Relatório sobre os PRs filtrados

Felipe Emerson de O. Calixto

2023-07-15

Overview dos dados

PRs

Total de PRs

```
data %>% count()
```

```
##     n
## 1 826
```

Quantidade de PRs por repositório

```
data %>%
    group_by(repo) %>%
    count()
```

```
## # A tibble: 3 x 2
## # Groups:   repo [3]
##   repo               n
##   <chr>          <int>
## 1 accumulo         781
## 2 commons-io        28
## 3 maven-surefire    17
```

Quantidade de PRs que possuem PR_commit como primeiro commit (geral)

```
## [1] "Quantidade de PRs com PR_commit como primeiro commit: 725"
```

```
## [1] "Quantidade de PRs com PR_commit nao sendo primeiro commit: 101"
```

```
## [1] "Porcentagem de PRs com PR_commit como primeiro commit: 87.77%"
```

```
## [1] "Porcentagem de PRs com PR_commit nao sendo primeiro commit: 12.23%"
```

Quantidade de PRs que possuem PR_commit como primeiro commit (por repo)

```
# Calcular a quantidade de PRs com is_pr_commit_first como true e o complemento agrupado por repo
qnt_PRs_true_por_repo <- data %>%
  group_by(repo) %>%
  summarize(qnt_PRs_true = sum(is_pr_commit_first == TRUE),
            qnt_PRs_false = sum(is_pr_commit_first == FALSE))

# Calcular as porcentagens correspondentes
qnt_PRs_true_por_repo <- qnt_PRs_true_por_repo %>%
  mutate(porcentagem_PRs_true = (qnt_PRs_true / (qnt_PRs_true + qnt_PRs_false)) * 100,
         porcentagem_PRs_false = (qnt_PRs_false / (qnt_PRs_true + qnt_PRs_false)) * 100)

qnt_PRs_true_por_repo
```

```
## # A tibble: 3 x 5
##   repo     qnt_PRs_true qnt_PRs_false porcentagem_PRs_true porcentagem_PRs_false
##   <chr>           <int>         <int>                <dbl>                 <dbl>
## 1 accumulo          693            88                 88.7                  11.3
## 2 commons~           22             6                 78.6                  21.4
## 3 maven-s~           10             7                 58.8                  41.2
```

Commits

Total de commits

```
sum(data$qnt_commits)
```

```
## [1] 4236
```

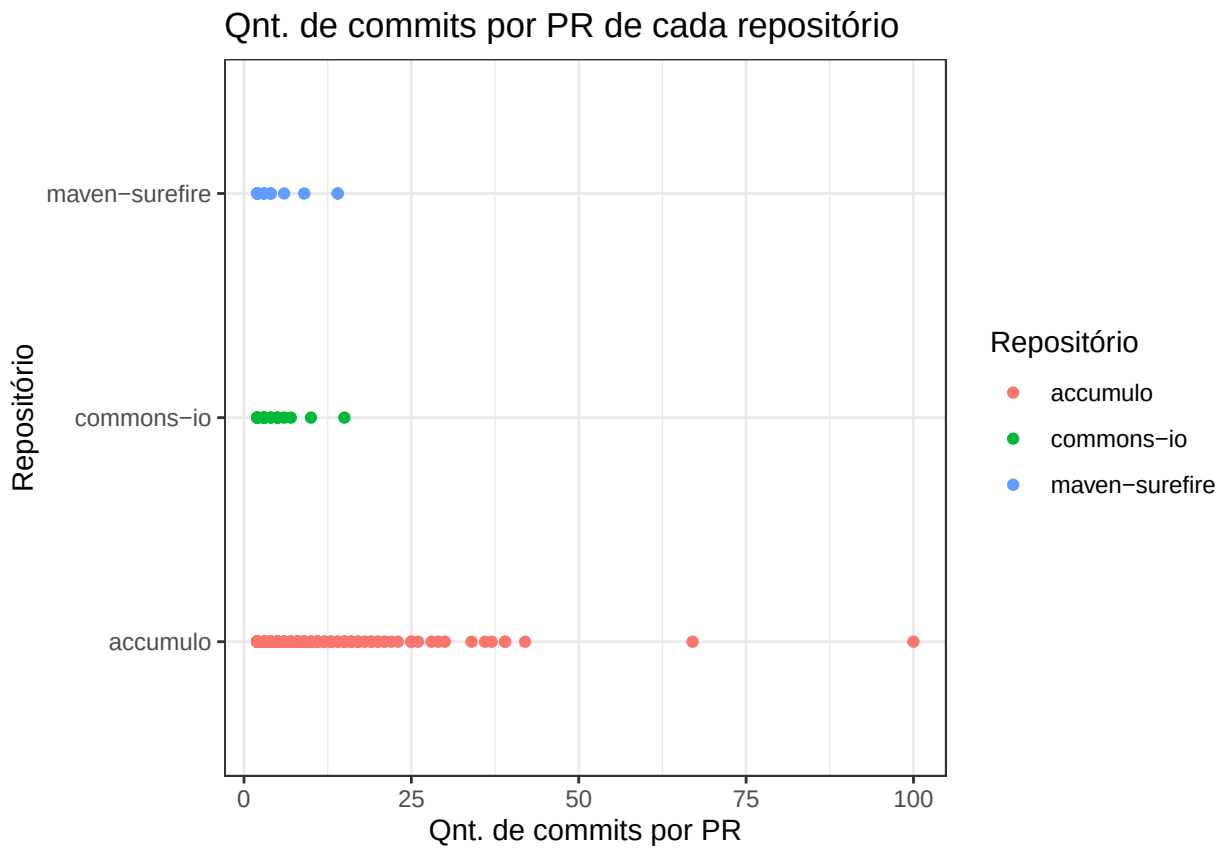
Total de commits por repo

```
data %>%
    group_by(repo) %>%
    summarize(total_commits = sum(qnt_commits))
```

```
## # A tibble: 3 x 2
##   repo           total_commits
##   <chr>                  <int>
## 1 accumulo                4048
## 2 commons-io               110
## 3 maven-surefire            78
```

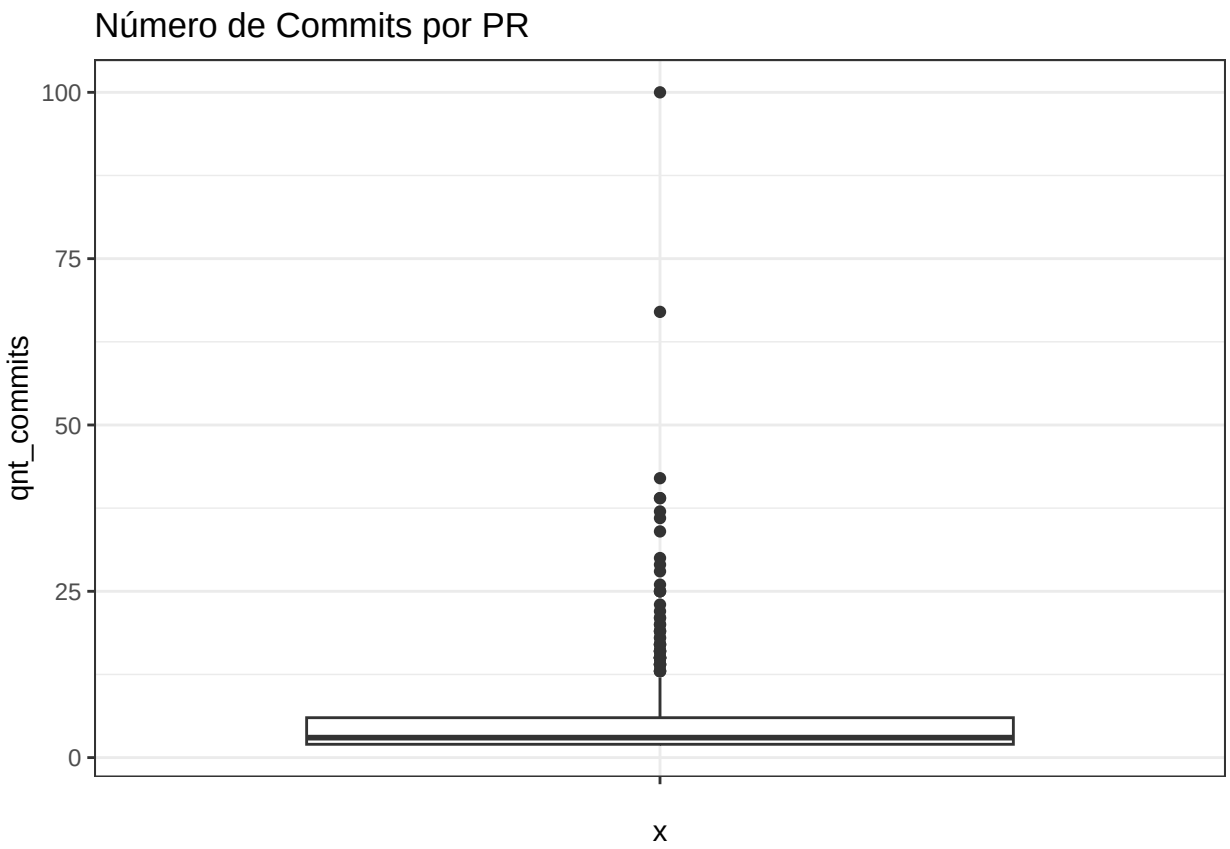
```
data %>%
    group_by(repo) %>%
    ggplot(aes(x=qnt_commits, y=repo)) +
    geom_point(aes(color = repo)) +
```

```
    labs(title ="Qnt. de commits por PR de cada repositório",
         color = "Repositório"
    ) +
    xlab("Qnt. de commits por PR") +
    ylab("Repositório")
```



Boxplot quantidade de commits por PR

```
ggplot(data, aes(x = "", y = qnt_commits)) +
  geom_boxplot() +
  labs(title = "Número de Commits por PR")
```



Sumário commits por PR

```
summary(data$qnt_commits)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   2.000   3.000   5.128   6.000 100.000
```

Frequência da quantidade de commits distintas por PR em ordem decrescente

```
# Calcular a porcentagem da frequência do número de commits distintos por PR e mostrar em ordem decresc
data %>%
  count(qnt_commits) %>%
  mutate(porcentagem = prop.table(n) * 100) %>%
  arrange(desc(porcentagem))
```

```
##    qnt_commits   n porcentagem
## 1            2 286  34.6246973
## 2            3 164  19.8547215
## 3            4 113  13.6803874
## 4            5  55   6.6585956
## 5            6  38   4.6004843
```

```
## 6            7  32   3.8740920
## 7            9  29   3.5108959
## 8            8  25   3.0266344
## 9           10  17   2.0581114
## 10          11  10   1.2106538
## 11          15   8   0.9685230
## 12          12   5   0.6053269
## 13          14   5   0.6053269
## 14          16   5   0.6053269
## 15          13   4   0.4842615
## 16          17   4   0.4842615
## 17          19   3   0.3631961
## 18          25   3   0.3631961
## 19          18   2   0.2421308
## 20          20   2   0.2421308
## 21          21   2   0.2421308
## 22          39   2   0.2421308
## 23          22   1   0.1210654
## 24          23   1   0.1210654
## 25          26   1   0.1210654
## 26          28   1   0.1210654
## 27          29   1   0.1210654
## 28          30   1   0.1210654
## 29          34   1   0.1210654
## 30          36   1   0.1210654
## 31          37   1   0.1210654
## 32          42   1   0.1210654
## 33          67   1   0.1210654
## 34         100   1   0.1210654
```