# Repositórios GitHub©

Daniel Venturini Luiz Augusto da Silva Silveira

#### Conjunto dos Dados

O conjunto de dados foi obtido através da API do GitHub, dos quais foram selecionados mil repositórios com maior quantidade de Estrelas e maior quantidade de Forks.

Nós temos dois conjuntos de dados: um com os mil repositórios com mais estrelas; e um com os mil repositórios com mais forks.

#### Conjunto dos Dados

Links utilizado para obtenção da *database* do git.

Repositórios maior número estrela:

https://api.github.com/search/repositories?q= stars%3A%3E9000&per\_page=100&page=1

Repositórios maior número de forks:

https://api.github.com/search/repositories?q=forks%3A%3E2000&per\_page=100&page=1

```
1039
 total count:
                             false
 incomplete results:
▼items:
                             28457823
       id:
                             "MDEw01JlcG9zaXRvcnkv0D01NzgvMw=="
       node id:
                             "freeCodeCamp"
                             "freeCodeCamp/freeCodeCamp"
       full name:
                             false
       private:
                             {-}
     owner:
                             "https://github.com/freeCodeCamp/freeCodeCamp"
       html url:
                             "The https://freeCodeCamp...ode and help nonprofits."
     description:
                             false
       fork:
                             35897
       size:
                             294175
       stargazers_count:
                             294175
       watchers count:
                             "JavaScript"
       language:
       has issues:
                             true
                             true
       has_projects:
       has downloads:
                             true
       has wiki:
                             false
                             false
       has pages:
                             14556
       forks count:
                             nul1
       mirror url:
                             false
       archived:
                             355
       open issues count:
     1 license:
                             {_}}
                             14556
       forks:
```

#### Problema para recuperar os dados

Recuperamos os dados diretamente pelo navegador.

O GitHub não retornava as informações dos mil repositórios de uma vez. O máximo que conseguimos era apenas de cem repositórios.

Depois descobrimos que o GitHub retorna os dados em páginas, assim, conseguimos recuperar todos os dados, cem por página.

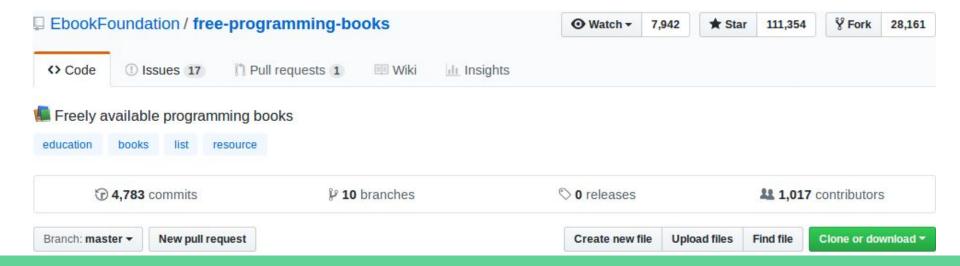
#### Obtenção dos dados

Para recuperar automaticamente os dados e transformá-los em CSV, foi desenvolvido uma pequena ferramenta em JavaScript que realizava as requisições, uma por página, e organiza todos em uma tabela para transformar em CSV.

### Problemas com alguns dados

Alguns campos que o GitHub retornava não continham valores. Exemplo disso é quando o GitHub não consegue identificar a linguagem do repositório, então este valor é retornado como "null"





#### Problemas com alguns dados

- 12.1% dos projetos da lista dos 1k com mais forks não possuem linguagem.
- 13.1% dos projetos da lista dos 1k com mais forks não possuem licença.

- 12.4% dos projetos da lista dos 1k com mais estrelas não possuem linguagem.
- 13.6% dos projetos da lista dos 1k com mais forks não possuem licença.

#### Base de Dados

#### Export to CSV

name	created_at	size	stargazers_c	forks_count	open_issues	license	linguagem	open_issues	fork
Spoon-Knife	2011-01-27T	1407736	10046	100237	12609	null	HTML	12609	false
tensorflow	2015-11-07T	257852	110400	67895	1794	Apache Licen	C++	1794	false
bootstrap	2011-07-29T	127848	127614	62277	479	MIT License	CSS	479	false
gitignore	2010-11-08T	1702	73033	34114	161	Creative Com		161	false
angular.js	2010-01-06T	99311	59132	29015	497	MIT License	JavaScript	497	false
free-program	2013-10-11T	4643	111589	28214	18	Other		18	false
models	2016-02-05T	572248	41864	25113	1338	Apache Licen	Python	1338	false
linux	2011-09-04T	2246249	63819	23223	231	Other	С	231	false
bitcoin	2010-12-19T	114301	34960	21150	841	MIT License	C++	841	false
spring-boot	2012-10-19T	80982	29265	20699	388	Apache Licen	Java	388	false

#### Possíveis visualizações

- Estabelecer qual relação entre projetos com Número de Estrelas e Número de Forks;
- Quantidade de issues afeta no crescimento do projeto, e consequentemente, na quantidade de estrelas em relação ao tempo;
- Quais linguagens mais utilizadas em relação a projetos mais relevantes (Número de Estrela/Número de Forks);
- Influência da licença do software em relação a número de forks realizado;
- Colaboradores que contribuem em diversos projetos;

#### **DUVIDAS**:

- QUESTÕES;
- INTERROGAÇÕES;
- INQUISIÇÕES;
- PERQUIRIÇÕES;
- OBJEÇÕES;
- CISÕES;
- INDAGAÇÕES; OU
- OBJETIVAÇÕES.

## FIM