.

RQ 04. Qual a relação entre o tamanho dos repositórios e as suas características de qualidade?

Hipótese 1 (Tamanho x CBO): Repositórios que tem um tamanho menor tendem a serem menos acoplados. Dessa forma, espera-se que quanto menor o tamanho do repositório, menor é o CBO.

Hipótese 2 (Tamanho x DIT): Repositórios que tem um tamanho menor tendem a serem menos classes e assim o número de heranças é menor. Dessa forma, espera-se que quanto menor o tamanho do repositório, menor é o DIT.

Hipótese 3 (Tamanho x LCOM): Repositórios que tem um tamanho menor tendem a serem menos mais fáceis de realizar manutenção e refatorar o seu código.. Dessa forma, espera-se que quanto menor o tamanho do repositório, melhor é o LCOM.

2 METODOLOGIA

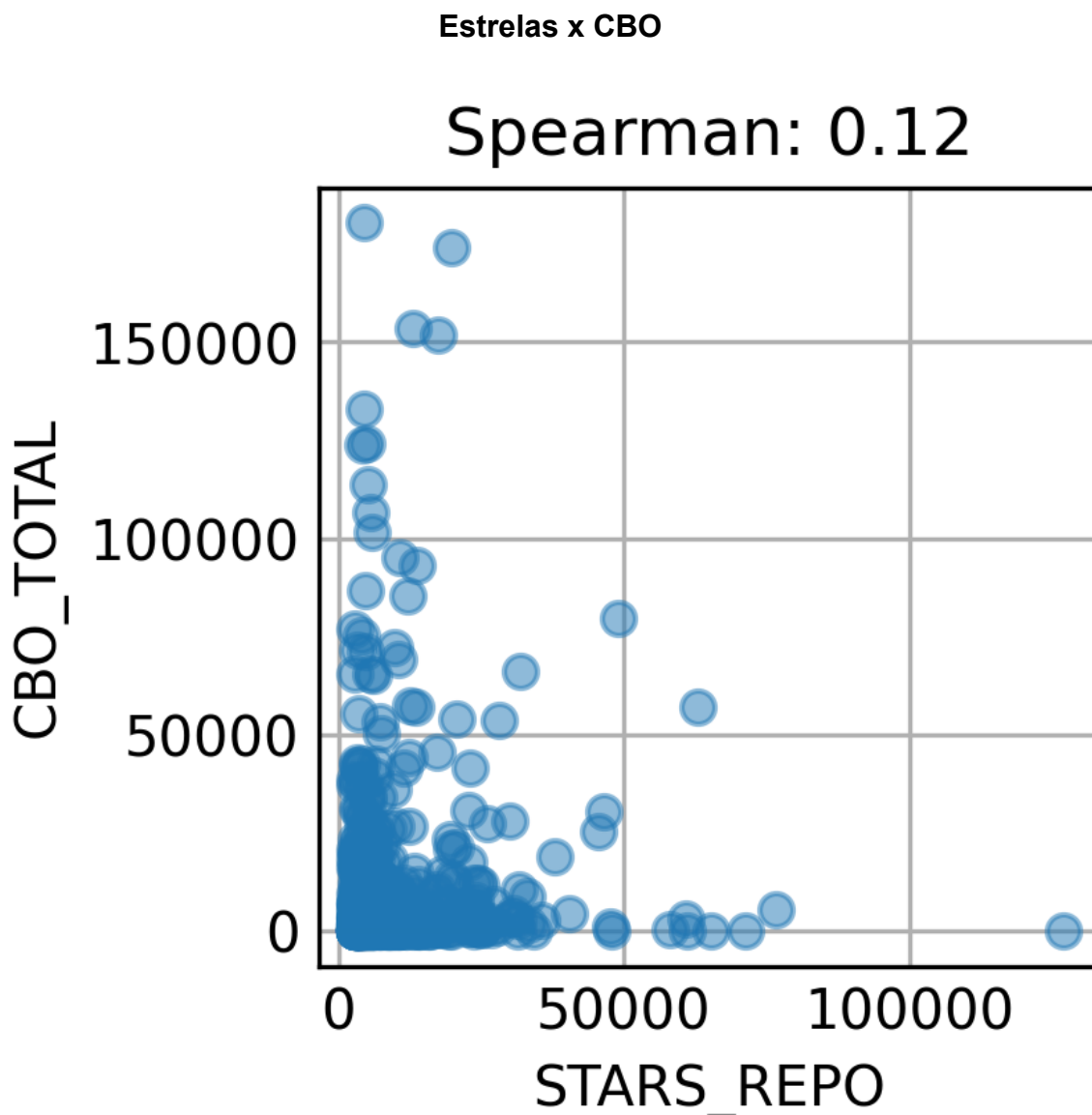
Para realizar a análise das hipóteses, foi desenvolvido um script em Python que realiza uma consulta em GraphQL dos 1000 repositórios mais populares em Java, que foram clonados e analisados na ferramenta CK.

Após a seleção dos repositórios, foi gerado um arquivo de análise com a mediana das métricas de qualidade CBO, DIT e LCOM de cada um deles. Com os valores das medianas, foram gerados gráficos Scatter Plot para analisar cada uma das hipóteses apresentadas na **seção 1.1**.

A data da coleta foi durante o período de 22/03/2022 a 26/03/2022. O código-fonte do script e o arquivo csv está contido no repositório: <https://github.com/padraorezende/Lab>

3 RESULTADOS OBTIDOS E ANÁLISES DOS RESULTADOS

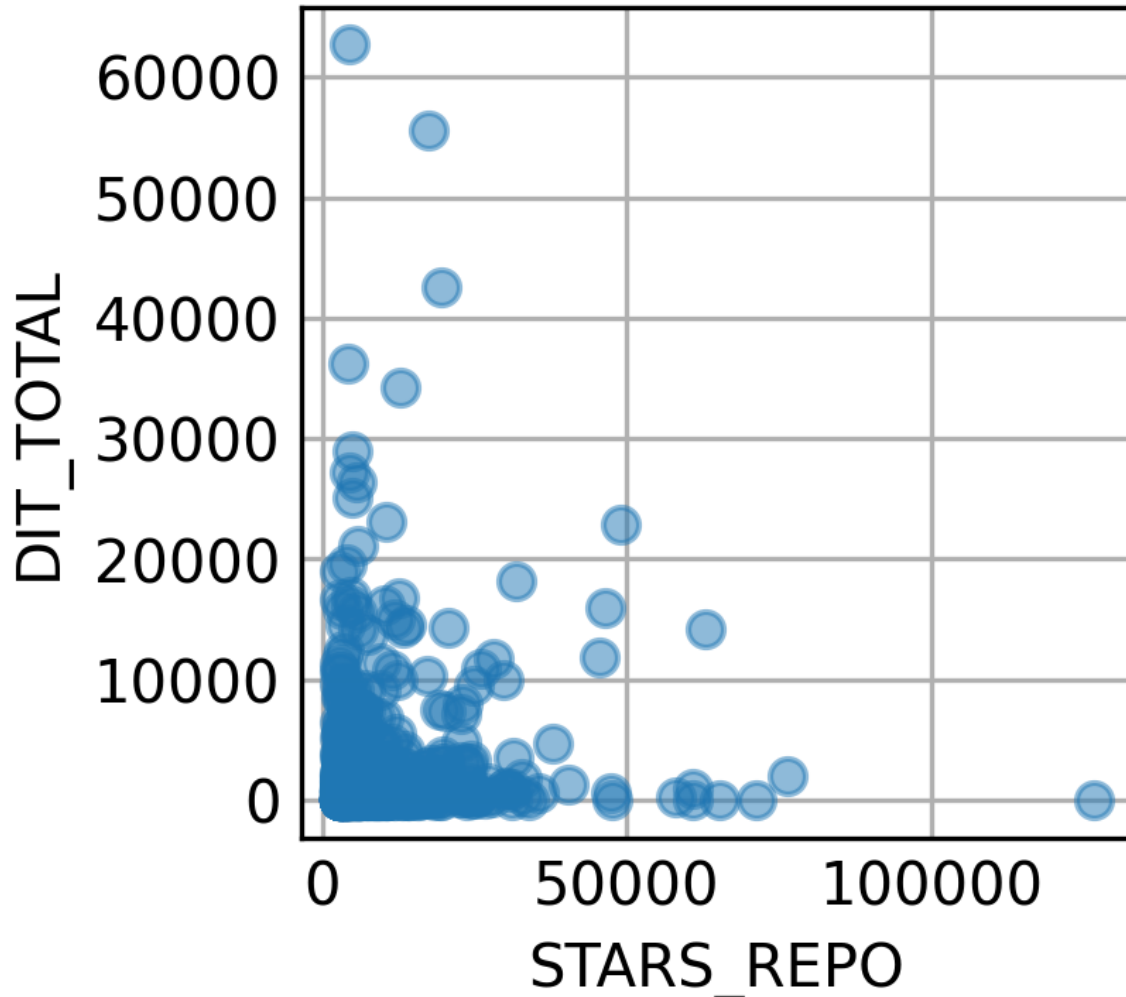
RQ1. Popularidade



A relação que percebemos entre as estrelas e o CBO, é que a grande maioria está próximo de zero não ultrapassando os 50000. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.12). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto mais popular é o repositório, menor é o CBO.

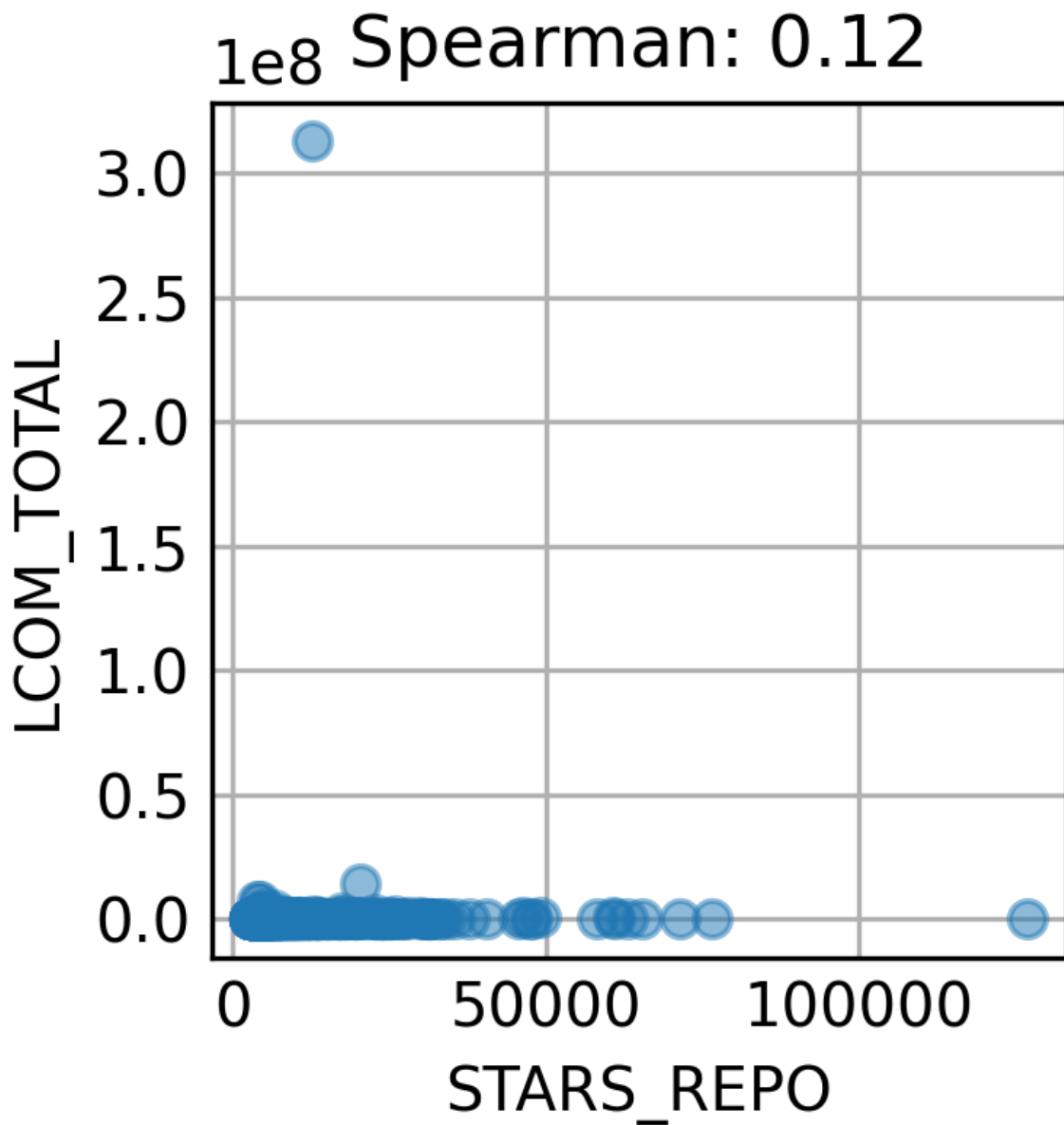
Estrelas x DIT

Spearman: 0.13



Pode-se perceber que a grande maioria dos resultados obtidos das Estrelas x DIT está próximo do zero e não ultrapassando o primeiro quadrante também.. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.13). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto mais popular o repositório, menor é o DIT.

Estrelas x LCOM

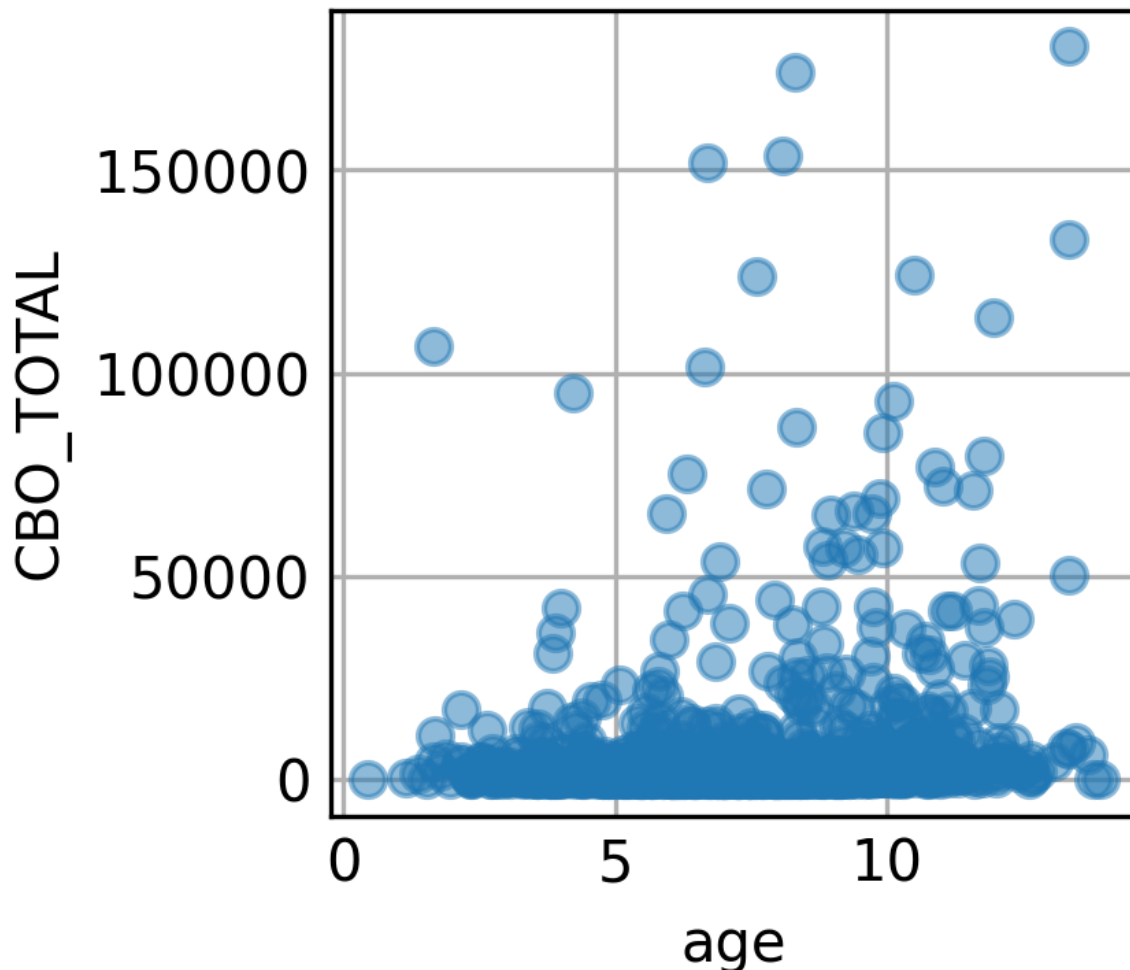


Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.12). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto mais popular é o repositório, menor é o LCOM.

RQ2. Maturidade

Maturidade x CBO

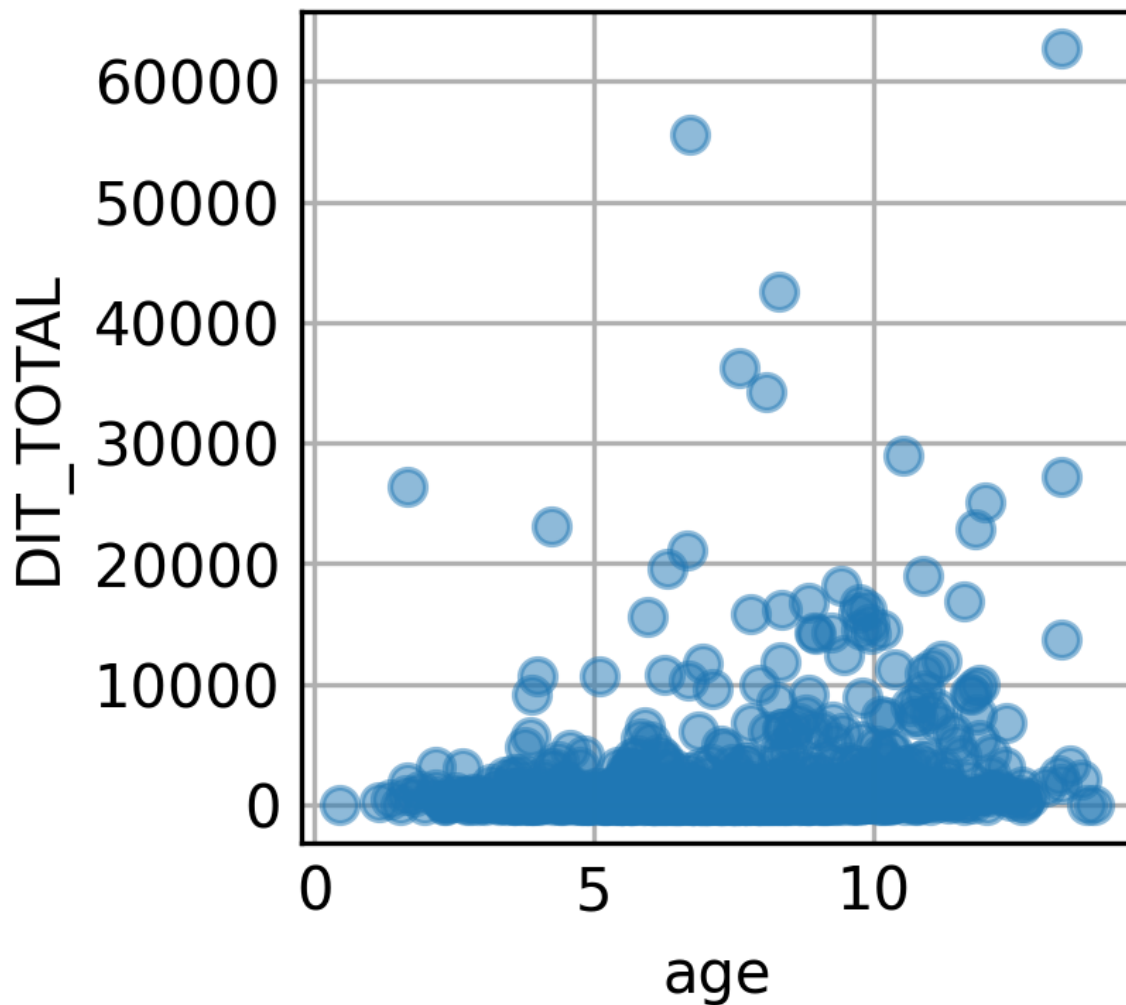
Spearman: 0.18



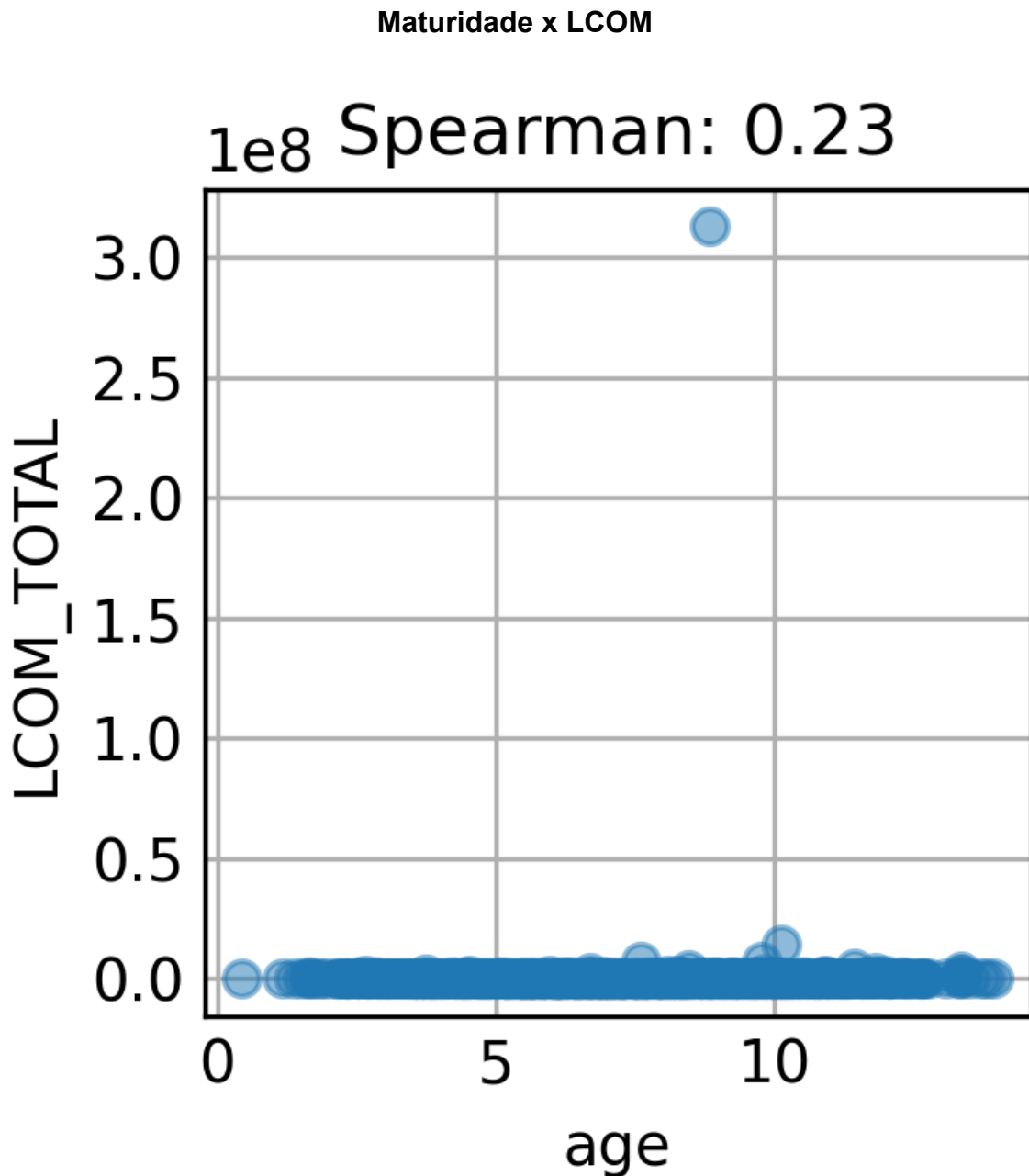
Pode-se ver no gráfico a Maturidade x CBO que a grande maioria não ultrapassa o quadrante 50000, mas chega a ter uma idade bastante variável entre o 0 a 10. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.18). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior atividade, menor é o CBO.

Maturidade x DIT

Spearman: 0.2

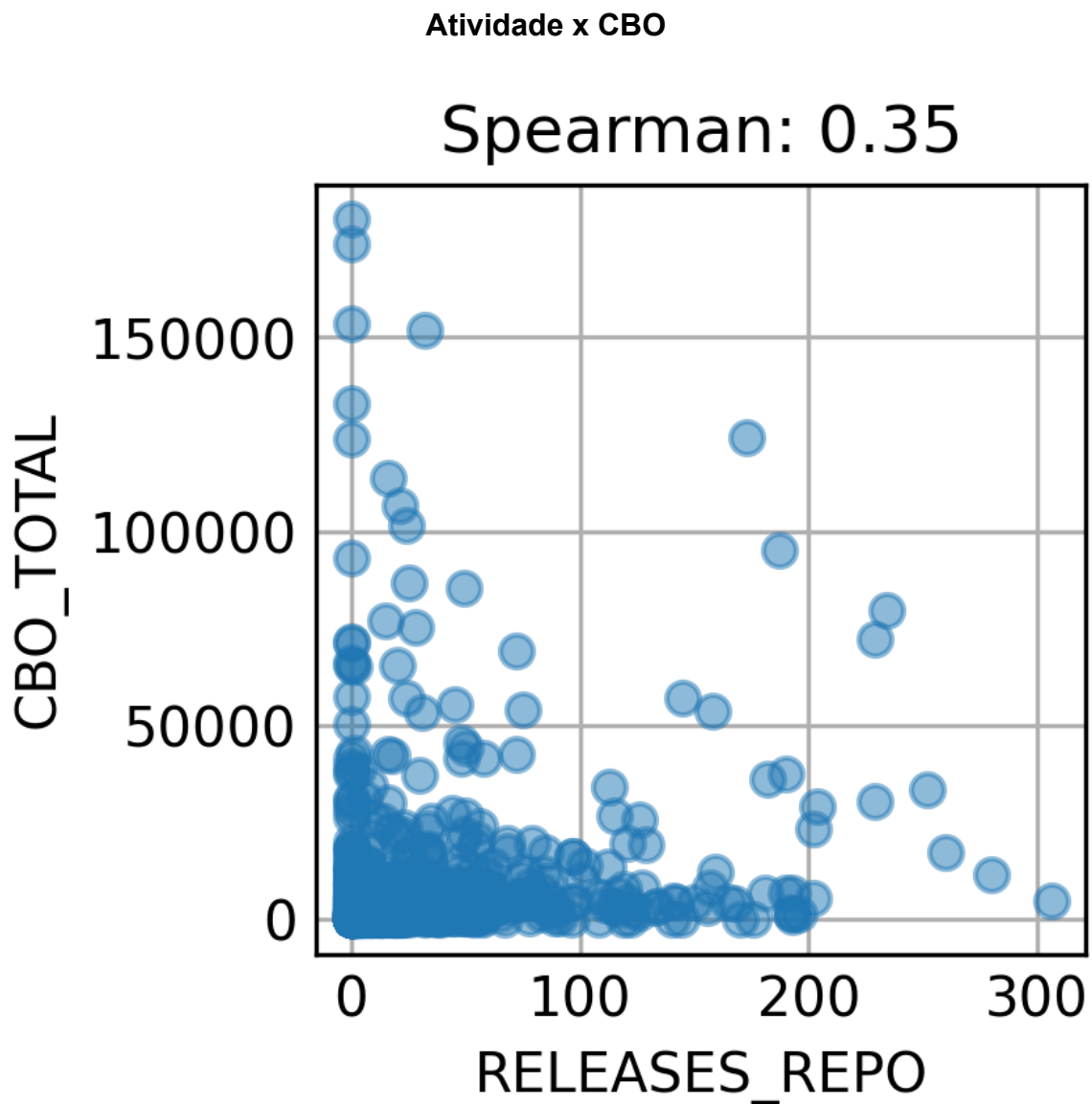


Pode-se ver no gráfico a Maturidade x DIT que a grande maioria não ultrapassa o quadrante 10000, mas chega a ter uma idade bastante variável entre 0 a 10. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.2). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior a maturidade do repositório, maior é o DIT.



Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.23). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior atividade, menor é o LCOM.

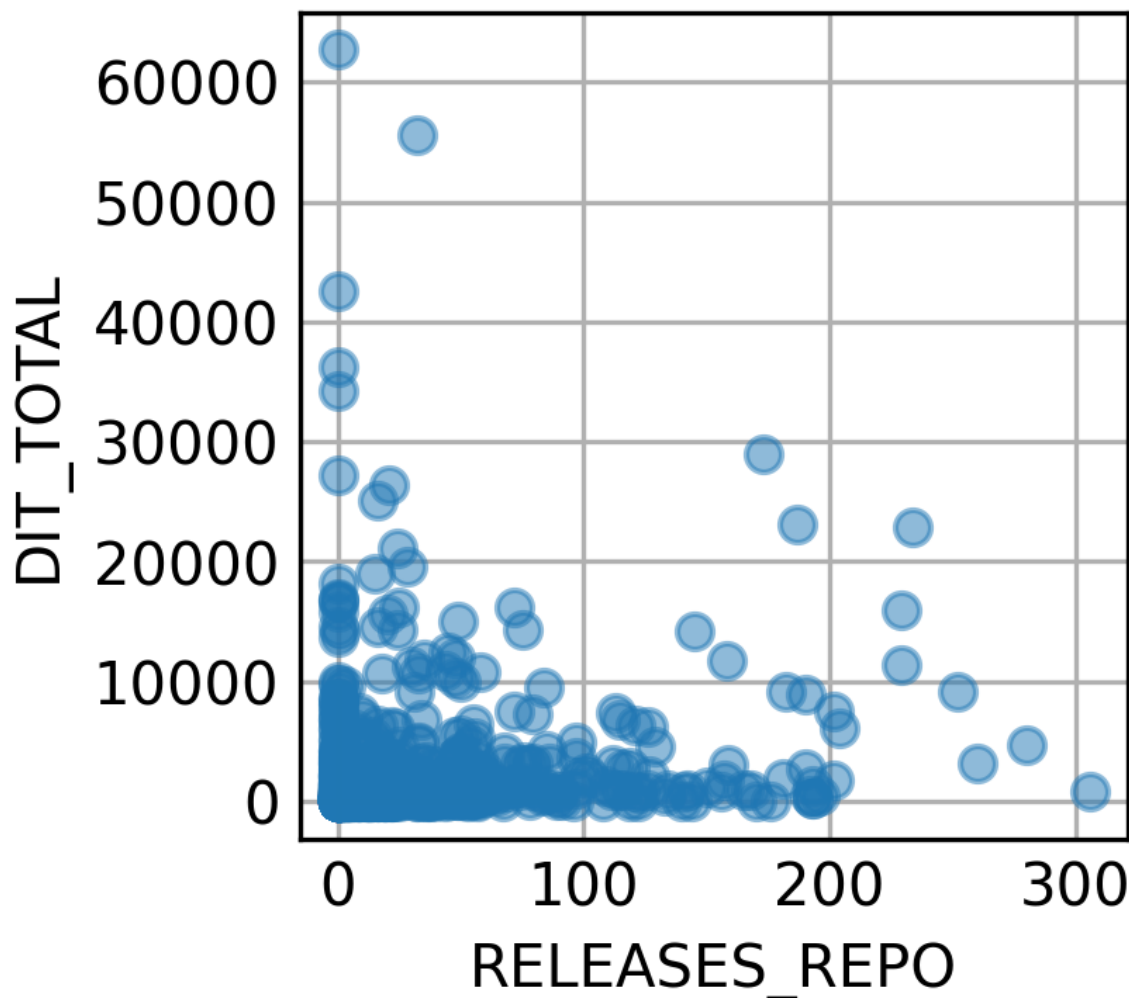
RQ3. Atividade



Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.35). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior atividade, menor é o CBO.

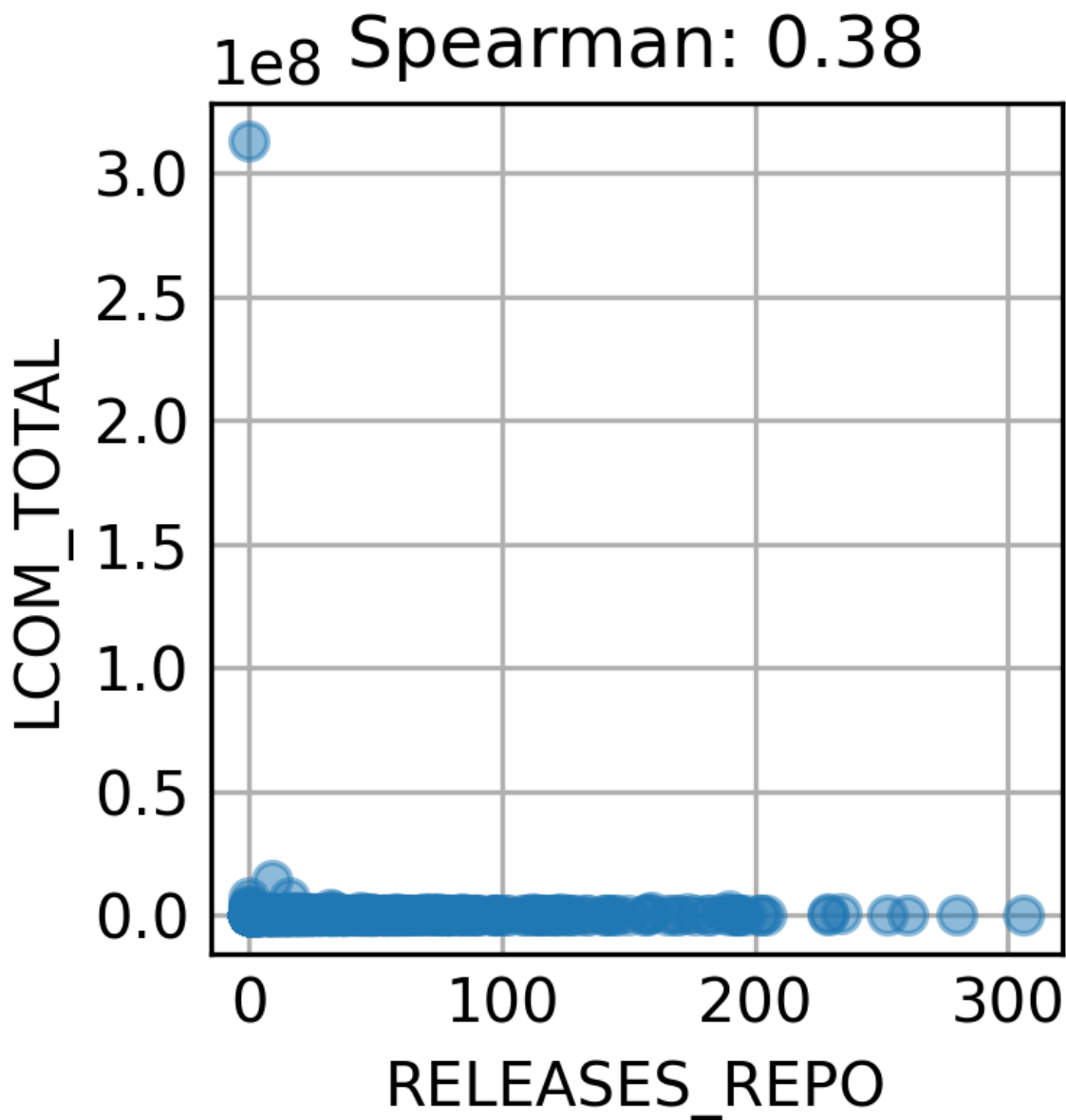
Atividade x DIT

Spearman: 0.32



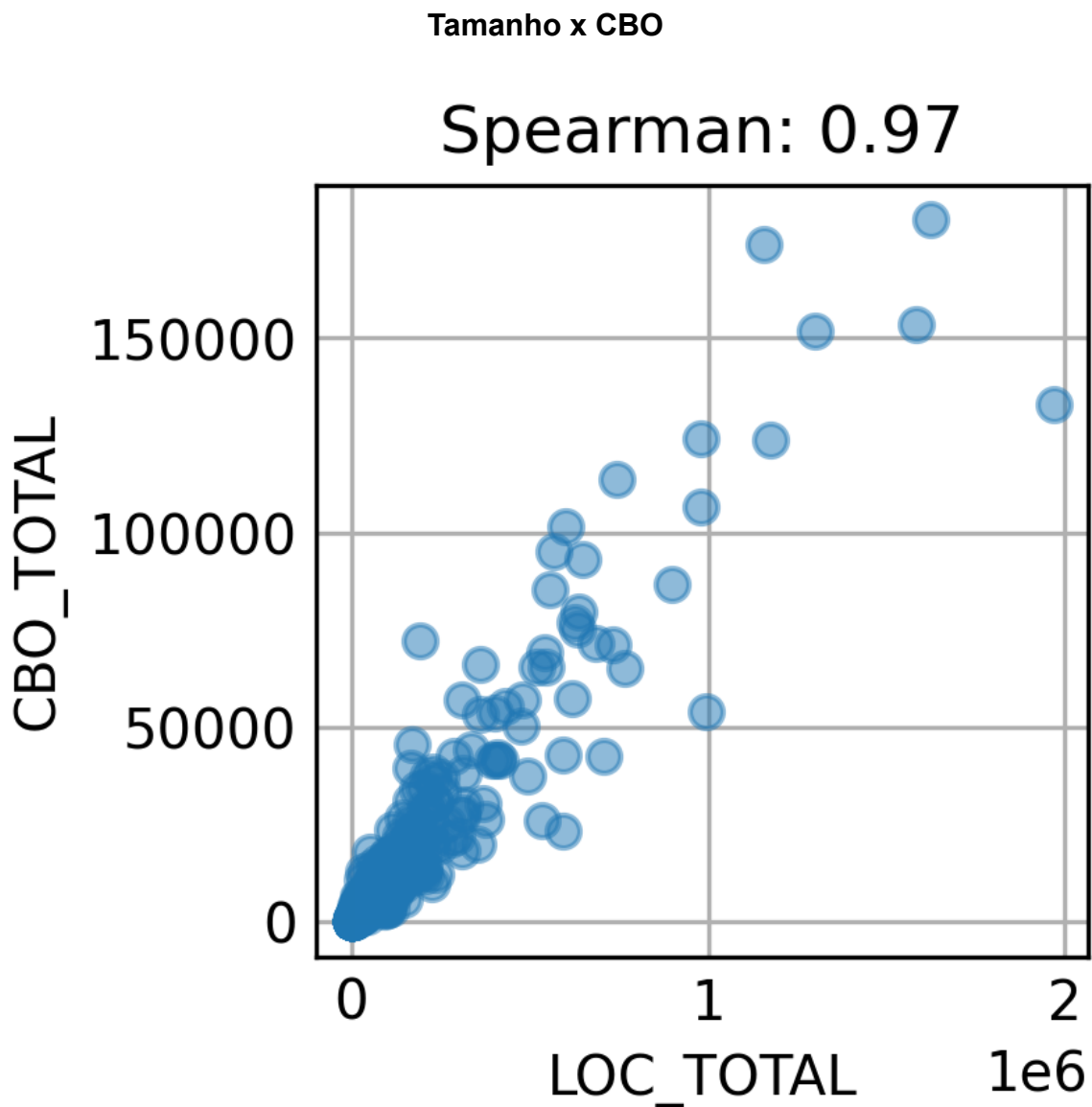
Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.32). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior atividade, menor é o DIT.

Atividade x LCOM



Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.38). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto maior atividade, menor é o LCOM.

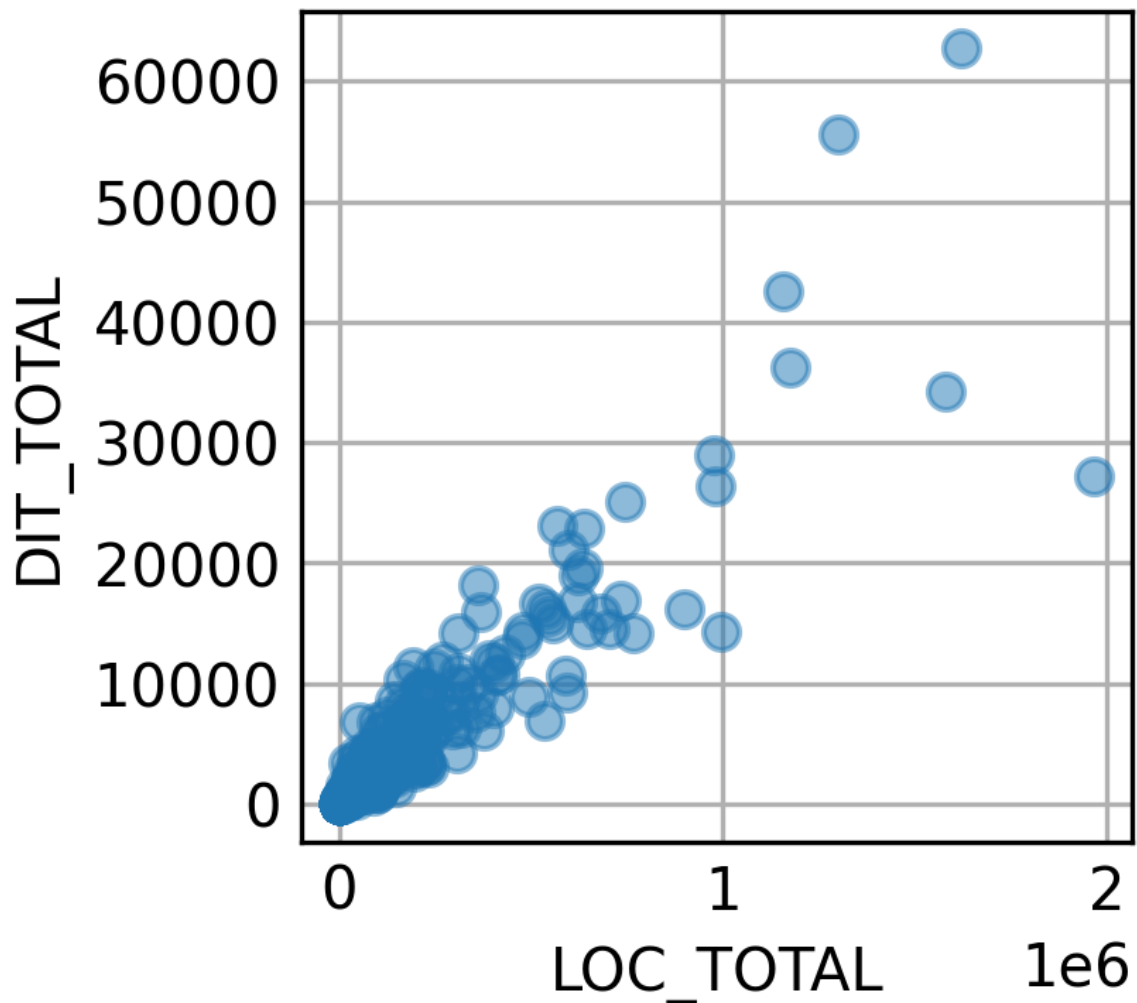
RQ4. Tamanho



Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.97). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto menor o tamanho do repositório, menor é o CBO.

Tamanho x DIT

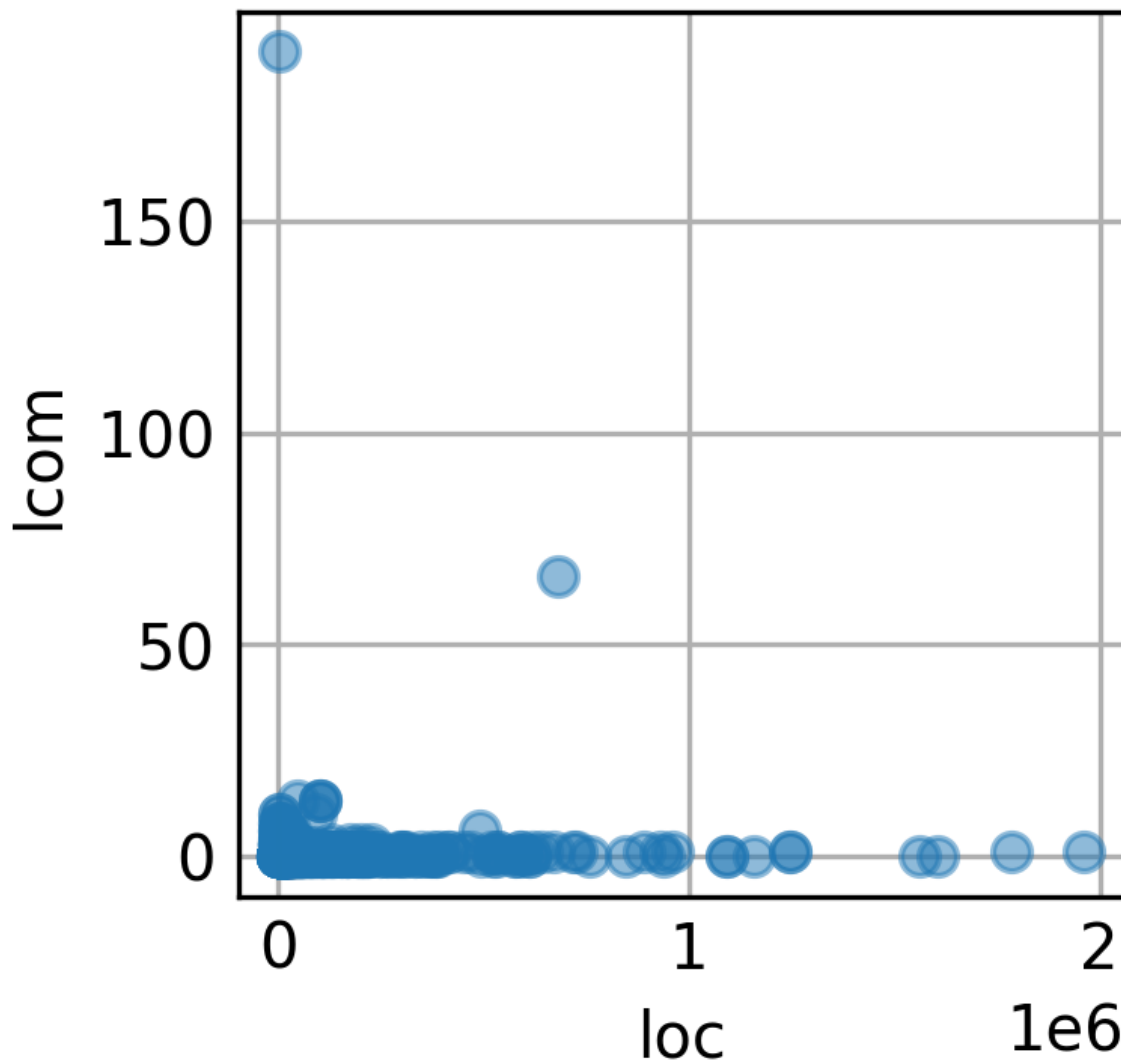
Spearman: 0.97



O Tamanho x DIT é possível de verificar que a grande maioria está localizada no primeiro quadrante entre 10000 DIT e entre o 0 a 1 do LOC. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.97). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto menor o tamanho do repositório, menor é o DIT.

Tamanho x LCOM

Spearman: 0.12



Não é possível obter uma relação significativa entre os dois aspectos. O coeficiente de Spearman calculado no gráfico mostra que a correlação positiva está muito fraca (0.12). Dessa maneira, não pode-se assumir que existe uma correlação totalmente direta e confirmar a hipótese inicial que era: quanto menor o tamanho do repositório, melhor é o LCOM.

Conclusão

A conclusão que o grupo chegou ao final deste trabalho é que os repositórios analisados ocasionalmente mostraram algum padrão com força o suficiente para que se fosse embasado ou negado as hipóteses levantadas, levando assim a diversidade dos mesmos, com relação a implementação e qualidade.

Ao finalizar todos os gráficos o grupo ficou decepcionado com os resultados alcançados, visto que, tendo mais dúvidas do que certezas gera incertezas nos padrões aplicados, contudo, isso mostra as melhorias que devem ser feitas para as minerações futuras, dando assim, uma motivação a mais para todos os integrantes.