Estudo de caso para Journal of the Brazilian Computer Society (JBCS)

5 de janeiro, 2019

RQ.1: Does our proposed approach, based on meta-programming, generate correct QDMs?

Não será respondida com R

RQ.2: Does the use of a DSL reduce the effort necessary to implement the test cases that target QDM generation?

Carrega os dados dos chamadores:

```
chamadores <- read.csv("chamadores.csv",
               header = TRUE, quote="\"",
               stringsAsFactors= TRUE, strip.white = TRUE)
```

Cria colunas adicionais:

```
chamadores <- mutate(chamadores, diferenca = java - dsl)

# o chamador dsl corresponde a qtos % do chamador java
chamadores <- mutate(chamadores, percent_dsl = (dsl * 100)/java)

# o chamador dsl eh qtos % menor do que o chamador java
chamadores <- mutate(chamadores, percent_menor = 100-percent_dsl)
```

Entendimento dos Dados:

Descrição dos atributos:

```
str(chamadores)
```

```
## 'data.frame':    200 obs. of  6 variables:
##  $ chamador     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ dsl          : int  10 10 10 11 15 9 18 12 9 11 ...
##  $ java         : int  26 25 32 52 41 16 21 25 11 23 ...
##  $ diferenca    : int  16 15 22 41 26 7 3 13 2 12 ...
##  $ percent_dsl  : num  38.5 40 31.2 21.2 36.6 ...
##  $ percent_menor: num  61.5 60 68.8 78.8 63.4 ...
```

Summary:

```
# dsl
summary(chamadores$dsl)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.000   8.000   9.000   9.675  11.000  18.000
```

```r
#resumo da dsl
chamadores %>%
  summarise(media = mean(dsl), desvio_padrao = sd(dsl), mediana = median(dsl))
```

```
##   media desvio_padrao mediana
## 1 9.675      2.014863       9
```

```r
#java
summary(chamadores$java)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   15.00   23.00   28.66   39.00  130.00
```

```r
#resumo do java
chamadores %>%
  summarise(media = mean(java), desvio_padrao = sd(java), mediana = median(java))
```

```
##   media desvio_padrao mediana
## 1 28.66      18.47107      23
```

```r
#diferenca
summary(chamadores$diferenca)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    5.00   13.00   18.98   27.50  121.00
```

```r
#resumo da diferenca
chamadores %>%
  summarise(media = mean(diferenca), desvio_padrao = sd(diferenca), mediana = median(diferenca))
```

```
##    media desvio_padrao mediana
## 1 18.985      18.24608      13
```

```r
#percent_menor
summary(chamadores$percent_menor)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.29   36.72   56.26   54.66   73.88   93.08
```

```r
#resumo do percent_menor
chamadores %>%
  summarise(media = mean(percent_menor), desvio_padrao = sd(percent_menor), mediana = median(percent_mer
```

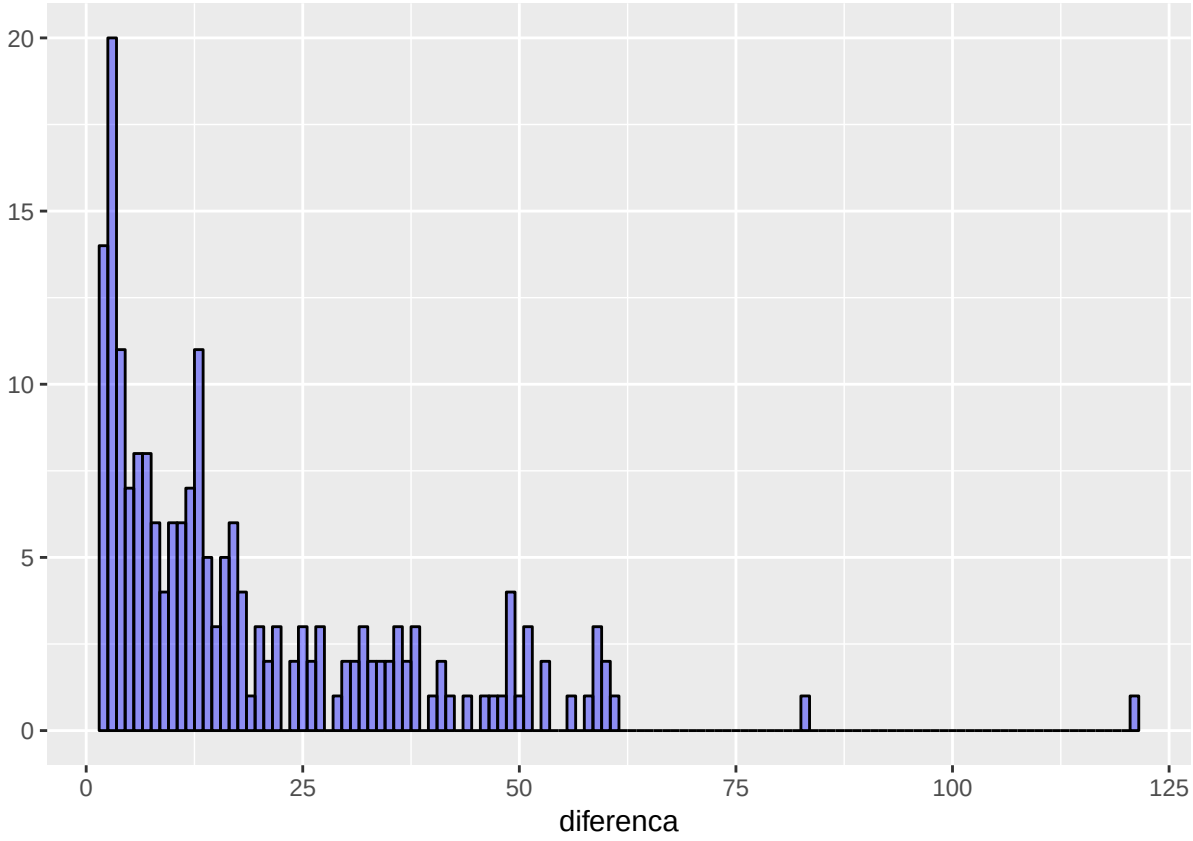
```
##      media desvio_padrao  mediana
## 1 54.65655      21.67773 56.26087
```

```r
#forte correlacao
cor(chamadores$java,chamadores$diferenca)
```

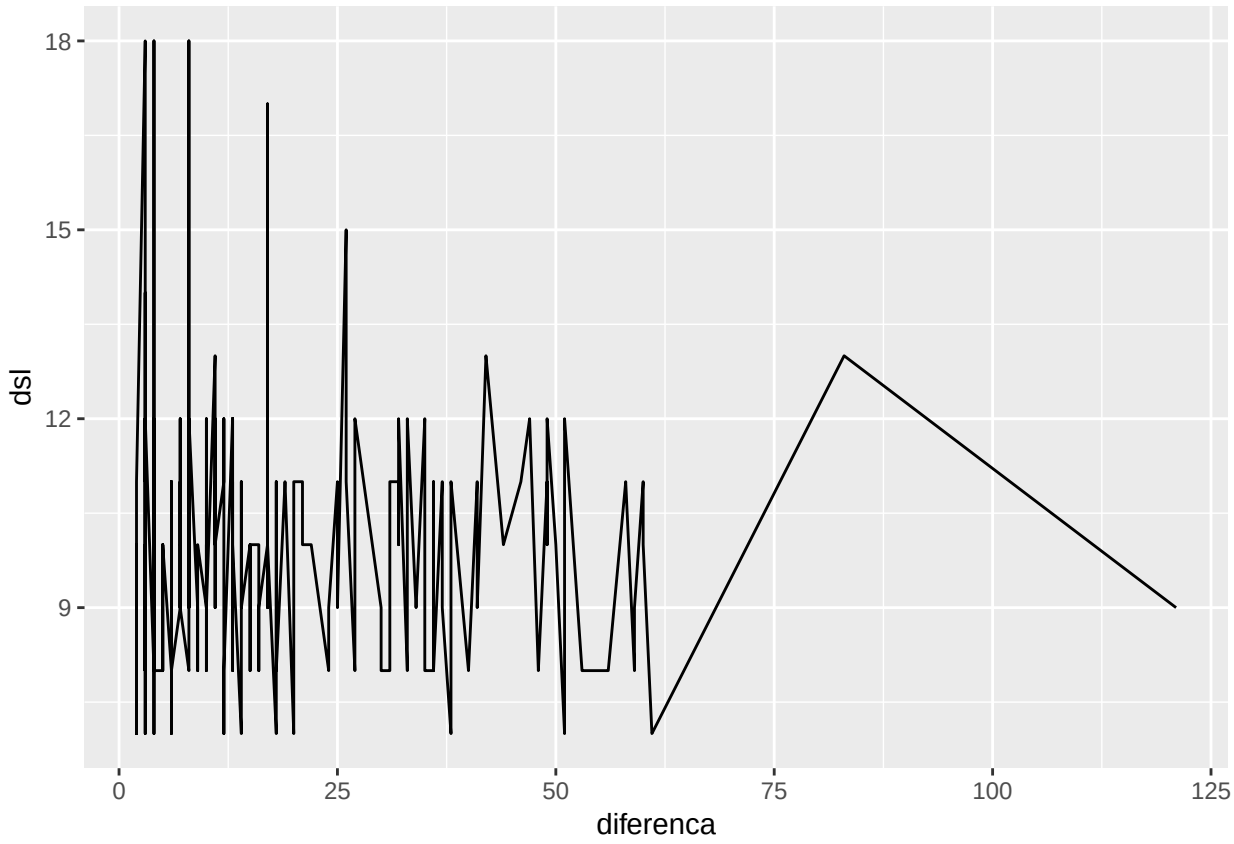
```
## [1] 0.9940523
```

Plots:

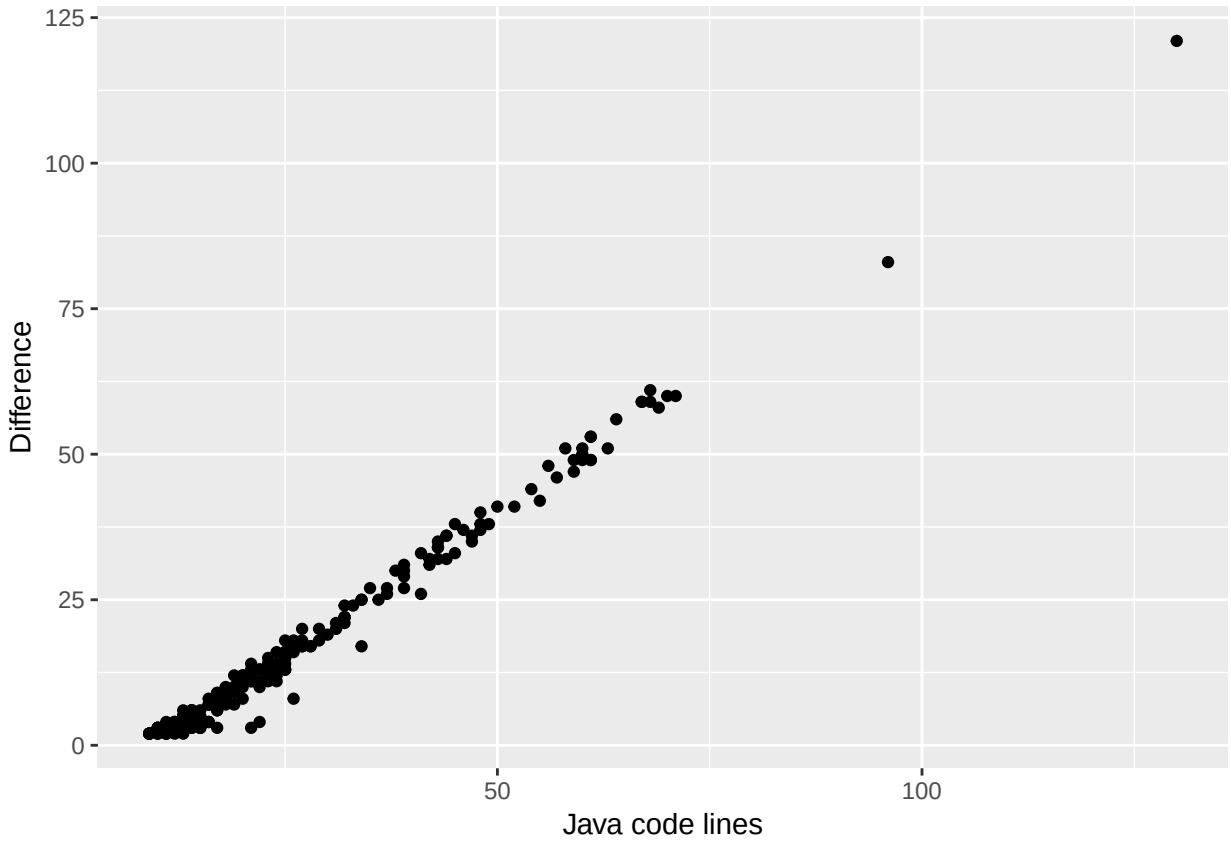
```r
#histograma da diferenca
qplot(x = diferenca, data = chamadores, geom = "histogram", binwidth = 1, fill=I("blue"), col=I("black"
```




```
#nao tem uma relacao clara entre diferenca e dsl
qplot(x = diferenca, y = dsl, data = chamadores, geom = "line")
```




```
#Quanto maior o chamador em Java maior a diferenca (mas na verdade eh relacionada a qtde de regras)
qplot(x = java, y = diferenca, data = chamadores, geom = "point", xlab = "Java code lines", ylab = "Dif
```



RQ3: Does our proposed approach, based on meta-programming, generate QDMs within a satisfactory time-frame?

Carrega os dados:

```
tempo <- read.csv("tempo.csv", header = TRUE)
```

Analise:

```
#correlacao fraca entre a qtde de cargos e o tempo de geracao
cor(tempo$cargos,tempo$X1000)
```

```
## [1] 0.1565016
```

```
cor(tempo$cargos,tempo$X2000)
```

```
## [1] 0.07159688
```

```
#sem correlacao entre os tempos (1000 e 2000 chamadores)
cor(tempo$X1000,tempo$X2000)
```

```
## [1] 0.3038842
```

```
#correlacao fraca entre a qtde total de militares e o tempo de geracao
cor(tempo$total,tempo$X1000)
```

```
## [1] 0.1532764
```

```r
cor(tempo$total,tempo$X2000)
```

```
## [1] 0.02690712
```

```r
#media dos tempos de geracao
mean(tempo$X1000)
```

```
## [1] 4925.57
```

```r
mean(tempo$X2000)
```

```
## [1] 9337.9
```

```r
#sumario dos tempos
summary(tempo$X1000)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4662    4833    4910    4926    4997    6185
```

```r
summary(tempo$X2000)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8650    9155    9284    9338    9504   10098
```

```r
#sumario de cargos e total
summary(tempo$cargos)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00   39.75   87.50  126.98  176.50 1235.00
```

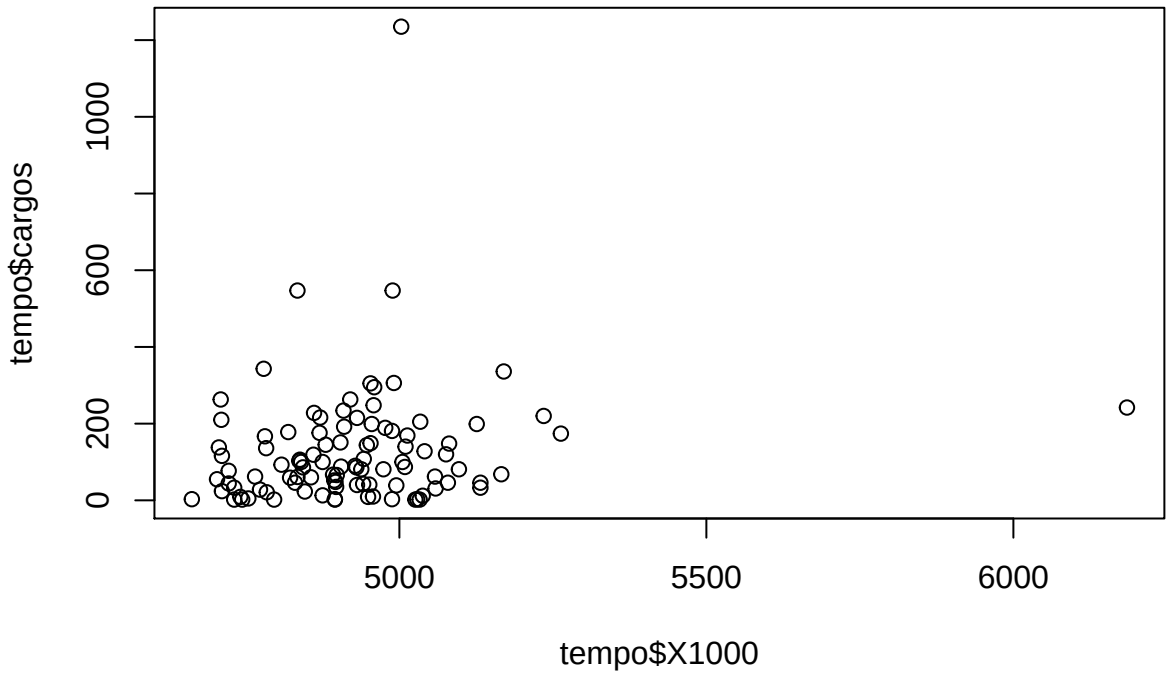
```r
summary(tempo$total)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     2.0    68.5   189.0   312.3   445.8  3068.0
```

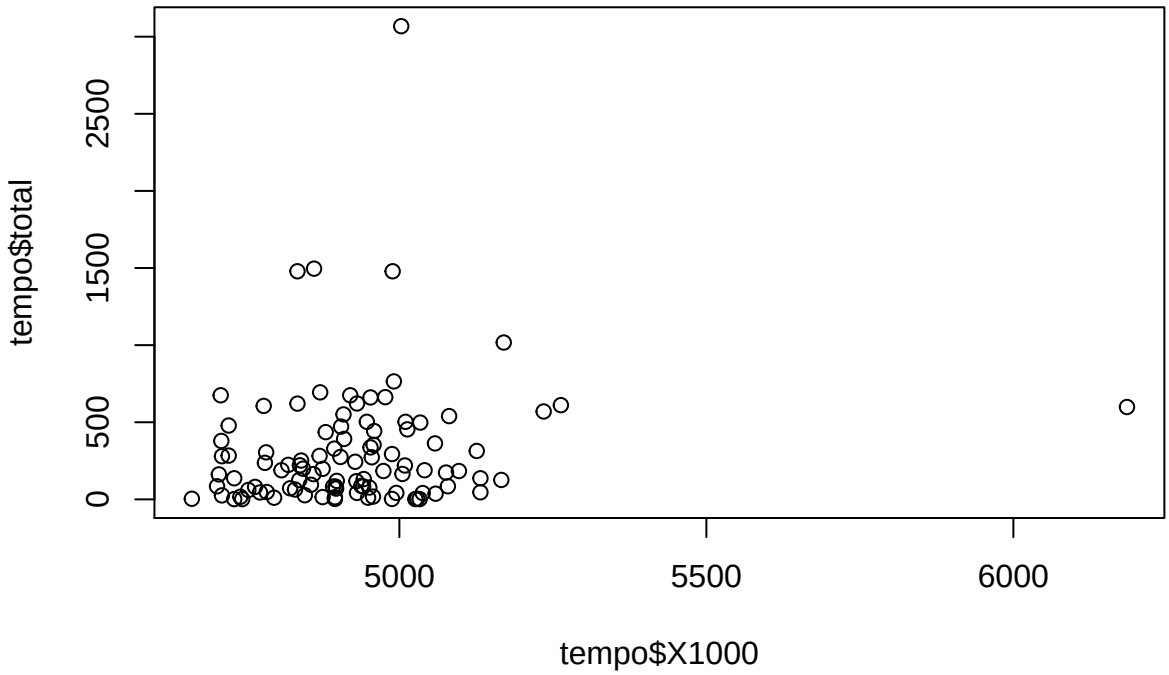
```r
#resumo do tempo (100 chamadores)
tempo %>%
  summarise(media = mean(X1000), desvio_padrao = sd(X1000), mediana = median(X1000), minimo=min(X1000),
```

```
##     media desvio_padrao mediana minimo maximo
## 1 4925.57      179.7946  4909.5   4662   6185
```

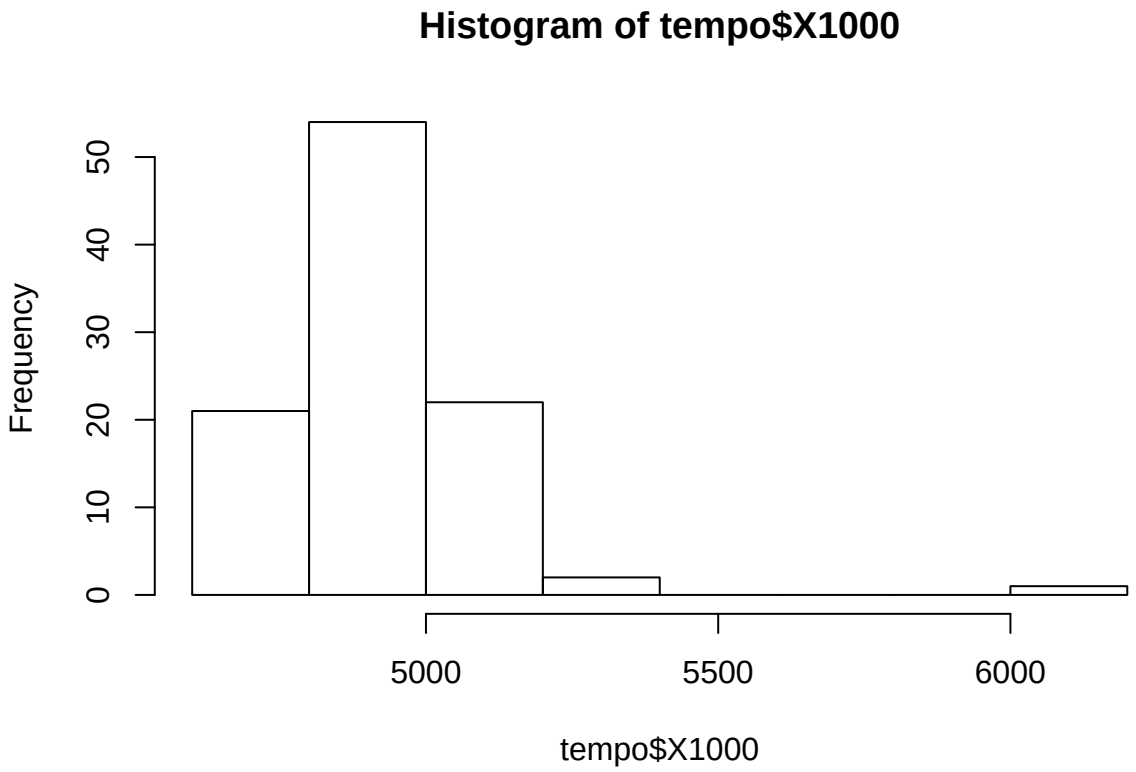
```r
#plots so pra ter nocao, mesmo sabendo q nao tem correlacao
plot(x = tempo$X1000, y = tempo$cargos)
```




```
plot(x = tempo$X1000, y = tempo$total)
```




```
#histograma do tempo (1000 chamadores)
hist(tempo$X1000)
```




```
#barplot(tempo$X1000)



#aux <- tempo[1,]
#aux
#c(aux[4:length(aux)])
#aux[4:length(aux)]

#tmp <- tempo[,4:length(tempo)]
#tmp



#plot(aux[4:length(aux)], pch = '|')
#plot(c(aux[4:length(aux)]))

#plot(aux[4:length(aux)])

#plot(x = qc, y = c200, data = aux, geom = "line")
```