

# Código Comentado

Fernando Bispo

25 agosto, 2021

## Contents

Motivação	1
Carregamento de pacotes	1
Importação do BD	1
Limpeza e manipulação do BD	2
Visualizando o BD pronto . . . . .	4
Criação de gráficos	4

## Motivação

O objetivo deste documento é gerar um **passo a passo comentado** da elaboração do código para a criação do relatório, a fim de possibilitar em uma consulta futura a análise do código feito e a metodologia que permitiu o sucesso da aplicação de algumas técnicas de programação, bem como avaliar os resultados e obter o entendimento da construção desse código.

## Carregamento de pacotes

```
library(dplyr)
library(ggplot2) #Criação de gráficos.
library(ggrepel) #Estilos de texto nos gráficos.
library(knitr) #Estruturação de tabelas.
library(kableExtra) #Estilização de tabelas.
```

## Importação do BD

Iniciei importando o BD para análise inicial e manipulação dos dados.

```
#Código omitido.
```

Visualizando os Bancos de dados

```
#Banco de dados de 2018.1
glimpse(disciplinas_2018_1)
```

```
## Rows: 96
## Columns: 18
## $ ID          <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, ~
```

```
## $ q1 <dbl> 5, 5, 5, 5, 5, 2, 5, 5, 2, 5, 3, 5, 5, 4, 5, 5, 5, 5, 5~
## $ q2 <dbl> 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5~
## $ q3 <dbl> 5, 5, 5, 5, 2, 4, 5, 5, 3, 5, 1, 5, 5, 5, 5, 5, 5, 5, 5~
## $ q4 <dbl> 5, 5, 5, 5, 2, 4, 5, 5, 3, 5, 1, 4, 5, 2, 5, 3, 5, 5, 5~
## $ q5 <dbl> 5, 5, 5, 5, 3, 1, 5, 5, 5, 5, 2, 5, 5, 4, 5, 5, 5, 5, 5~
## $ q6 <dbl> 5, 5, 5, 5, 3, 4, 5, 5, 3, 5, 2, 4, 5, 3, 5, 4, 5, 5, 5~
## $ q7 <dbl> 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5~
## $ q8 <dbl> 5, 5, 5, 5, 4, 4, 5, 5, 3, 5, 2, 5, 5, 3, 5, 2, 5, 5, 5~
## $ q9 <dbl> 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5~
## $ q10 <dbl> 5, 5, 2, 5, 5, 3, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5~
## $ q11 <dbl> 4, 5, 5, 5, 5, 2, 4, 5, 5, 5, 4, 5, 4, 5, 2, 5, 3, 1, 1~
## $ q12 <dbl> 5, 2, 5, 5, 5, 4, 4, 5, 4, 5, 4, 4, 4, 5, 5, 5, 4, 5, 5~
## $ comentariodisc <chr> NA, NA, NA, NA, NA, NA, "Melhor professora !", NA, NA, ~
## $ MATRICULA <dbl> 201517019, 217220657, 201517019, 201517019, 217220657, ~
## $ DISCIPLINA <chr> "ISCA82", "MAT229", "MAT045", "MAT224", "MATD43", "MATD~
## $ TURMA <dbl> 90900, 10100, 20200, 10100, 10000, 10000, 10100, 10000,~
## $ SITUACAO <chr> "aprovado", "aprovado", "aprovado", "aprovado", "aprova~
```

```
#Banco de dados de 2018.2
glimpse(disciplinas_2018_2)
```

```
## Rows: 58
## Columns: 18
## $ ID <dbl> 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,~
## $ q1 <dbl> 0, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 0, 3, 0, 5, 5, 5, 3, 5,~
## $ q2 <dbl> 0, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 2, 0, 5, 0, 5, 5, 5, 3, 5,~
## $ q3 <dbl> 0, 5, 4, 5, 5, 5, 4, 4, 5, 5, 5, 4, 0, 5, 0, 5, 5, 5, 1, 5,~
## $ q4 <dbl> 0, 4, 3, 5, 5, 5, 4, 4, 5, 5, 4, 3, 0, 4, 0, 5, 5, 5, 3, 4,~
## $ q5 <dbl> 0, 4, 4, 5, 4, 5, 5, 4, 5, 5, 4, 3, 0, 5, 0, 5, 5, 5, 4, 2,~
## $ q6 <dbl> 0, 4, 3, 5, 4, 5, 4, 3, 5, 5, 5, 4, 0, 4, 0, 5, 5, 5, 2, 4,~
## $ q7 <dbl> 0, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 2, 0, 5, 0, 5, 5, 5, 3, 3,~
## $ q8 <dbl> 0, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 0, 3, 0, 5, 5, 5, 4, 5,~
## $ q9 <dbl> 0, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 0, 5, 0, 5, 5, 5, 4, 5,~
## $ q10 <dbl> 0, 3, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 0, 5, 0, 5, 5, 3, 3, 5,~
## $ q11 <dbl> 0, 5, 5, 5, 5, 4, 4, 3, 5, 3, 4, 4, 0, 5, 0, 3, 3, 3, 5, 3,~
## $ q12 <dbl> 0, 1, 1, 1, 1, 3, 5, 3, 5, 5, 4, 5, 0, 5, 0, 5, 5, 2, 5, 5,~
## $ comentdisc <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "Professor ~
## $ MATRICULA <dbl> 216115595, 216115595, 216115595, 216115595, 216115595, 2161~
## $ DISCIPLINA <chr> "MATD57", "MATD51", "MATD50", "MATD47", "MATD48", "MAT191",~
## $ TURMA <dbl> 10000, 10000, 10100, 10100, 10100, 20200, 10000, 400, 10000~
## $ SITUACAO <chr> "reprovadoFalta", "aprovado", "aprovado", "aprovado", "apro~
```

## Limpeza e manipulação do BD

Antes de unir os BDs visualizarei os nomes das VAs afim de identificar possíveis divergências.

```
colnames(disciplinas_2018_1)
```

```
## [1] "ID" "q1" "q2" "q3"
## [5] "q4" "q5" "q6" "q7"
## [9] "q8" "q9" "q10" "q11"
## [13] "q12" "comentariodisc" "MATRICULA" "DISCIPLINA"
## [17] "TURMA" "SITUACAO"
```

```
colnames(disiplinas_2018_2)
```

```
## [1] "ID"          "q1"          "q2"          "q3"          "q4"
## [6] "q5"          "q6"          "q7"          "q8"          "q9"
## [11] "q10"         "q11"         "q12"         "comentdisc"  "MATRICULA"
## [16] "DISCIPLINA"  "TURMA"       "SITUACAO"
```

Constatei que o nome da variável Comentários da Disciplina estava diferente nos dois BDs, então renomeei em ambos os BDs e criei a VA **SEMESTRE**

```
#Renomeação das VAs.
```

```
disiplinas_2018_1 <- disciplinas_2018_1 %>%
  rename(COMENTARIOS_DISC = comentariodisc)
```

```
disiplinas_2018_2 <- disciplinas_2018_2 %>%
  rename(COMENTARIOS_DISC = comentdisc)
```

```
#Criação da VA SEMESTRE.
```

```
disiplinas_2018_1 <- disciplinas_2018_1 %>%
  mutate(SEMESTRE = 2018.1)
```

```
disiplinas_2018_2 <- disciplinas_2018_2 %>%
  mutate(SEMESTRE = 2018.2)
```

Após esse procedimento uni os dois BDs

```
#União dos BDs
```

```
disiplinas_2018 <- full_join(disiplinas_2018_1, disciplinas_2018_2)
```

```
## Joining, by = c("ID", "q1", "q2", "q3", "q4", "q5", "q6", "q7", "q8", "q9", "q10", "q11", "q12", "COMENTARIOS_DISC")
```

Percebi que as variáveis do BD não eram numéricas, então converti as VAs do BD para fator e algumas poucas para caractere.

```
#Transformação do BD para fator.
```

```
disiplinas_2018 <- disciplinas_2018 %>%
  mutate_all(as.factor)
```

```
#Transformando algumas variaveis do BD em caractere e alterando o formato de strings.
```

```
disiplinas_2018 <- disciplinas_2018 %>%
  mutate(COMENTARIOS_DISC = as.character(COMENTARIOS_DISC),
         DISCIPLINA = as.character(DISCIPLINA),
         SITUACAO = stringr::str_to_title(SITUACAO))
```

```
glimpse(disiplinas_2018) #Visualização da estrutura.
```

```
## Rows: 154
```

```
## Columns: 19
```

```
## $ ID      <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
## $ q1      <fct> 5, 5, 5, 5, 5, 2, 5, 5, 2, 5, 3, 5, 5, 4, 5, 5, 5, 5,~
## $ q2      <fct> 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5,~
## $ q3      <fct> 5, 5, 5, 5, 2, 4, 5, 5, 3, 5, 1, 5, 5, 5, 5, 5, 5,~
## $ q4      <fct> 5, 5, 5, 5, 2, 4, 5, 5, 3, 5, 1, 4, 5, 2, 5, 3, 5, 5,~
## $ q5      <fct> 5, 5, 5, 5, 3, 1, 5, 5, 5, 5, 2, 5, 5, 4, 5, 5, 5, 5,~
## $ q6      <fct> 5, 5, 5, 5, 3, 4, 5, 5, 3, 5, 2, 4, 5, 3, 5, 4, 5, 5,~
## $ q7      <fct> 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5,~
## $ q8      <fct> 5, 5, 5, 5, 4, 4, 5, 5, 3, 5, 2, 5, 5, 3, 5, 2, 5, 5,~
## $ q9      <fct> 5, 5, 5, 5, 5, 2, 5, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5,~
```

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Comentários	Matrícula	Disciplina	Si
5	5	5	5	5	5	5	5	5	5	4	5	NA	201517019	ISCA82	Ap
5	5	5	5	5	5	5	5	5	5	5	2	NA	217220657	MAT229	Ap
5	5	5	5	5	5	5	5	5	2	5	5	NA	201517019	MAT045	Ap
5	5	5	5	5	5	5	5	5	5	5	5	NA	201517019	MAT224	Ap
5	5	2	2	3	3	5	4	5	5	5	5	NA	217220657	MATD43	Ap
2	3	4	4	1	4	4	4	2	3	2	4	NA	201517019	MATD41	Ap
5	5	5	5	5	5	5	5	5	5	4	4	Melhor professora !	201510582	MAT224	Ap
5	5	5	5	5	5	5	5	5	5	5	5	NA	217220657	MATD44	Ap
2	5	3	3	5	3	5	3	5	5	5	4	NA	218123951	MAT198	Re
5	5	5	5	5	5	5	5	5	5	5	5	NA	217220657	MATD45	Ap

```
## $ q10      <fct> 5, 5, 2, 5, 5, 3, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, ~
## $ q11      <fct> 4, 5, 5, 5, 5, 2, 4, 5, 5, 5, 4, 5, 4, 5, 2, 5, 3, 1, ~
## $ q12      <fct> 5, 2, 5, 5, 5, 4, 4, 5, 4, 5, 4, 4, 4, 5, 5, 5, 4, 5, ~
## $ COMENTARIOS_DISC <chr> NA, NA, NA, NA, NA, NA, "Melhor professora !", NA, NA~
## $ MATRICULA <fct> 201517019, 217220657, 201517019, 201517019, 217220657~
## $ DISCIPLINA <chr> "ISCA82", "MAT229", "MAT045", "MAT224", "MATD43", "MA~
## $ TURMA     <fct> 90900, 10100, 20200, 10100, 10000, 10000, 10100, 1000~
## $ SITUACAO  <chr> "Aprovado", "Aprovado", "Aprovado", "Aprovado", "Apro~
## $ SEMESTRE  <fct> 2018.1, 2018.1, 2018.1, 2018.1, 2018.1, 2018.1, 2018.~
```

Identifiquei Variáveis que não trariam informação, então exclui do BD.

```
disciplinas_2018 <- disciplinas_2018 %>%
  select(-ID, -TURMA)
```

Para poder identificar e mensurar como os alunos avaliam as disciplinas do departamento, categorizei as disciplinas que pertenciam ao departamento e as que não pertencia através da criação da variável DEPARTAMENTO.

```
disciplinas_2018 <- disciplinas_2018 %>%
  mutate(DEPARTAMENTO = ifelse(DISCIPLINA %in% c("ADM001", "ISCA82", "FCHC45",
                                                "LETA09", "MAT045", "MAT174",
                                                "MATE56"), "Não_DEst", "DEst"))
```

## Visualizando o BD pronto

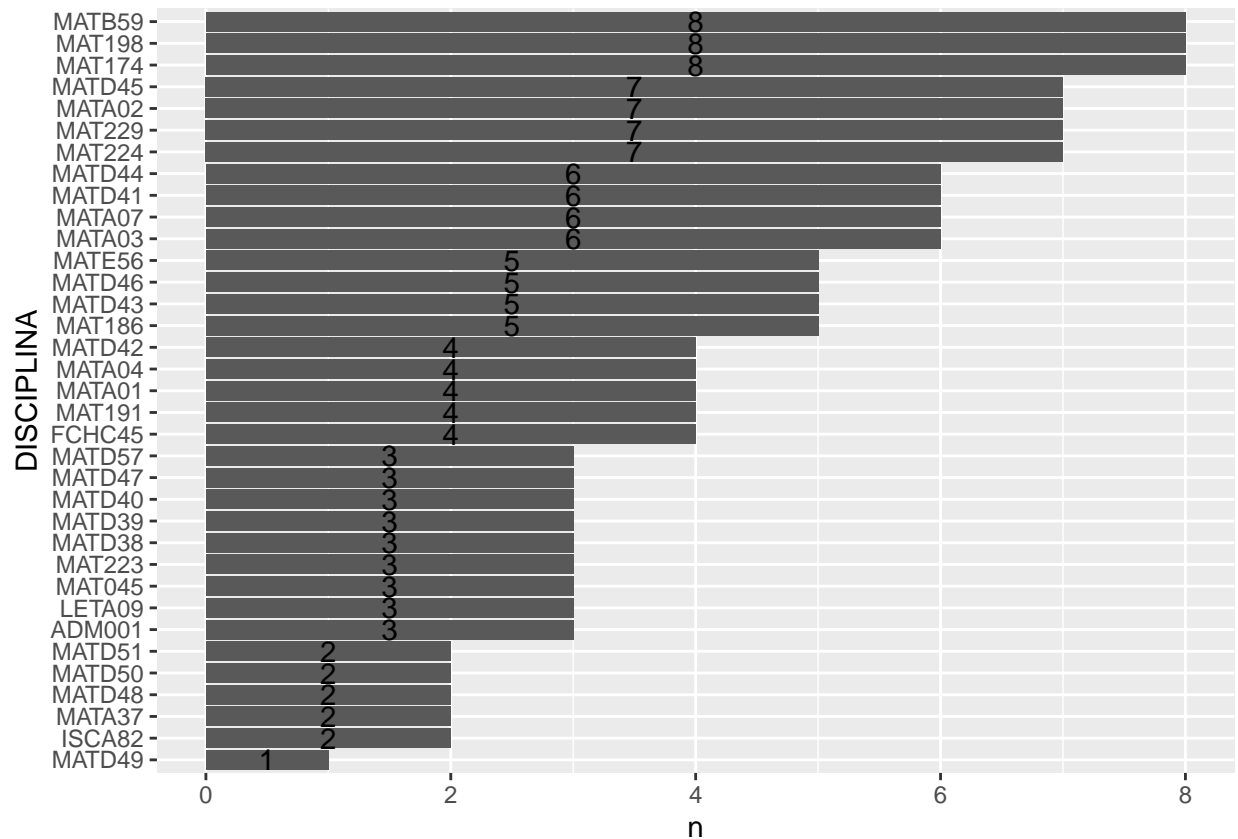
Após todas essas manipulações o BD ficou com a seguinte estrutura:

```
kable(head(disciplinas_2018,10), col.names = c("Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8", "Q9", "Q10", "Q11", "Q12", "Comentários", "Matrícula", "Disciplina", "Situação"),
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

## Criação de gráficos

Tentando replicar os gráficos feitos no relatório iniciei com o das disciplinas.

```
disciplinas_2018 %>%
  count(DISCIPLINA) %>%
  # top_n(10, n) %>% Estou na dúvida se devo listar todas as disciplinas ou apenas as que possuem mais
  mutate(DISCIPLINA = forcats::fct_reorder(DISCIPLINA, n)) %>%
  ggplot(aes(x = DISCIPLINA)) +
  geom_col(aes(y = n), show.legend = FALSE) +
  geom_text(aes(y = n/2, label = n)) +
  coord_flip()
```



Após ler o relatório que a senhora desenvolveu, tentei replicar um dos gráficos e então surgiu a primeira dúvida: Como inserir os valores das respectivas porcentagens no gráfico de forma a deixalos centralizados nos seus respectivos locais?

```
disciplinas_2018 %>%
  group_by(SEMESTRE) %>%
  count(SITUACAO) %>%
  mutate(Fr = n/sum(n)*100) %>%
  ggplot(aes(x = SEMESTRE, y = Fr, fill = SITUACAO)) +
  geom_col() +
  geom_text_repel(aes(label = round(Fr,0))) +
  labs(x = "Semestre", y = "Frequência (%)")
```

