

Laboratório 2

Análise de Métricas de Qualidade em Repositórios Java Open-Source

Pedro Henrique Moreira
Lucca
Jhonata Dias



Contexto e Motivação

- Qualidade impacta confiabilidade, manutenção e evolução de software.
- Em Java OSS, métricas CK (CBO, DIT, LCOM) avaliam acoplamento, hierarquia, coesão.
- Investigar como atributos do projeto (popularidade, idade, releases, LOC) se relacionam com qualidade estrutural.



Questões de Pesquisa

RQ01: Popularidade × Qualidade (CBO, DIT, LCOM)

RQ02: Maturidade (idade) × Qualidade

RQ03: Atividade (releases) × Qualidade

RQ04: Tamanho (LOC) × Qualidade

Métricas Utilizadas

Processo:

- **Popularidade:** estrelas (stargazers)
- **Atividade:** n° de releases
- **Maturidade:** idade em anos
- **Tamanho:** LOC

Qualidade (CK):

- CBO (acoplamento), DIT (profundidade de herança), LCOM (falta de coesão)

Amostra e Coleta

- 1.000 repositórios Java populares (GitHub GraphQL).
- Métricas de processo coletadas na mineração (exceto LOC).
- Ferramenta CK para extrair CBO, DIT, LCOM por classe.
- Agregação por repositório: média aritmética.

Tratamento de Dados (Robustez)

- Remoção de outliers acima do 96º percentil.
- Trimmed mean (média truncada) para mitigar extremos (Tukey/Huber).
- Ex.: LCOM dp ~1700 → ~36, distribuição mais estável.

Análise Estatística e Gráficos

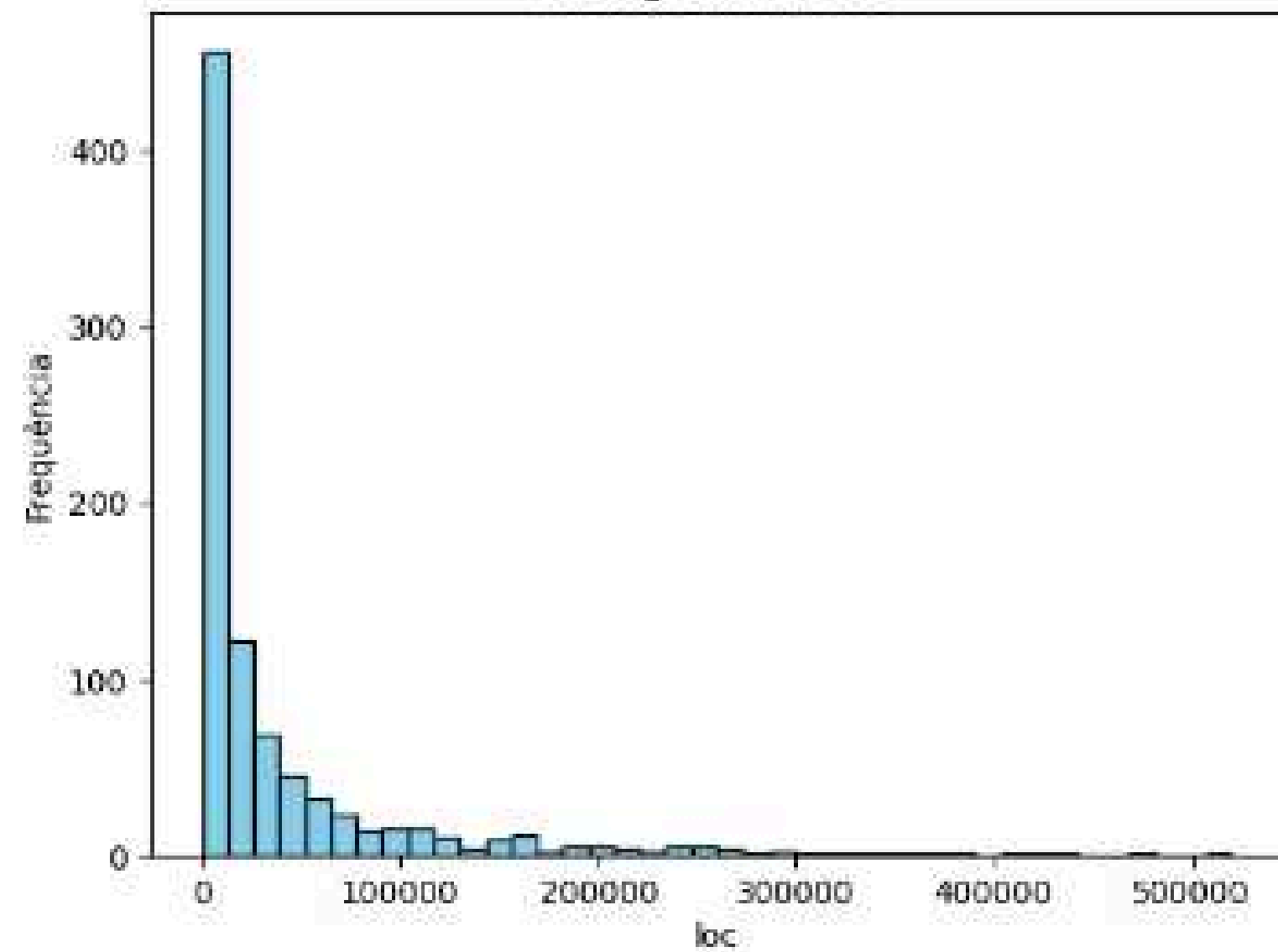
- **Medidas:** média, mediana, desvio-padrão.
- **Correlações:** Pearson e Spearman.
- **Gráficos:** histogramas, boxplots, scatter, heatmaps de correlação.

Sumário dos Resultados (Estatísticas)

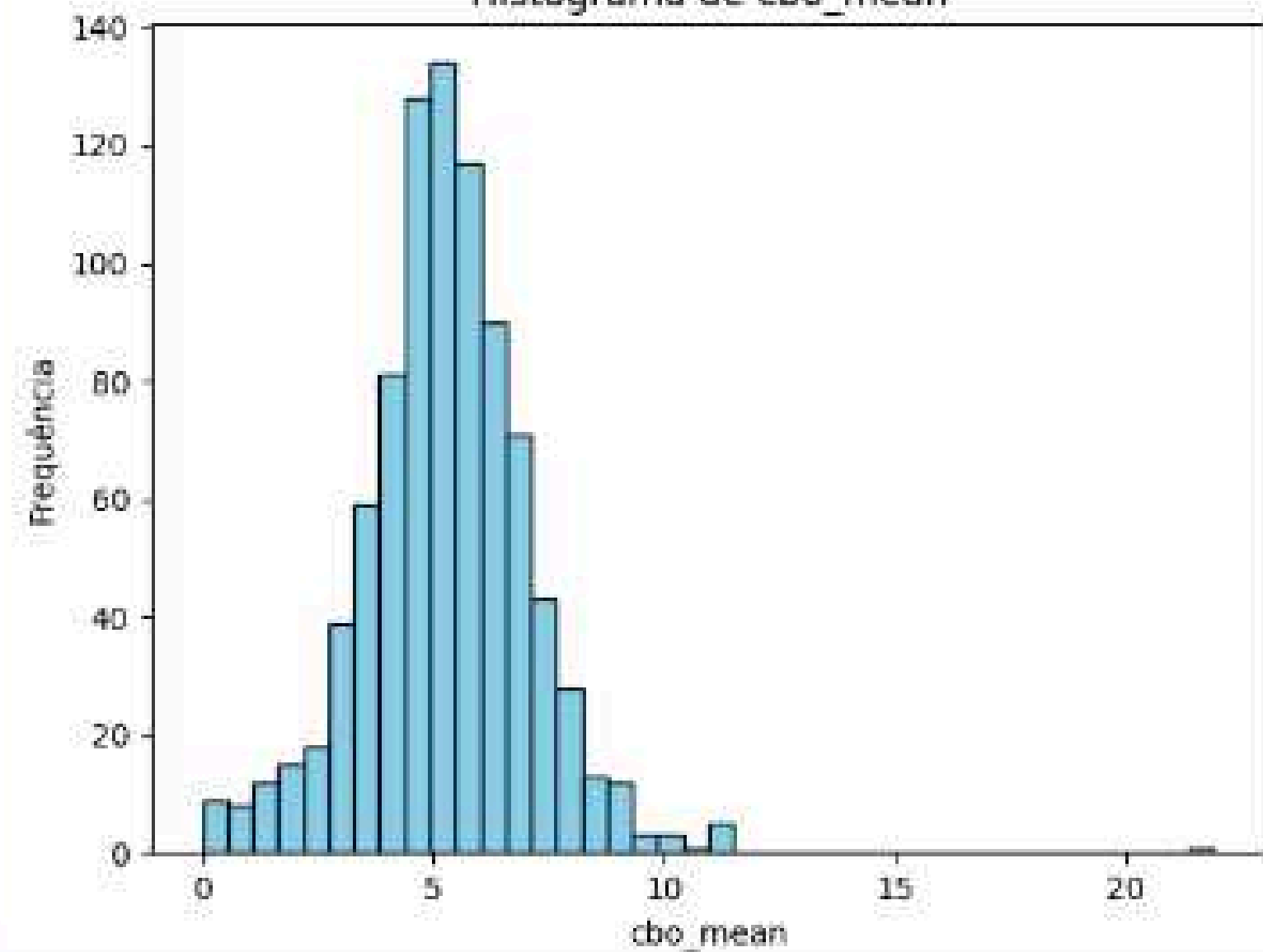
- **Releases:** média 37.10; mediana 10; dp 85.01
- **Estrelas:** média 9060.99; mediana 5614.5; dp 10363.73
- **Idade (anos):** média 9.05; mediana 9; dp 2.99
- **LOC:** média 43380.95; mediana 12473; dp 76875.07
- **CBO_mean:** 5.27; DIT_mean: 1.45; LCOM_mean: 33.45

métrica	média	mediana	desvio-padrão	moda
releases	37.104494	10.000000	85.012613	0.0
número de estrelas	9060.993258	5614.500000	10363.733200	3504.0
anos	9.047191	9.000000	2.987416	9.0
LOC	43380.949438	12473.000000	76875.067554	5.0
CBO média	5.266339	5.233520	1.835075	0.0
DIT média	1.449279	1.380952	0.356645	1.0
LCOM média	33.453302	21.413314	36.626377	0.0

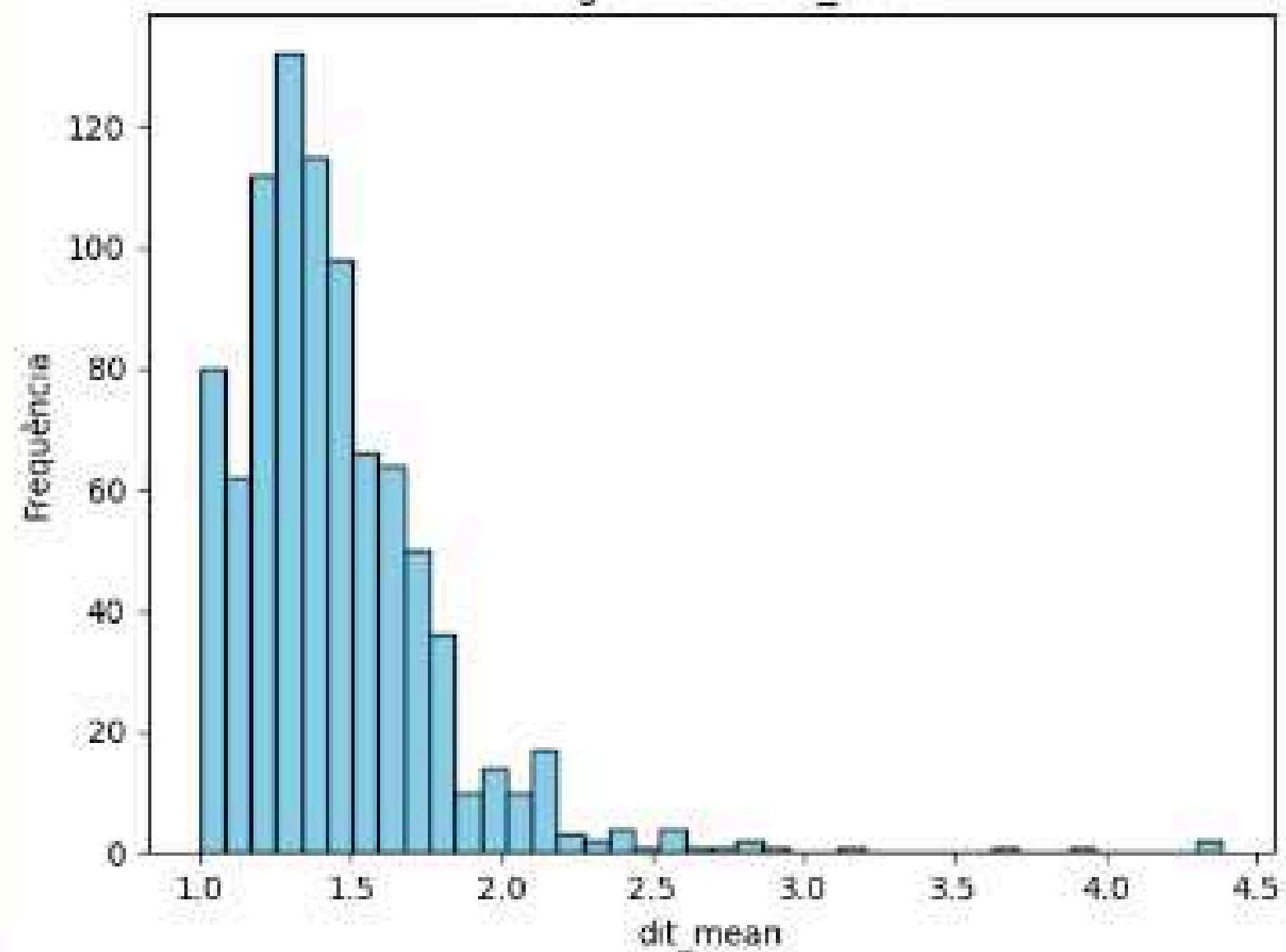
Histograma de loc



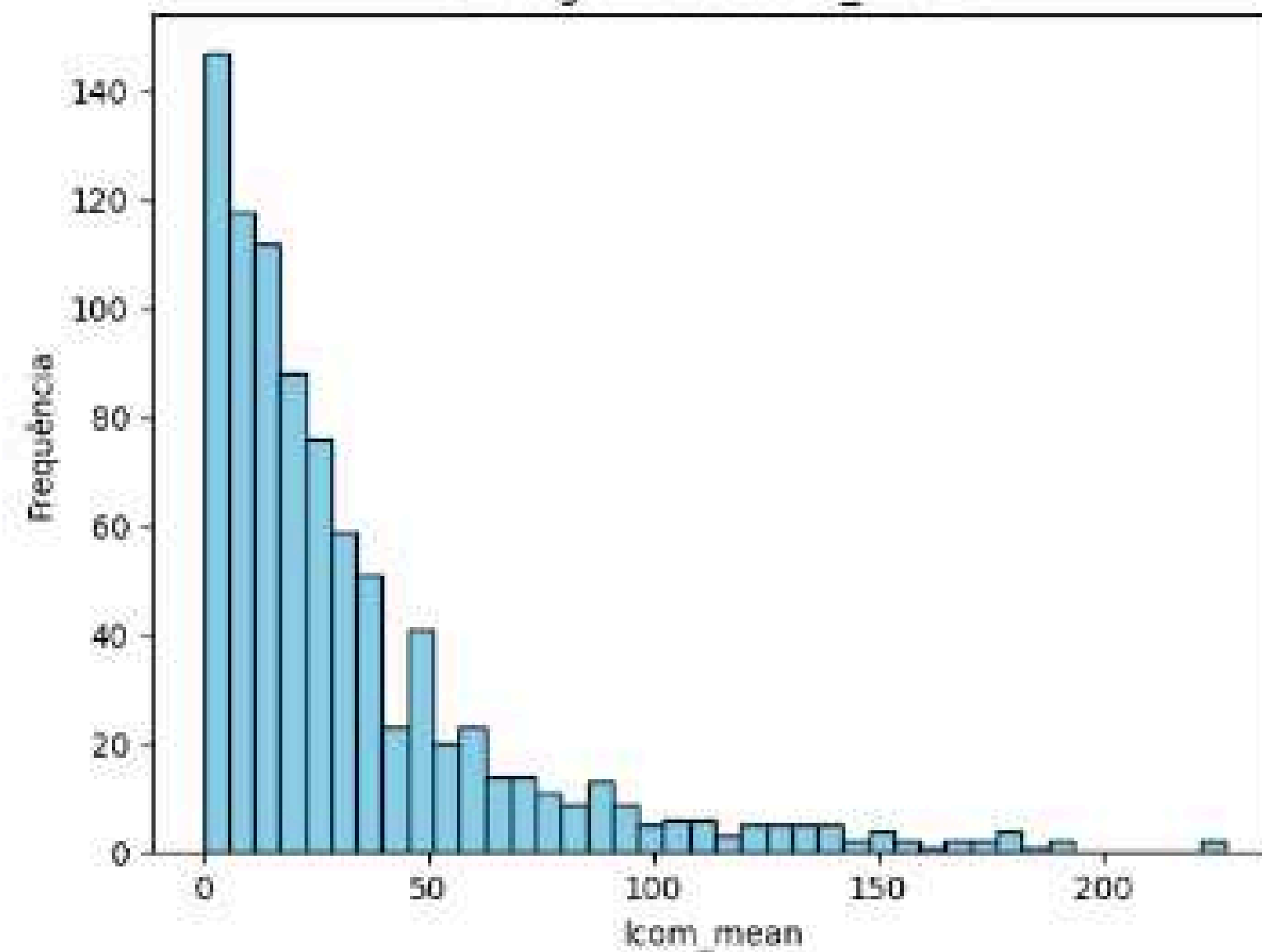
Histograma de cbo_mean

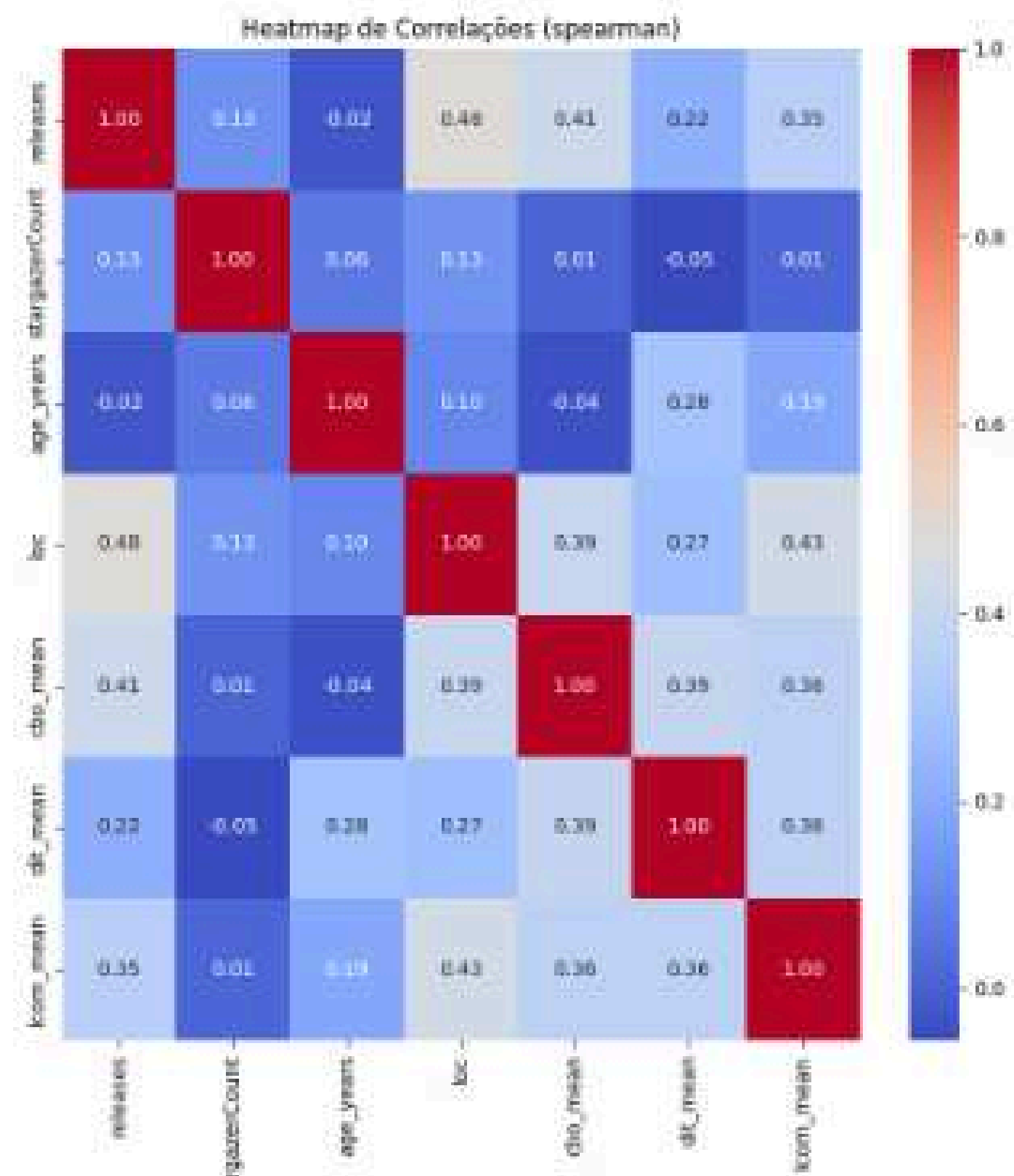
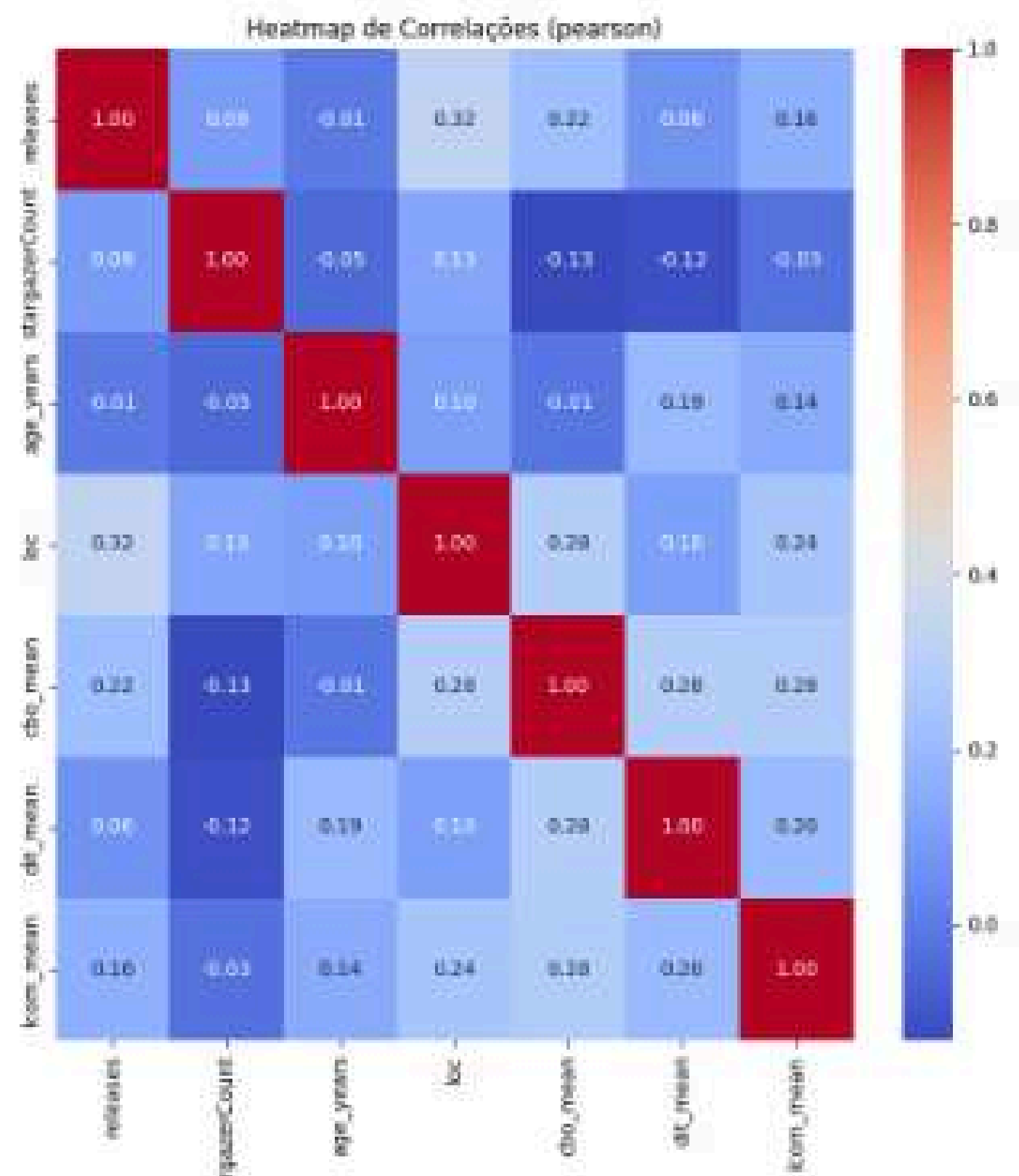


Histograma de dit_mean



Histograma de lcom_mean

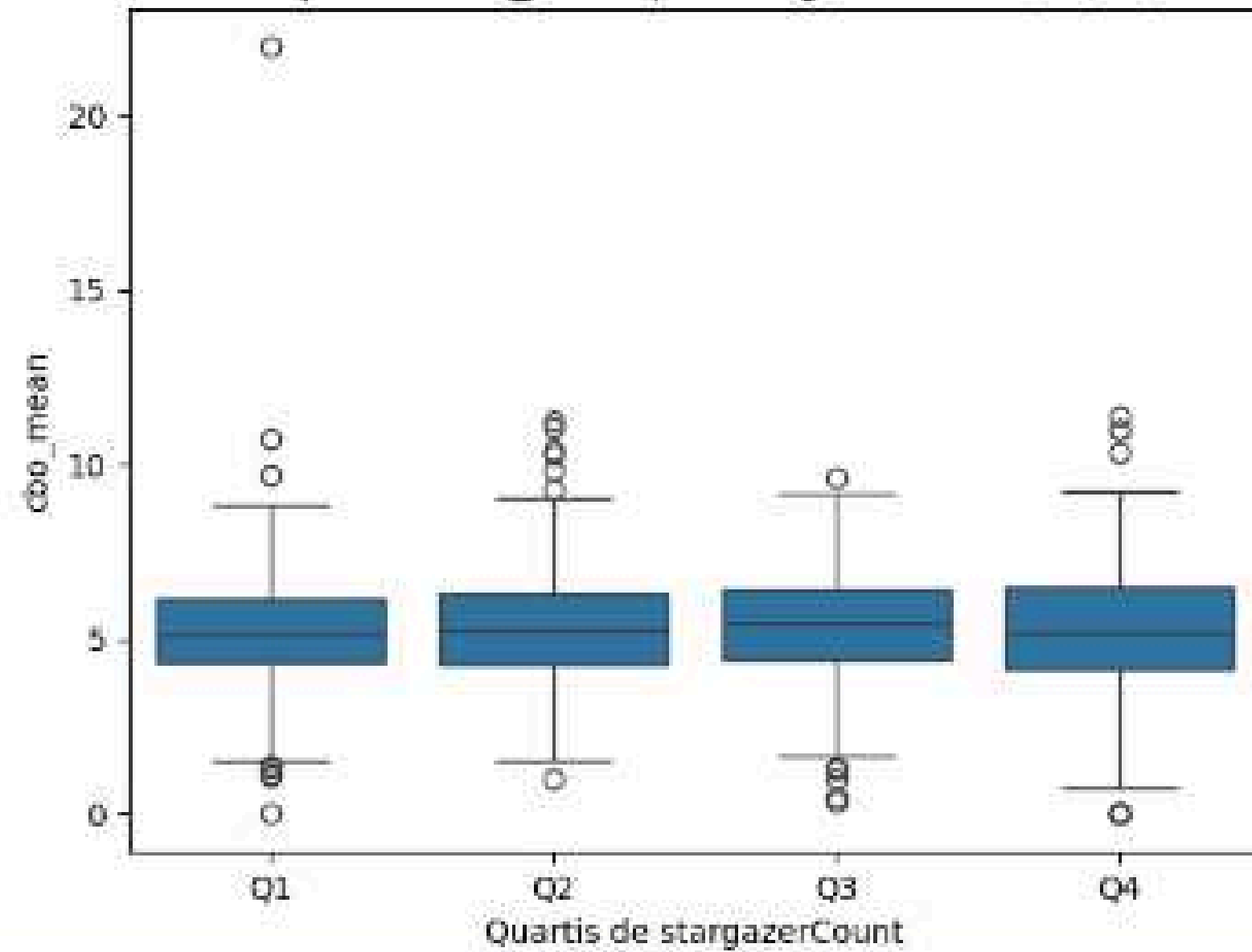




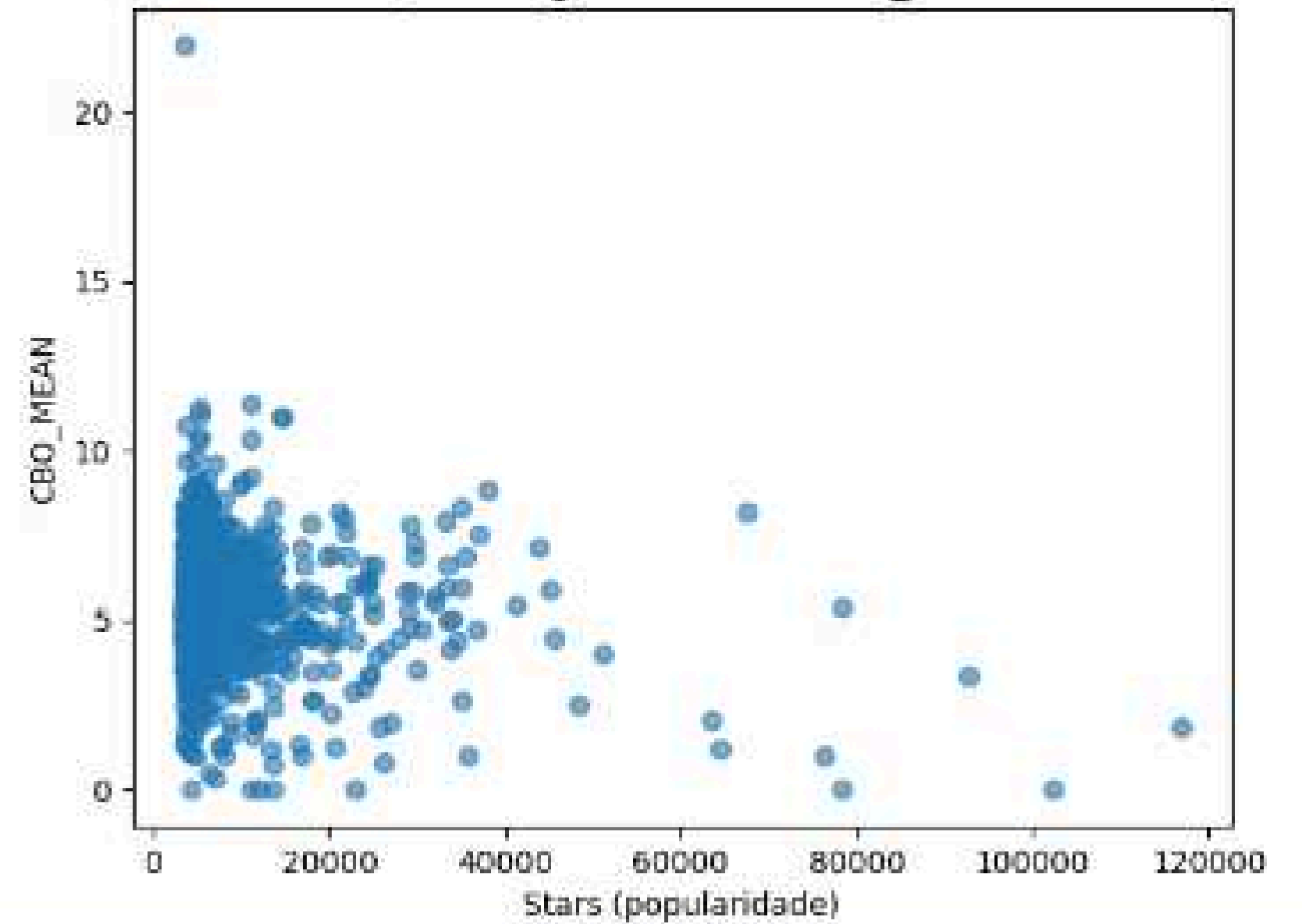
RQ01: Popularidade × Qualidade

- **CBO × estrelas:** -0.13 (Pearson), +0.01 (Spearman)
- **DIT × estrelas:** -0.11, -0.05
- **LCOM × estrelas:** -0.02, +0.01
- **Conclusão:** popularidade ~ não se relaciona à qualidade estrutural.

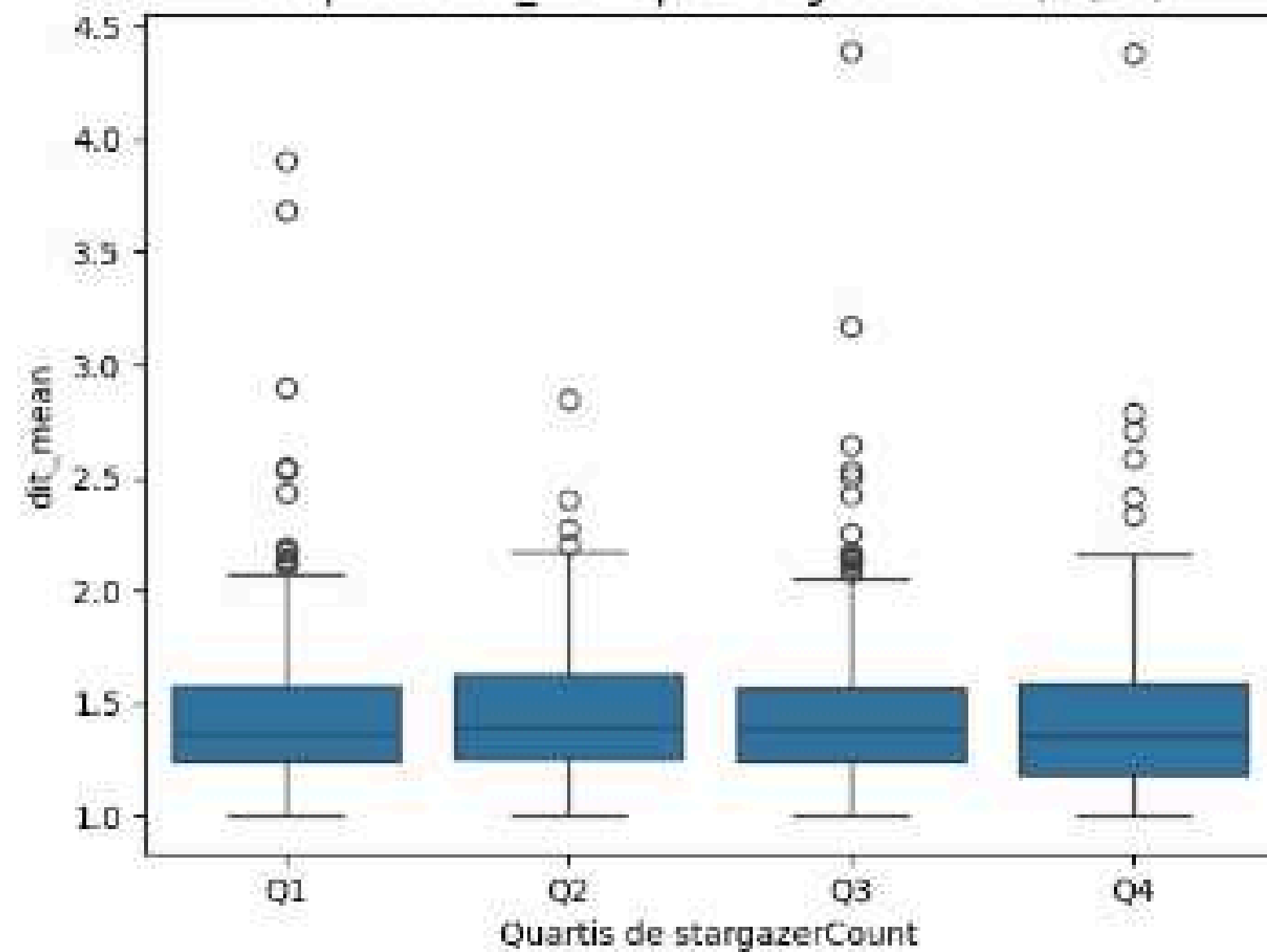
Boxplot de cbo_mean por stargazerCount (RQ01)



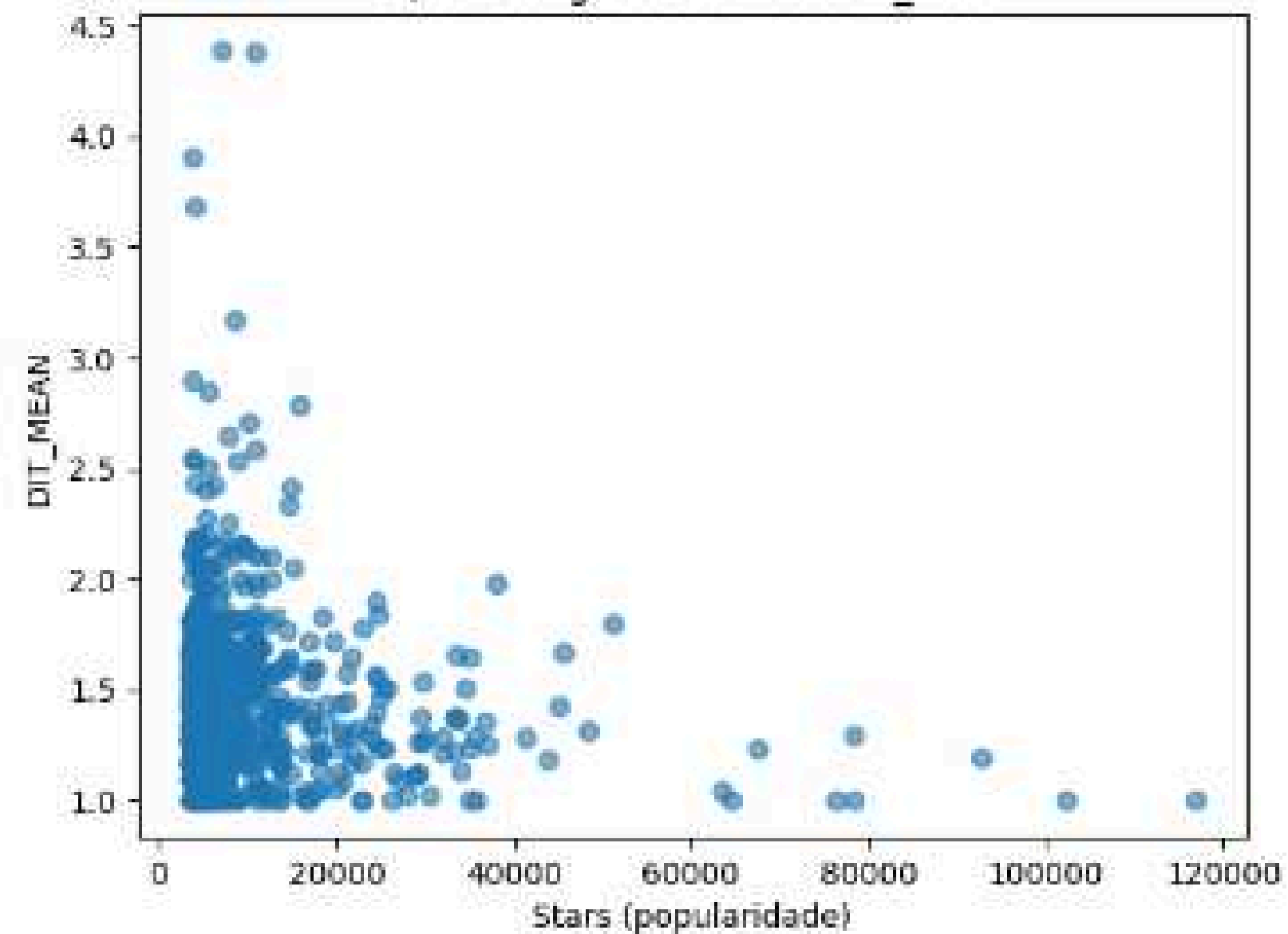
RQ01: stargazerCount vs cbo_mean



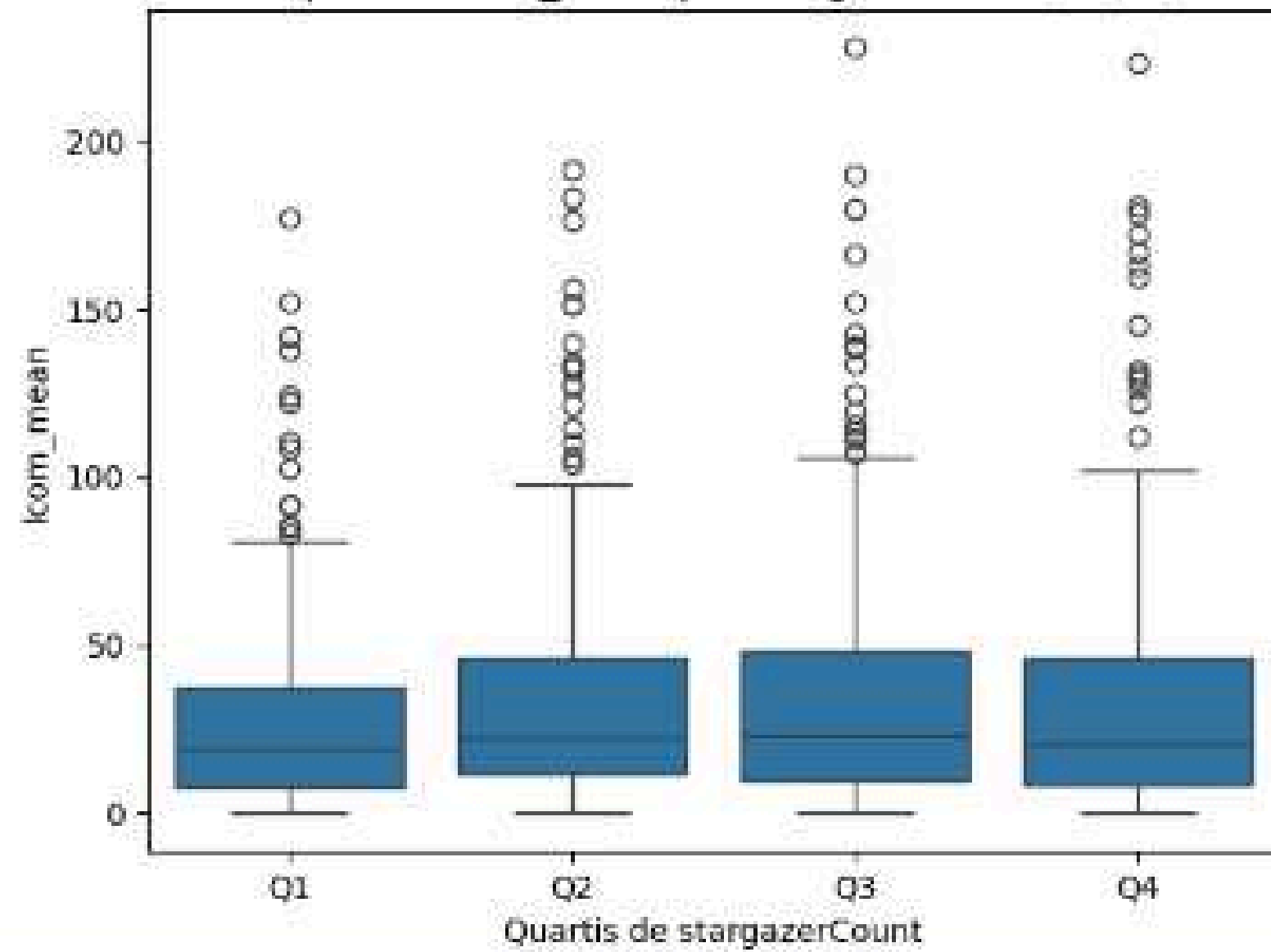
Boxplot de dit_mean por stargazerCount (RQ01)



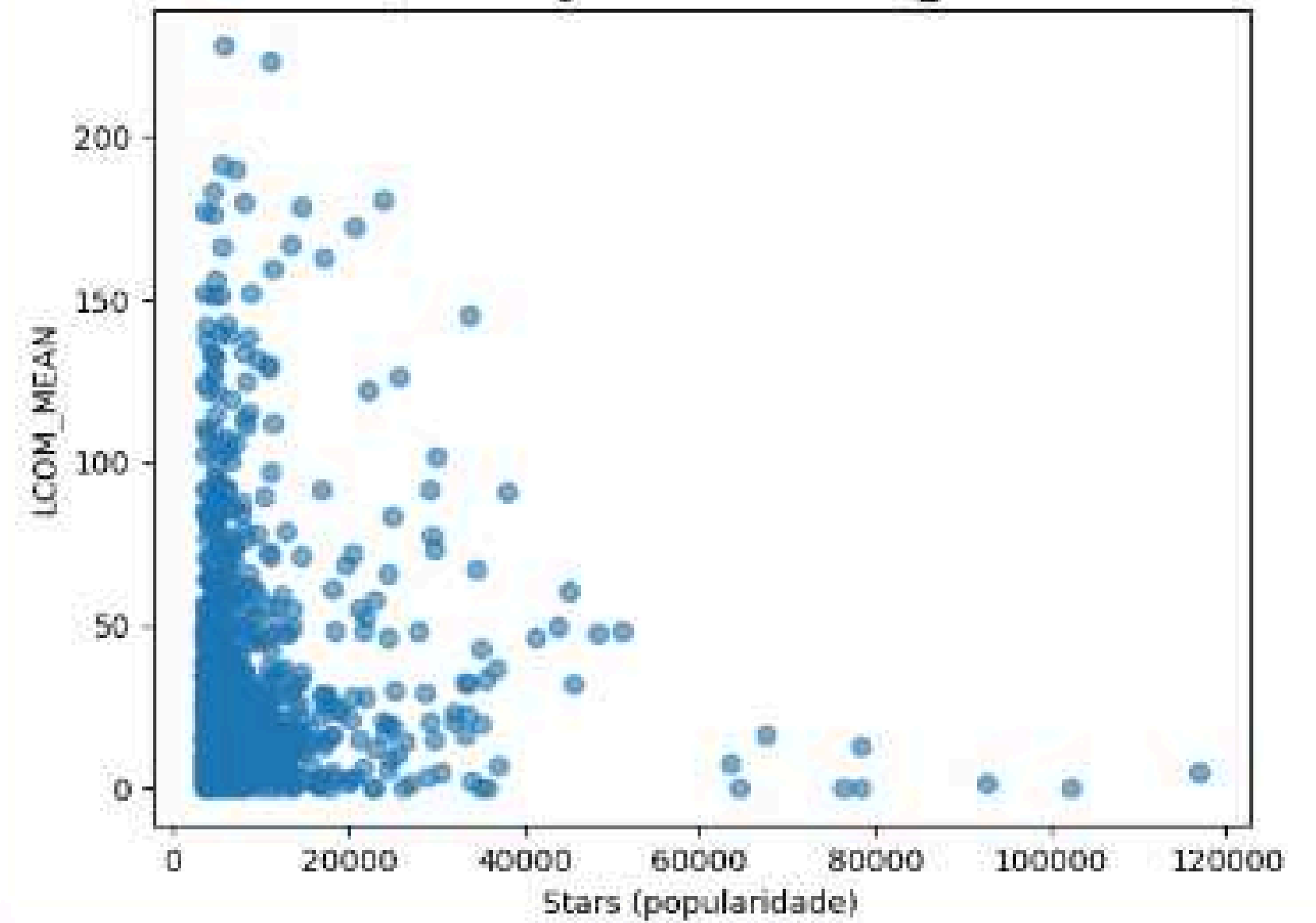
RQ01: stargazerCount vs dit_mean



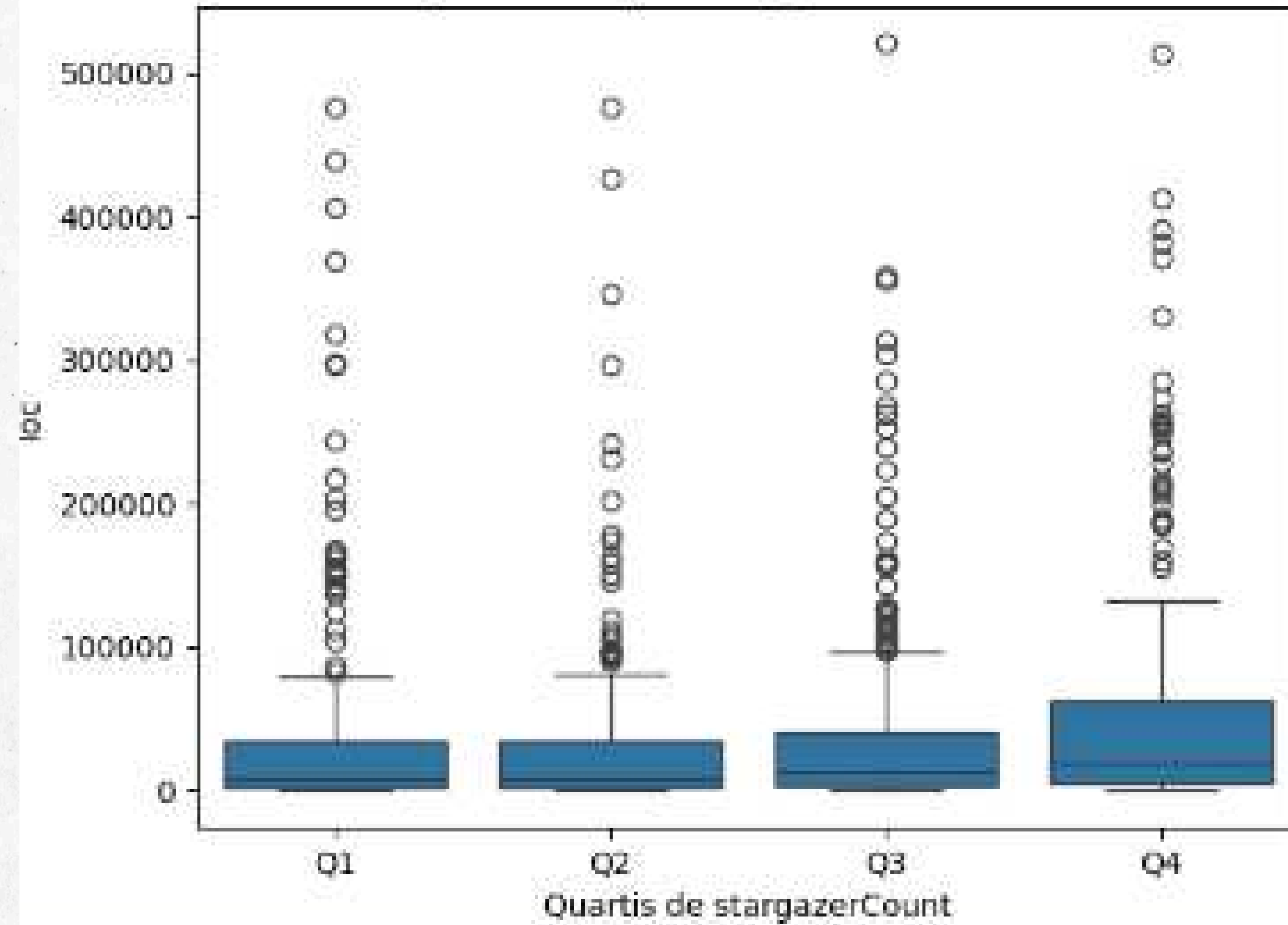
Boxplot de lcom_mean por stargazerCount (RQ01)



RQ01: stargazerCount vs lcom_mean



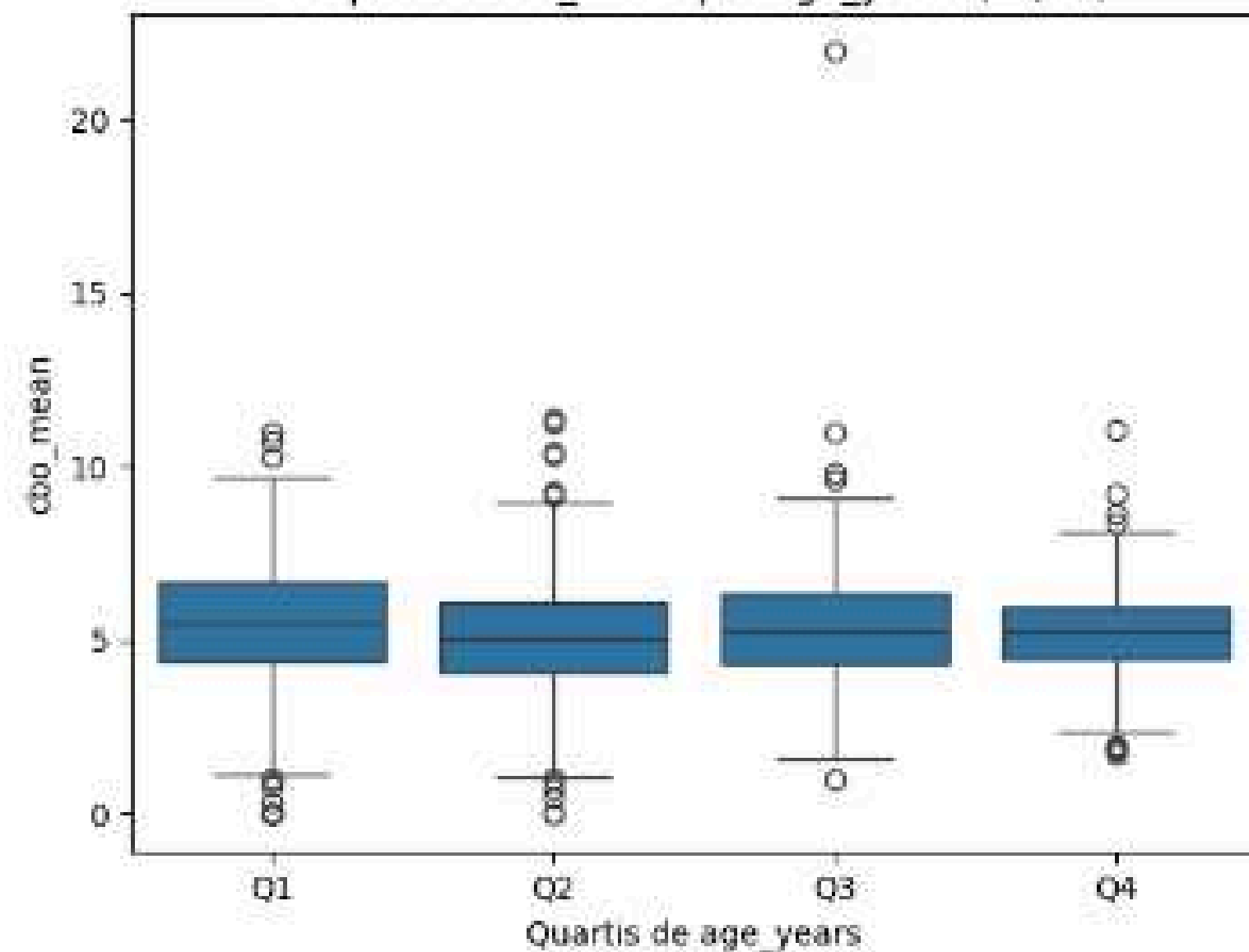
Boxplot de loc por stargazerCount (RQ01)



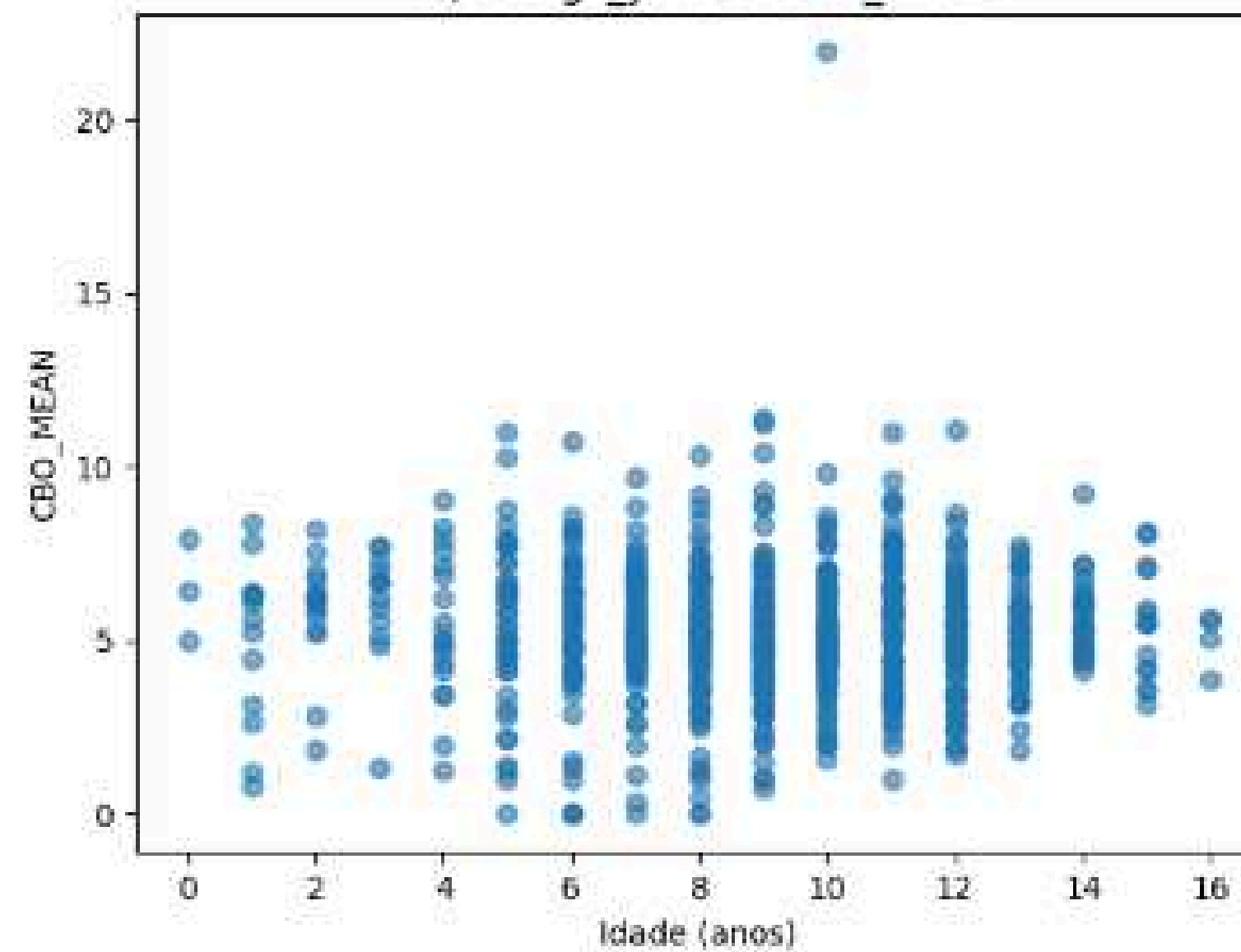
RQ02: Maturidade (Idade) × Qualidade

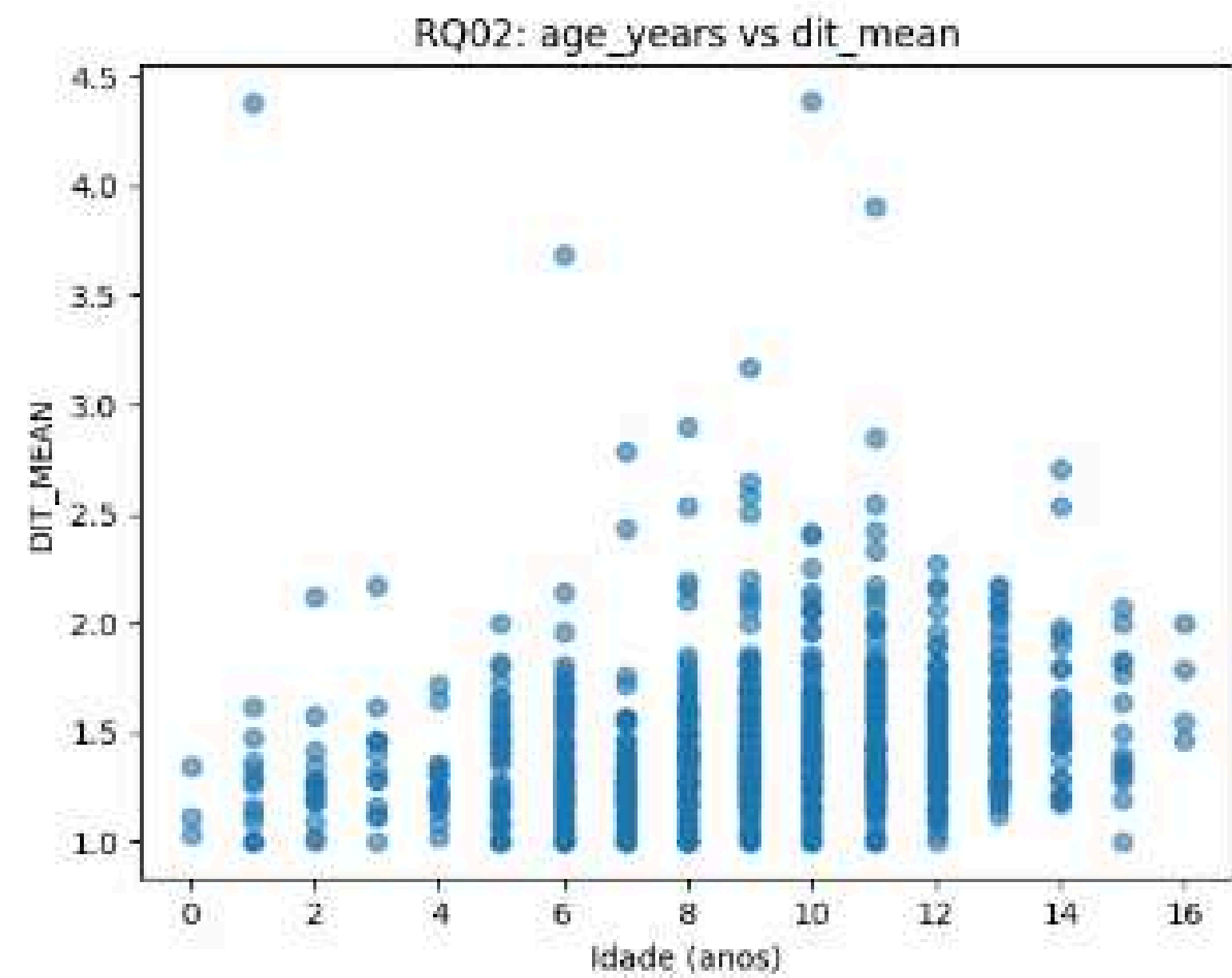
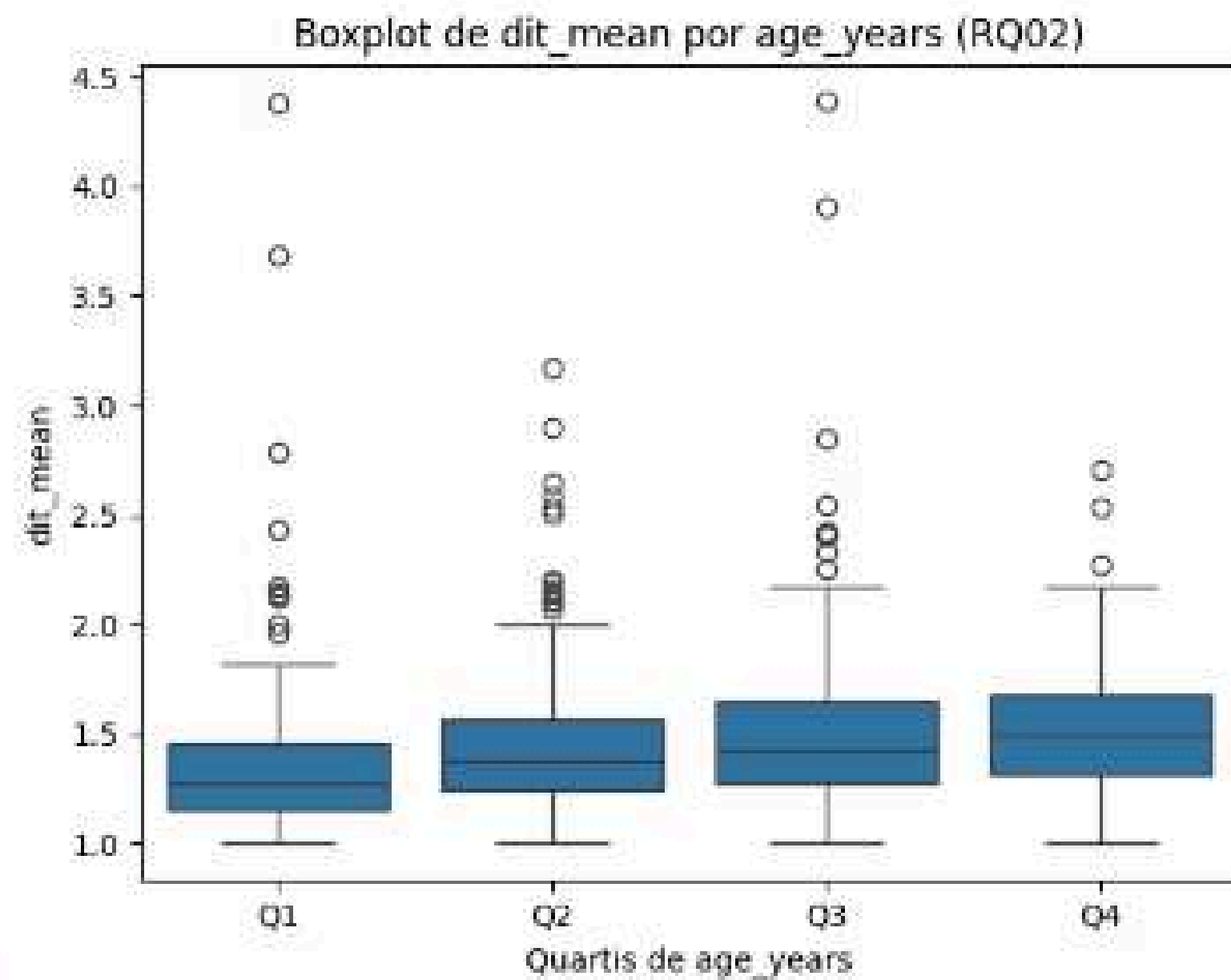
- **CBO × idade:** -0.01, -0.04 (quase nulo)
- **DIT × idade:** +0.19, +0.28 (tendência positiva/moderada)
- **LCOM × idade:** +0.14, +0.19 (positiva fraca-moderada)
- **Conclusão:** repositórios mais antigos tendem a aumentar complexidade (DIT/LCOM).

Boxplot de cbo_mean por age_years (RQ02)

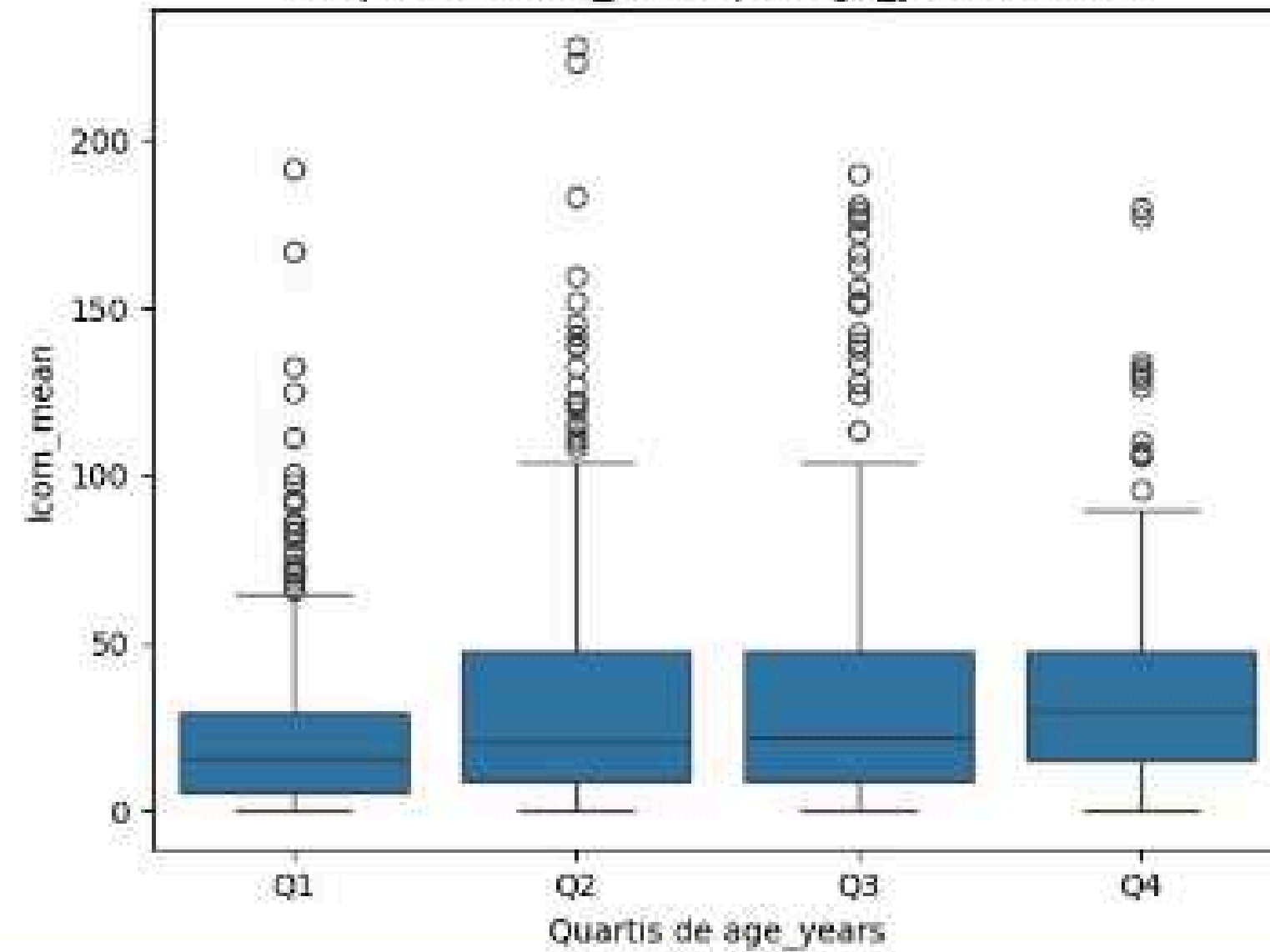


RQ02: age_years vs cbo_mean

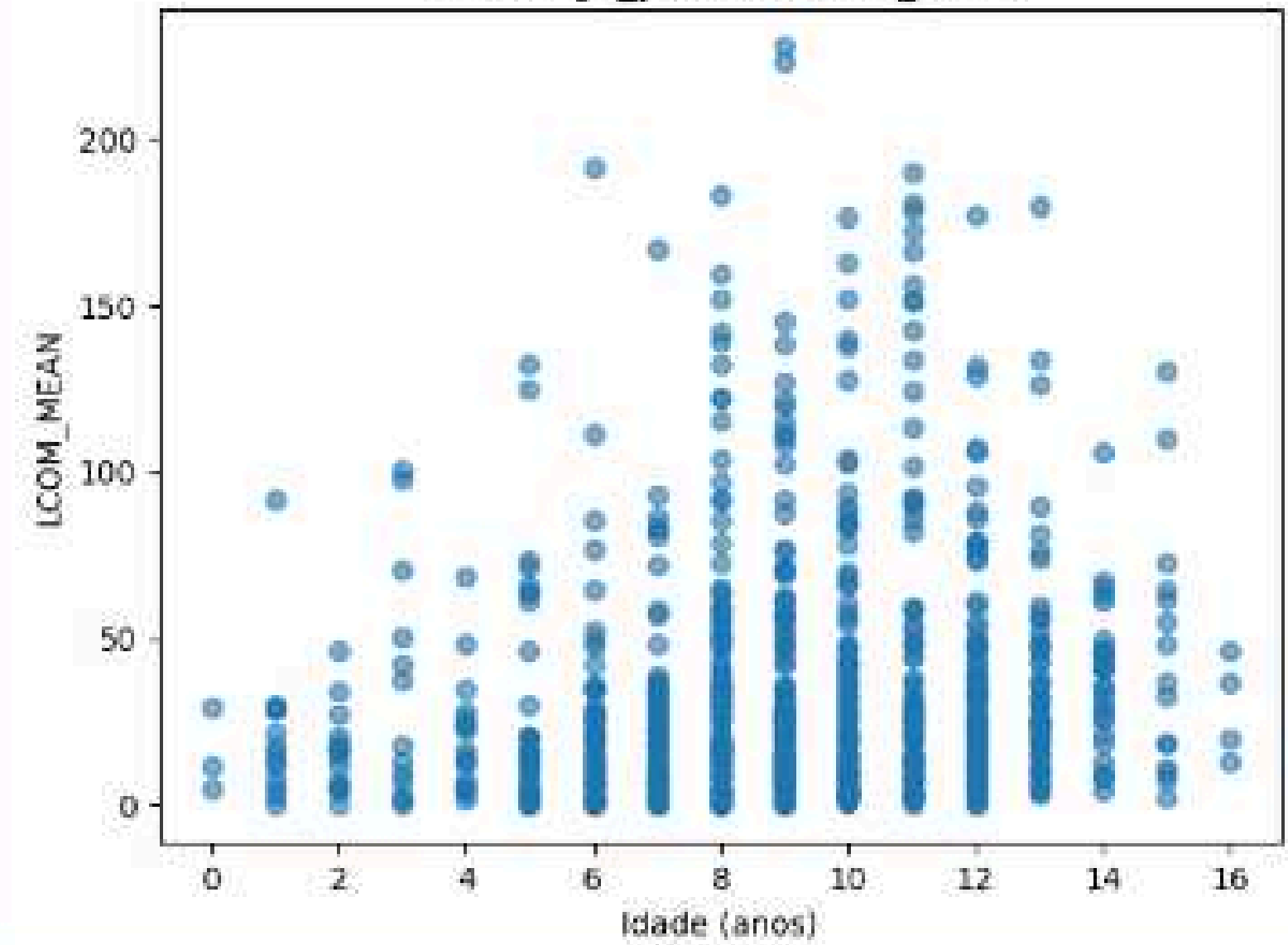




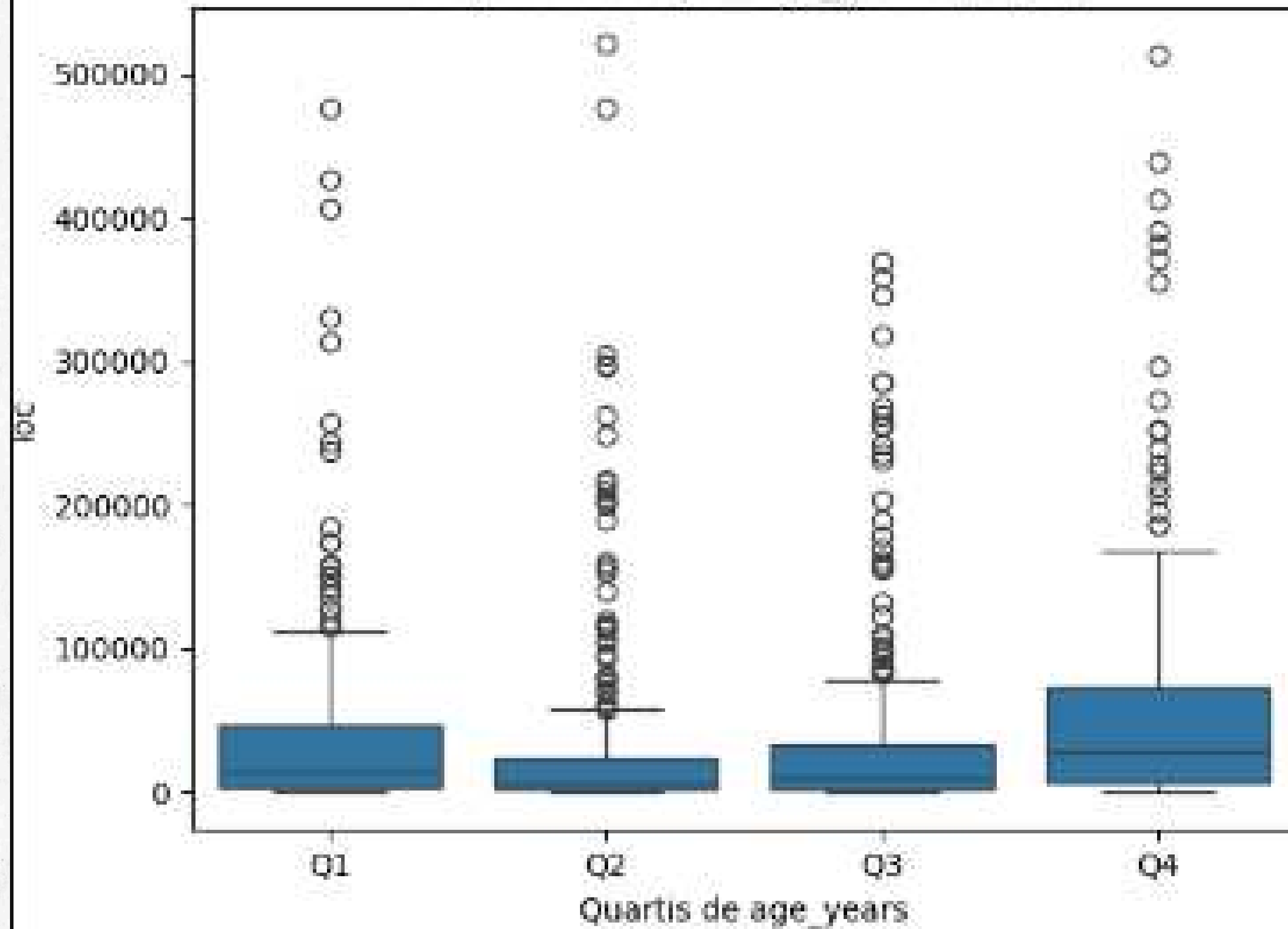
Boxplot de lcom_mean por age_years (RQ02)



RQ02: age_years vs lcom_mean



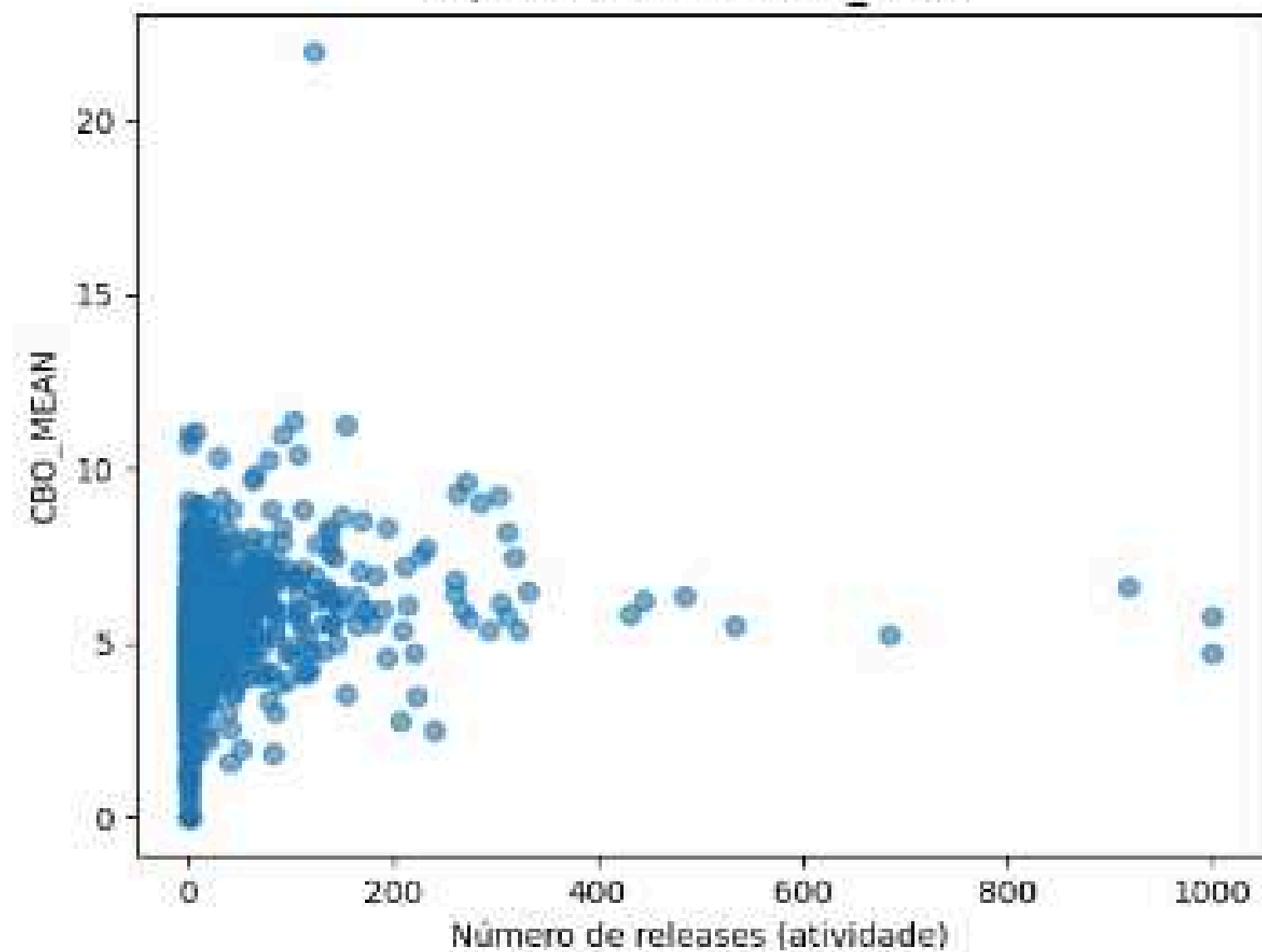
Boxplot de loc por age_years (RQ02)



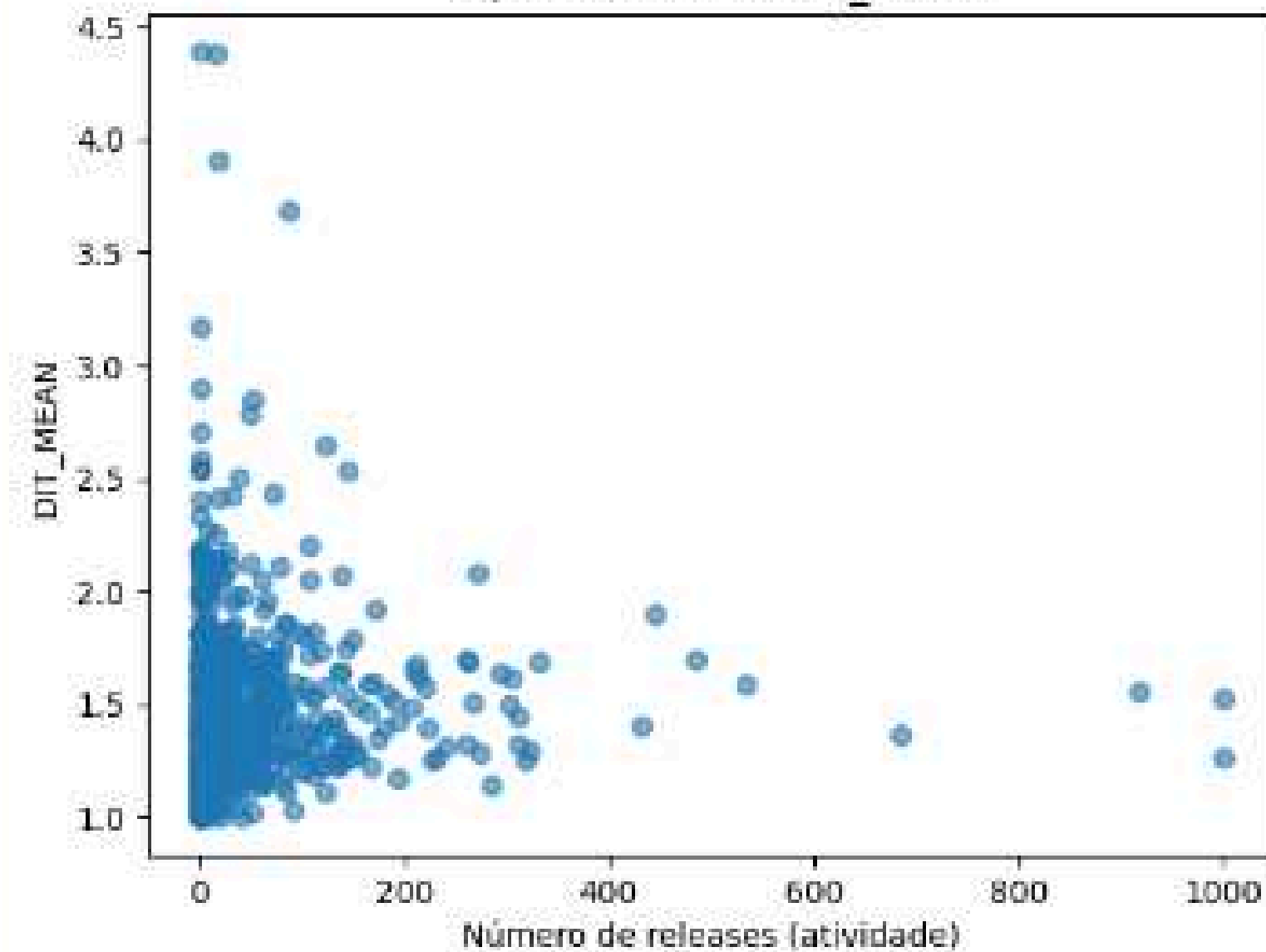
RQ03: ReRQ03: Atividade (Releases) × Qualidadeleases

- **CBO × releases:** +0.21, +0.41 (positiva; Spearman forte)
- **DIT × releases:** +0.06, +0.21
- **LCOM × releases:** +0.16, +0.35
- **Conclusão:** mais releases → maior acoplamento e menor coesão.

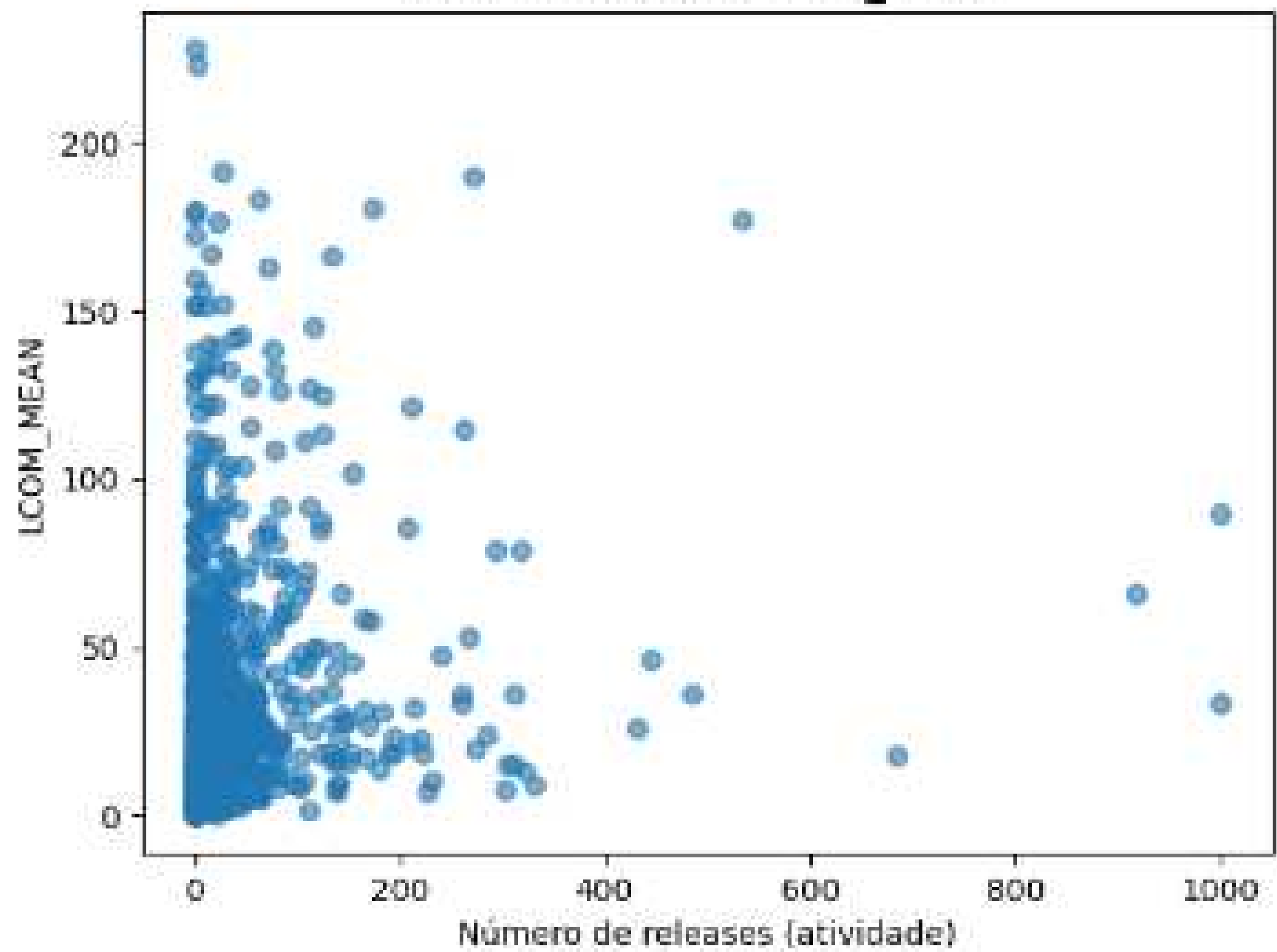
RQ03: releases vs cbo_mean



RQ03: releases vs dit_mean



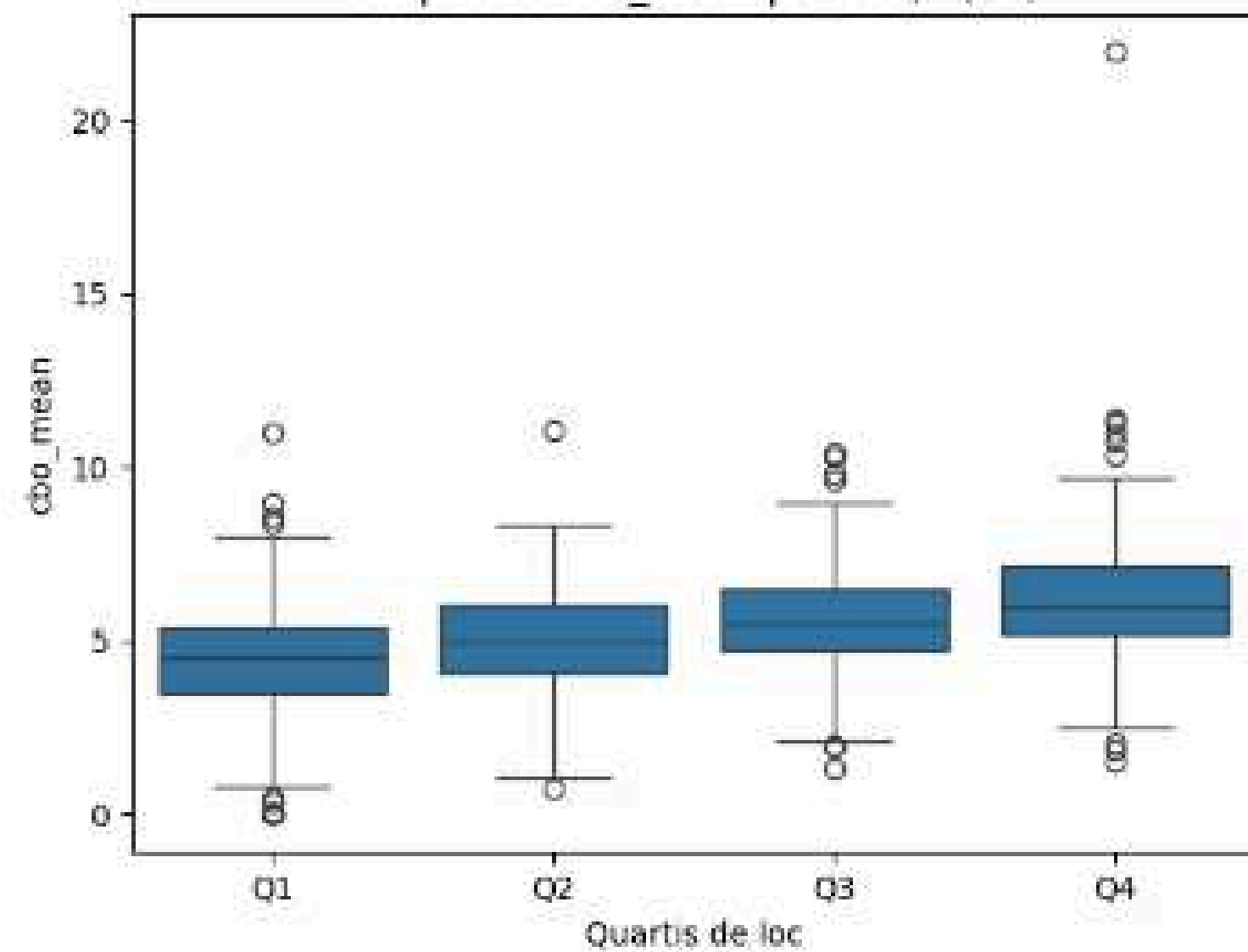
RQ03: releases vs lcom_mean



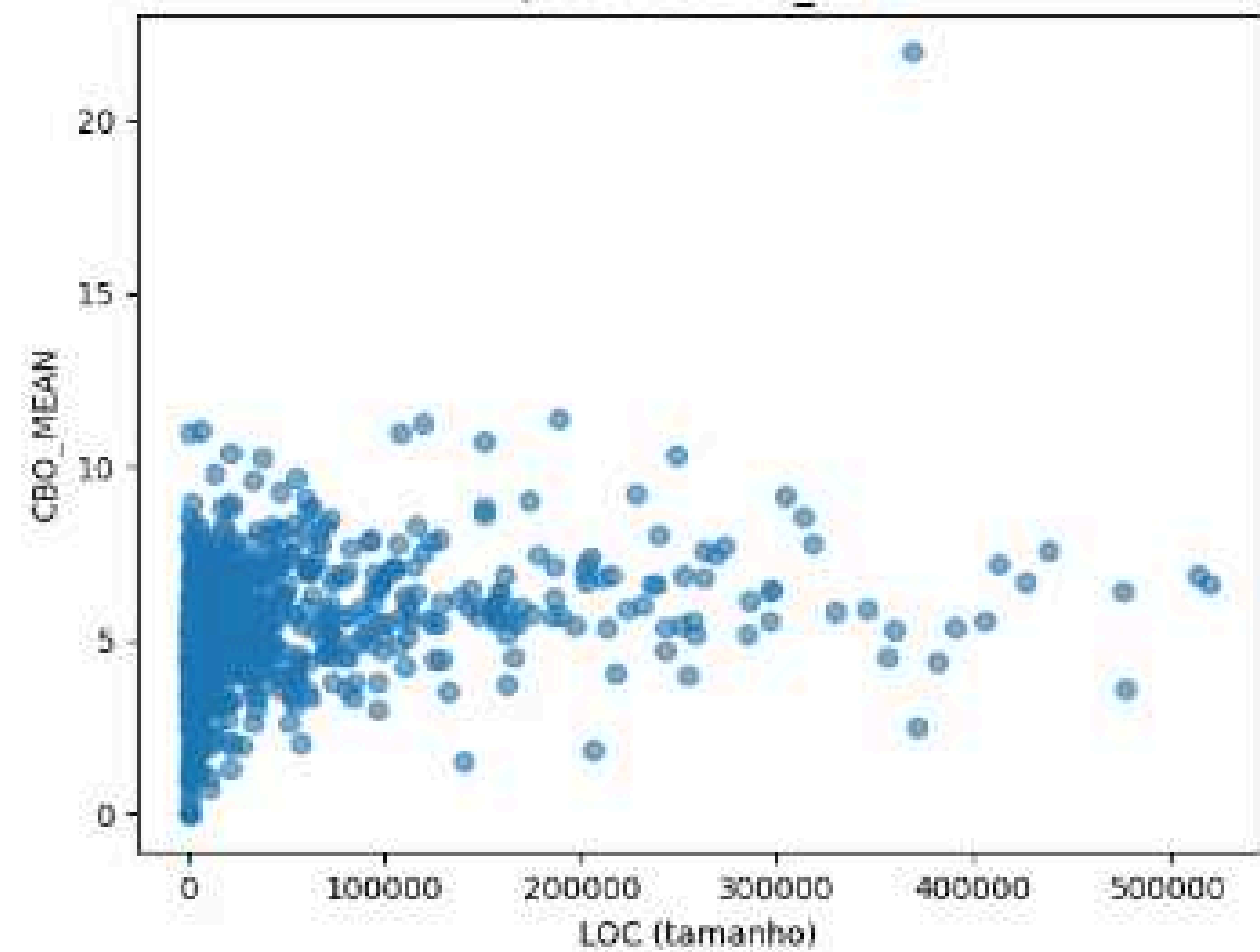
RQ04: Tamanho (LOC) × Qualidade

- **CBO × LOC:** +0.28, +0.39
- **DIT × LOC:** +0.10, +0.27
- **LCOM × LOC:** +0.24, +0.43
- **Conclusão:** sistemas maiores → pior qualidade estrutural (CK).

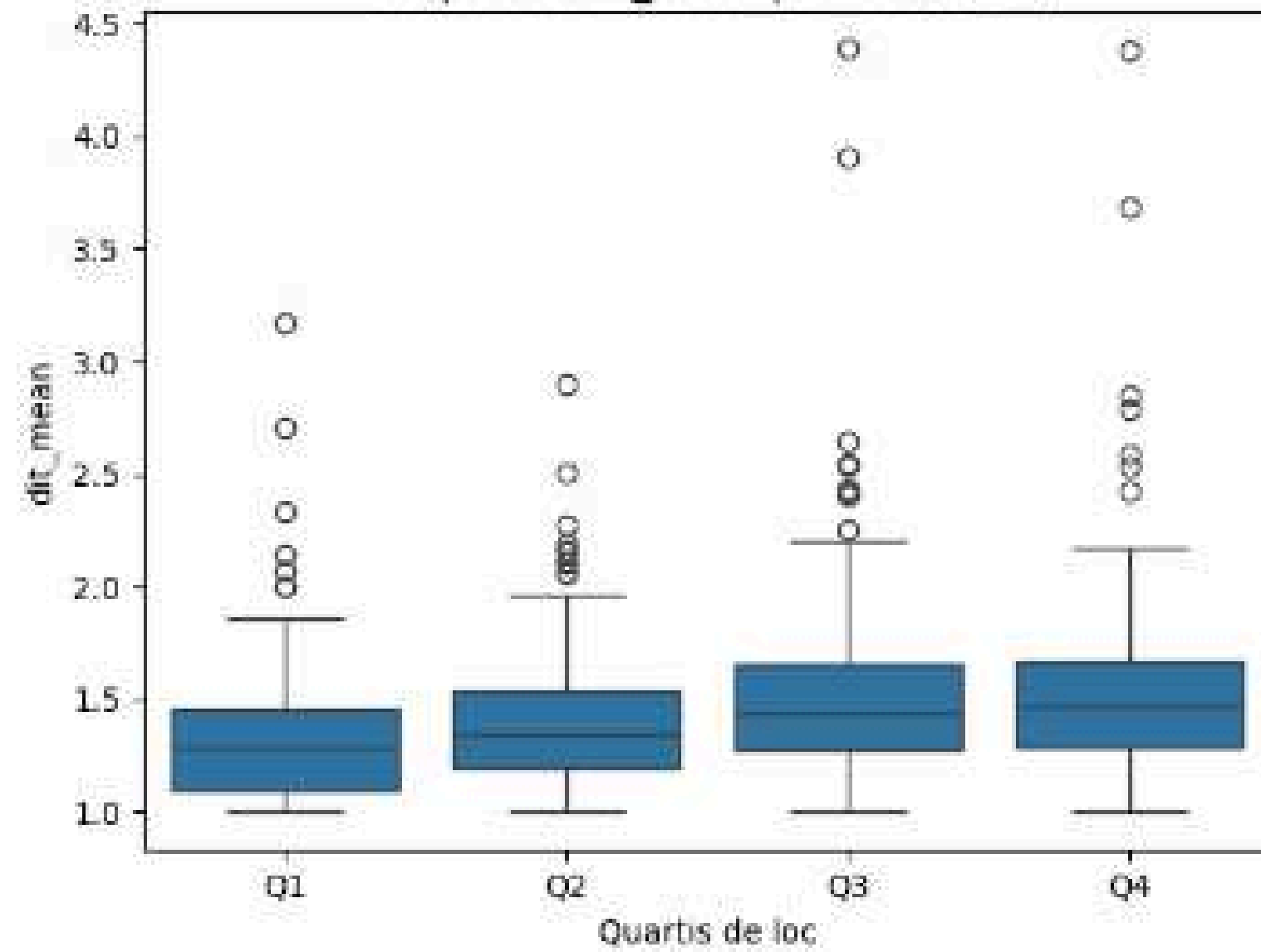
Boxplot de cbo_mean por loc (RQ04)



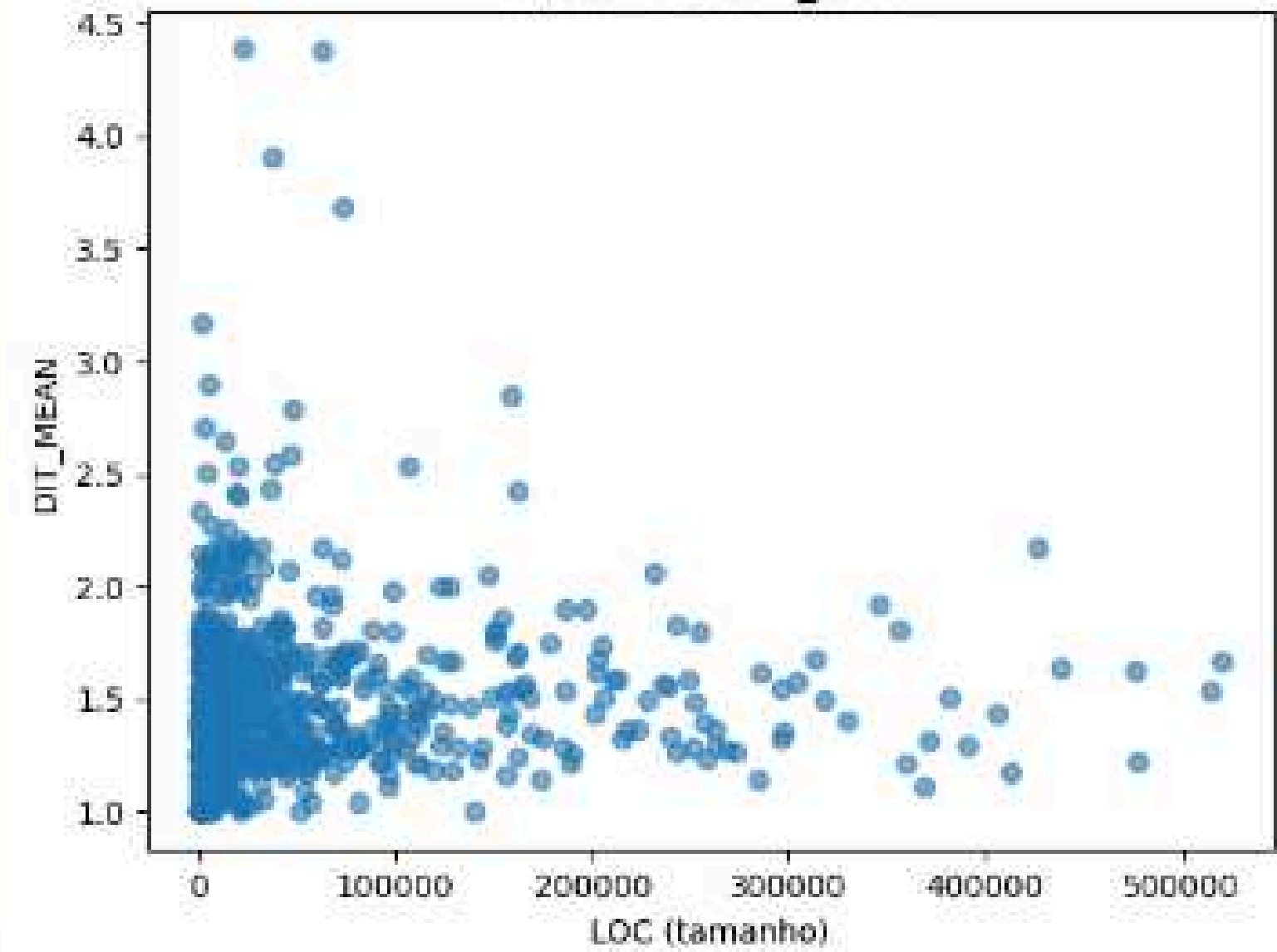
RQ04: loc vs cbo_mean



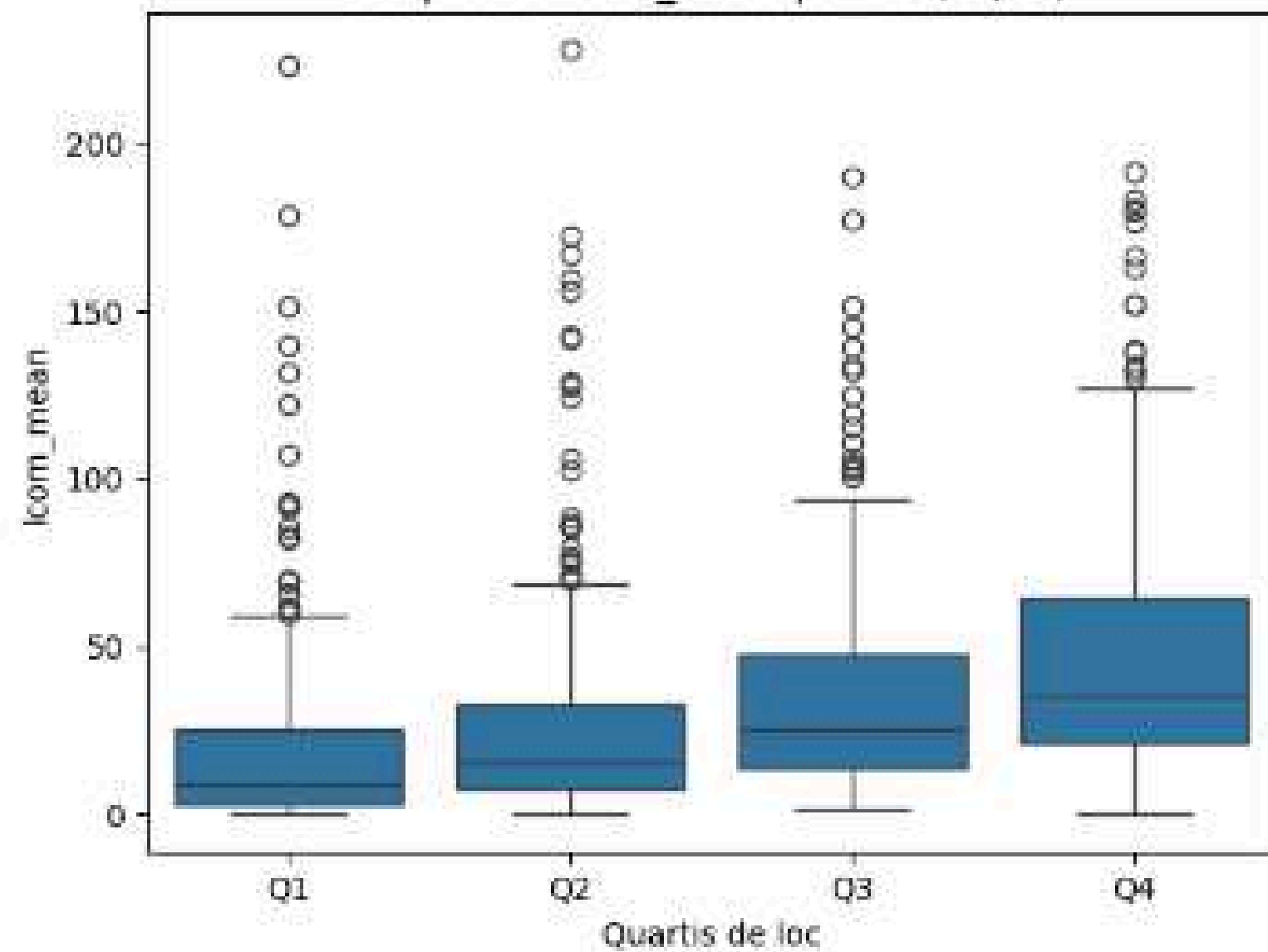
Boxplot de dit_mean por loc (RQ04)



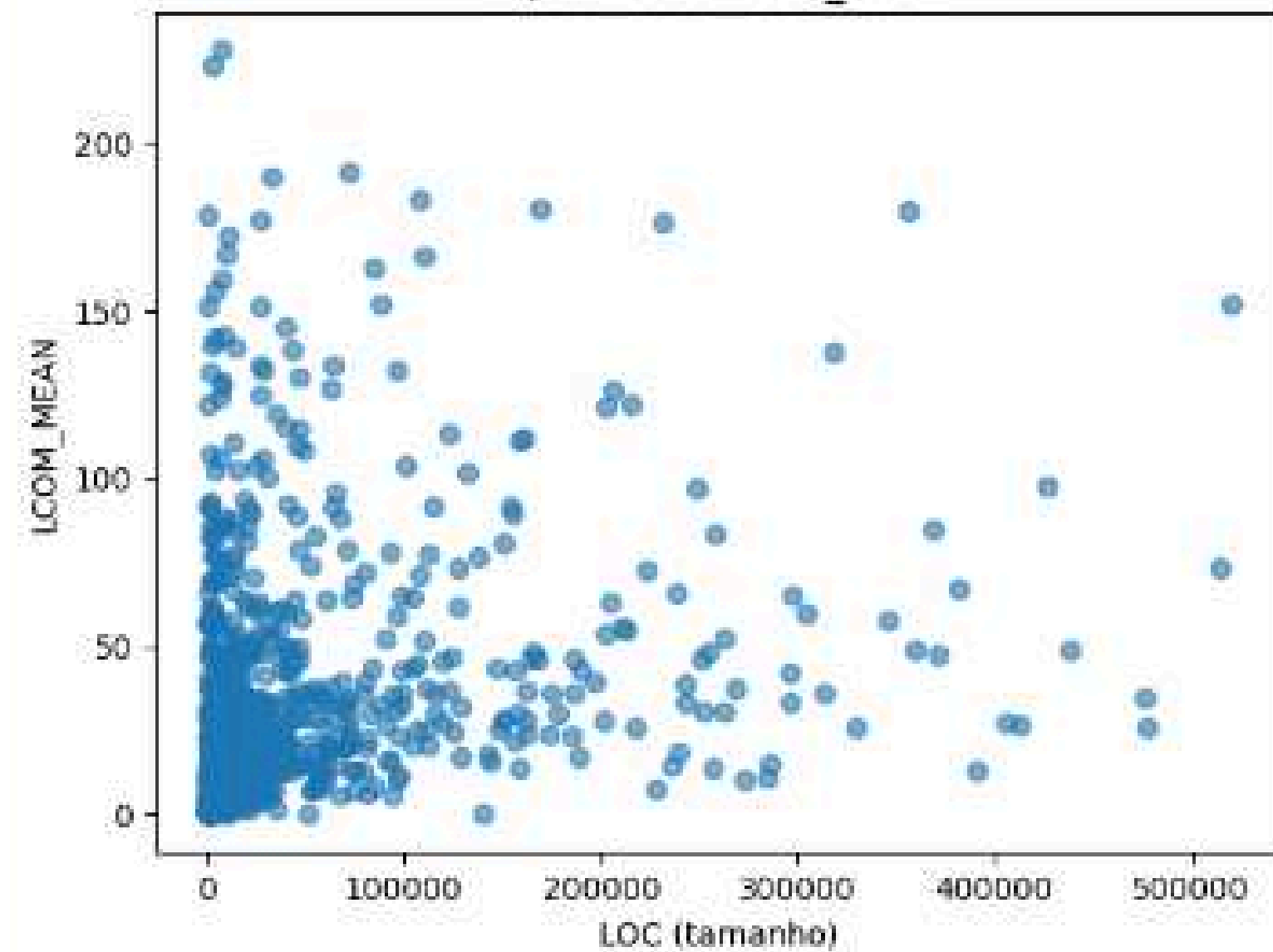
RQ04: loc vs dit_mean



Boxplot de lcom_mean por loc (RQ04)



RQ04: loc vs lcom_mean



Confronto com Hipóteses (Resumo)

- **RQ01 (H01-H03):** não confirmadas (popularidade não melhora CK).
- **RQ02 (H04-H06):** H04/H06 contrariadas; H05 não indica estabilidade (DIT \uparrow).
- **RQ03 (H07-H09):** H07/H09 contrariadas; H08 não estabiliza DIT.
- **RQ04 (H10-H12):** confirmadas (tamanho piora CK).

Implicações Práticas

- **Popularidade \neq qualidade interna:** não usar estrelas como proxy de qualidade.
- **Projetos antigos e muito ativos:** atenção a acoplamento/coesão.
- **Governança de crescimento:** modularização, boundaries claros, revisões arquiteturais.
- Refatoração contínua e limites de complexidade por módulo.

Limitações e Ameaças à Validade

- Métricas agregadas por média podem mascarar variação intra-repositório.
- Correlação não implica causalidade.
- Foco em projetos populares (viés de seleção).
- **Ferramenta CK e heurísticas:** possíveis erros de extração/medição.

Trabalhos Futuros

- Analisar distribuições intra-repositório (por módulo/pacote).
- Explorar modelos multivariados e normalizações por KLOC.
- Incluir outras métricas (complexidade ciclomática, RFC, WMC).
- Estudos longitudinais (séries temporais) por releases.

Conclusões

- Crescimento e manutenção tendem a degradar estrutura (CBO/DIT/LCOM).
- Popularidade não se associa à qualidade estrutural.
- Idade, releases e LOC correlacionam-se com maior complexidade e menor coesão.
- Recomenda-se monitoramento contínuo de CK em projetos grandes



OBRIGADO!