



**Laboratório #01**  
**Características de Sistemas Populares GitHub**

**Instruções:**

1. A coleta de dados pode ser feita através da linguagem de programação da sua escolha;
2. O código desenvolvido faz parte da entrega;
3. Código plagiado implicará na anulação da submissão.

**Objetivo:**

Neste laboratório, vamos estudar as principais características de sistemas populares open-source. Dessa forma, vamos analisar como eles são desenvolvidos, com que frequência recebem contribuição externa, com qual frequência lançam releases, entre outras características. Para tanto, colete os dados indicados a seguir para os **1.000 repositórios com maior número de estrelas no GitHub** e discuta os valores obtidos.

**Questões de Pesquisa:**

**RQ 01.** Sistemas populares são maduros/antigos?

Métrica: idade do repositório (calculado a partir da data de sua criação)

**RQ 02.** Sistemas populares recebem muita contribuição externa?

Métrica: total de pull requests aceitas

**RQ 03.** Sistemas populares lançam releases com frequência?

Métrica: total de releases

**RQ 04.** Sistemas populares são atualizados com frequência?

Métrica: tempo até a última atualização (calculado a partir da data de última atualização)

**RQ 05.** Sistemas populares são escritos nas linguagens mais populares do mercado?<sup>1</sup>

Métrica: linguagem primária de cada um desses repositórios

---

<sup>1</sup> Compare os resultados com os dados dessa reportagem:

<https://www.devmedia.com.br/top-10-linguagens-de-programacao-mais-usadas-no-mercado/39635>

**RQ 06.** Sistemas populares possuem um alto percentual de issues fechadas?

Métrica: razão entre número de issues fechadas pelo total de issues

### **Relatório Final:**

Para cada uma das questões de pesquisa anteriores, faça uma sumarização dos dados obtidos através de valores medianos<sup>2</sup>. Mesmo que de forma informal, elabore hipóteses sobre o que você espera de resposta e tente analisar a partir dos valores obtidos. Para valores de categoria (ex.: linguagem de programação), elabore uma contagem por categoria, para facilitar suas descobertas.

Elabore um documento que apresente (i) uma introdução simples com hipóteses informais; (ii) a metodologia que você utilizou para responder às questões de pesquisa; (iii) os resultados obtidos para cada uma delas; (iii) a discussão sobre o que você esperava como resultado (suas hipóteses) e os valores obtidos.

### **Bônus (+ 2 pontos):**

Divida os resultados obtidos nas RQs 02-04 por linguagem e analise como esses valores se comportam de acordo com as linguagens de cada repositório. Ou seja, acrescente ao seu trabalho a seguinte questão:

**RQ. 07:** Sistemas escritos em linguagens mais populares recebem mais contribuição externa, lançam mais releases e são atualizados com mais frequência?

Dica: compare os resultados para os sistemas com as linguagens da reportagem com os resultados de sistemas em outras linguagens.

### **Entrega:**

Submeta no SGA, até o dia 08/09, uma pasta compactada (.zip) com os seguintes artefatos:

1. Os dados coletados num arquivo .csv;
2. Os scripts de coleta de dados;
3. O relatório final em formato .pdf.

Valor total: 12.5 pontos

---

<sup>2</sup> <https://www.alura.com.br/artigos/media-ou-mediana-entendendo-cada-uma>