

Dúvida sobre implementação da ferramenta para AED e elaboração de relatório

Fernando Bispo

02/07/2021

Intrudução

No intuito de refazer os passos de elaboração do relatório confeccionado inicialmente com ferramentas da Microsoft, estou utilizando os conhecimentos adquiridos até aqui com programação em R a fim de obter proficiência na manipulação da ferramenta. Iniciei a tratativa do banco de dados (BD) fornecido com as respostas do Questionário SABE Disciplinas 2018 e exponho aqui meu desenvolvimento no intuito de obter maiores orientações de como melhorar meu desempenho e auxílio para aplicar a teoria obtida até o momento na faculdade.

Bicliotecas Utiizadas

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(ggrepel)
```

Importação do BD

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

```
disciplinas_2018_1 <- readr::read_csv2("BD/SABE - Disciplinas 2018.1 e 2018.2/disciplinas_2018_1.csv")

## i Using '\',\'' as decimal and '\'.\'' as grouping mark. Use `read_delim()` for more control.
##
## -- Column specification -----
## cols(
##   ID = col_double(),
##   q1 = col_double(),
##   q2 = col_double(),
##   q3 = col_double(),
##   q4 = col_double(),
```

```
## q5 = col_double(),
## q6 = col_double(),
## q7 = col_double(),
## q8 = col_double(),
## q9 = col_double(),
## q10 = col_double(),
## q11 = col_double(),
## q12 = col_double(),
## comentariodisc = col_character(),
## MATRICULA = col_double(),
## DISCIPLINA = col_character(),
## TURMA = col_double(),
## SITUACAO = col_character()
## )

disciplinas_2018_2 <- readr::read_csv2("BD/SABE - Disciplinas 2018.1 e 2018.2/disciplinas_2018_2.csv")

## i Using '\',\'' as decimal and '\'.\'' as grouping mark. Use `read_delim()` for more control.
##
## -- Column specification -----
## cols(
##   ID = col_double(),
##   q1 = col_double(),
##   q2 = col_double(),
##   q3 = col_double(),
##   q4 = col_double(),
##   q5 = col_double(),
##   q6 = col_double(),
##   q7 = col_double(),
##   q8 = col_double(),
##   q9 = col_double(),
##   q10 = col_double(),
##   q11 = col_double(),
##   q12 = col_double(),
##   comentdisc = col_character(),
##   MATRICULA = col_double(),
##   DISCIPLINA = col_character(),
##   TURMA = col_double(),
##   SITUACAO = col_character()
## )
```

Após importação analisei os nomes das variáveis com o intuito de de juntar os dois BDs e para não haver confusão entre eles criei a variável SEMESTRE para obter essa diferenciação.

```
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(disciplinas_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 a variável comentários da disciplina estava com nome diferente, então renomeei em ambos os BDs e criei a VA SEMESTRE

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

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

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

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

Após esse procedimento uni os dois BDs

```
disciplinas_2018 <- full_join(disciplinas_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

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

#Transformando algumas variaveis do BD em caractere e alterando o formato de strings.
disciplinas_2018 <- disciplinas_2018 %>%
  mutate(COMENTARIOS_DISC = as.character(COMENTARIOS_DISC),
         DISCIPLINA = as.character(DISCIPLINA),
         SITUACAO = stringr::str_to_title(SITUACAO))

glimpse(disciplinas_2018)
```

```
## 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,~
## $ 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,~
## $ q10         <fct> 5, 5, 2, 5, 5, 3, 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.~
```

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()
```

Relatório-de-desenvolvimento_files/figure-latex/unnamed-chunk-7-1.pdf

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 (%)")
```

Relatório-de-desenvolvimento_files/figure-latex/unnamed-chunk-8-1.pdf