

# Relatório sobre as issues

Felipe Emerson de O. Calixto

2023-07-14

## Overview dos dados de entrada

```
glimpse(data)
```

```
## Rows: 52
## Columns: 9
## $ resolved_past      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 12, 0, 12, 14, 0, 1,~
## $ resolved_inside    <int> 0, 1, 5, 0, 0, 0, 0, 0, 0, 0, 0, 5, 0, 0, 6, 0, 0, 0,~
## $ not_resolved_past  <int> 10000, 9999, 9992, 10000, 10000, 10000, 10000, 999~
## $ not_resolved_inside <int> 0, 0, 3, 0, 0, 0, 0, 0, 1, 3, 8, 22, 11, 7, 27, 0, 1,~
## $ total_resolved     <int> 0, 1, 5, 0, 0, 0, 0, 0, 0, 0, 7, 17, 0, 12, 20, 0, 1,~
## $ total_not_resolved <int> 10000, 9999, 9995, 10000, 10000, 10000, 10000, 100~
## $ total              <int> 10000, 10000, 10000, 10000, 10000, 10000, 10000, 1~
## $ repo               <chr> "accumulo", "accumulo", "accumulo", "accumulo", "a~
## $ pr_number          <int> 1379, 1414, 1545, 1667, 1706, 1758, 1772, 1795, 19~
```

## Sumário

Tamanho da amostra por repositório (cada unidade da amostra é um PR)

```
data %>%
  group_by(repo) %>%
  summarise(count = n())
```

```
## # A tibble: 3 x 2
##   repo      count
##   <chr>      <int>
## 1 accumulo    32
## 2 commons-io  14
## 3 maven-surefire 6
```

Qnt. de issues por PR por repositório

```
data %>%
  group_by(repo) %>%
  ggplot(aes(y=total, x=repo)) +
  geom_point(aes(color = repo)) +
  labs(title = "Qnt. de issues por PR de cada repositório",
        color = "Repositório")
  ) +
  xlab("Qnt. de issues por PR") +
  ylab("Repositório")
```



## Média da porcentagem de issues novas não resolvidas

Obs.: Considerando apenas PRs que possuem issues novas.

```
sprintf("A média da porcentagem de issues novas não resolvidas nos PRs que possuem issues novas é %.2f%%")
```

```
## [1] "A média da porcentagem de issues novas não resolvidas nos PRs que possuem issues novas é 82.25%"
```

## Métricas

Considerando as métricas: - Porcentagem de issues antigas resolvidas - Porcentagem de issues novas resolvidas  
- Porcentagem de issues novas não resolvidas - Porcentagem de issues resolvidas

```
# Exibir sumário das métricas
data_grouped %>%
  summarise(
    resolved_past_percent = mean(resolved_past_percent, na.rm = TRUE),
    resolved_inside_percent = mean(resolved_inside_percent, na.rm = TRUE),
    not_resolved_inside_percent = mean(not_resolved_inside_percent, na.rm = TRUE),
    total_resolved_percent = mean(total_resolved_percent, na.rm = TRUE)
  )
```

```
## # A tibble: 52 x 6
## # Groups:   repo [3]
##   repo      pr_number resolved_past_percent resolved_inside_percent
##   <chr>      <int>          <dbl>          <dbl>
## 1 accumulo      1379            0            NaN
## 2 accumulo      1414            0            1
## 3 accumulo      1545            0           0.625
## 4 accumulo      1667            0            NaN
## 5 accumulo      1706            0            NaN
## 6 accumulo      1758            0            NaN
## 7 accumulo      1772            0            NaN
## 8 accumulo      1795            0            0
## 9 accumulo      1947            0            0
## 10 accumulo     2234           0.000941            0
## # i 42 more rows
## # i 2 more variables: not_resolved_inside_percent <dbl>,
## #   total_resolved_percent <dbl>
```

## Análise das issues não resolvidas

### Taxa de Remoção de Violações (TRV)

```
data %>%
  group_by(repo) %>%
  summarise(not_resolved_past_sum = sum(not_resolved_past),
            total_sum = sum(total_not_resolved + resolved_past)
  , trv = (total_sum - not_resolved_past_sum) / not_resolved_past_sum * 100)
```

```
## # A tibble: 3 x 4
##   repo      not_resolved_past_sum total_sum   trv
##   <chr>          <int>      <int> <dbl>
## 1 accumulo      297926    298901 0.327
## 2 commons-io     32151    32192 0.128
## 3 maven-surefire   6356     6356 0
```

Os projetos accumulo e o commons-io, apresentaram TRV de 0,33% e 0,13%, respectivamente. Isto é, há uma redução bem pequena no débito técnico. A amostra do maven-surefire não apresentou nenhuma issue resolvida, provavelmente por conta da amostra ser muito pequena (7).

## Taxa de Remoção de Violações Velhas (TRV velhas)

```
data %>%
  group_by(repo) %>%
  summarize(
    soma_total = sum(not_resolved_past + resolved_past),
    soma_not_resolved_past = sum(not_resolved_past)
  ) %>%
  mutate(trvv = ((soma_total - soma_not_resolved_past) / soma_total) * 100)
```

```
## # A tibble: 3 x 4
##   repo          soma_total soma_not_resolved_past  trvv
##   <chr>          <int>          <int>    <dbl>
## 1 accumulo      298593          297926  0.223
## 2 commons-io    32159           32151  0.0249
## 3 maven-surefire 6356            6356    0
```

A TRV de violações velhas deu 0,22% e 0,02%, respectivamente, para os projetos accumulo e commons-io. Valores bem baixos de redução de violações velhas. A amostra do maven-surefire não apresentou nenhuma issue velha resolvida, provavelmente por conta da amostra ser muito pequena (7).

## Taxa de Remoção de Violações novas (TRV novas)

```
data %>%
  group_by(repo) %>%
  summarize(
    soma_total = sum(not_resolved_inside + resolved_inside),
    soma_not_resolved_inside = sum(not_resolved_inside)
  ) %>%
  mutate(trvn = ((soma_total - soma_not_resolved_inside) / soma_total) * 100)
```

```
## # A tibble: 3 x 4
##   repo          soma_total soma_not_resolved_inside  trvn
##   <chr>          <int>          <int>    <dbl>
## 1 accumulo      337            308    8.61
## 2 commons-io    40             33   17.5
## 3 maven-surefire 0              0    NaN
```

A TRV de violações novas deu 8,60% e 17,5%, respectivamente, para os projetos accumulo e commons-io. Valores dentro do intervalo de redução de violações do artigo original. A amostra do maven-surefire não apresentou nenhuma issue nova resolvida, provavelmente por conta da amostra ser muito pequena (7).