Análise das variáveis Saresp Questionário - moda por escola

Série 3EM

Livia Kobayashi

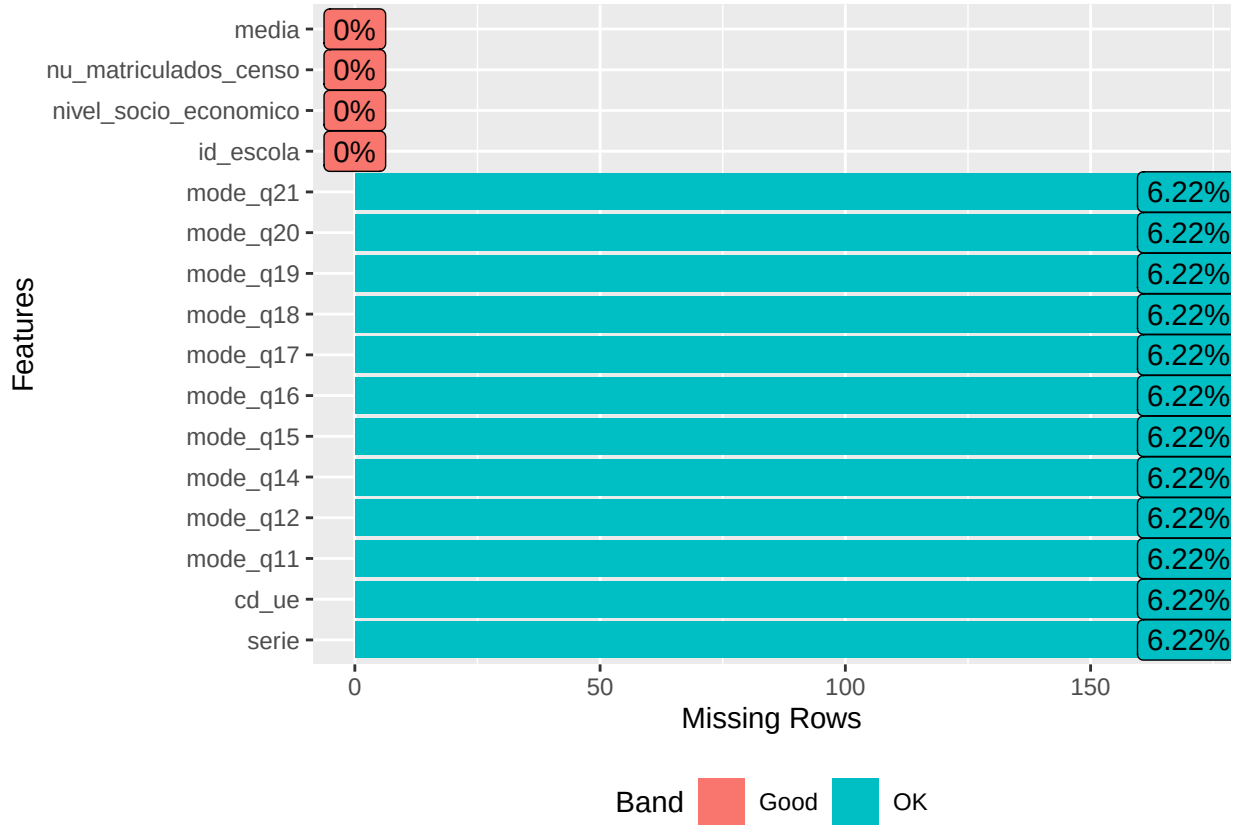
10 junho 2021

```
library(tidyverse)
library(DataExplorer)
library(gridExtra)
library(grid)
library(caret)
library(ggcorrplot)
library(vcd)
df_publico <- read.csv2("../books/df_publico.csv")
#book <- read.csv2("../books/saresp3em_mode.csv")
book <- read.csv2(params$book)
```

```
##   id_serie
## 1      3EM
```

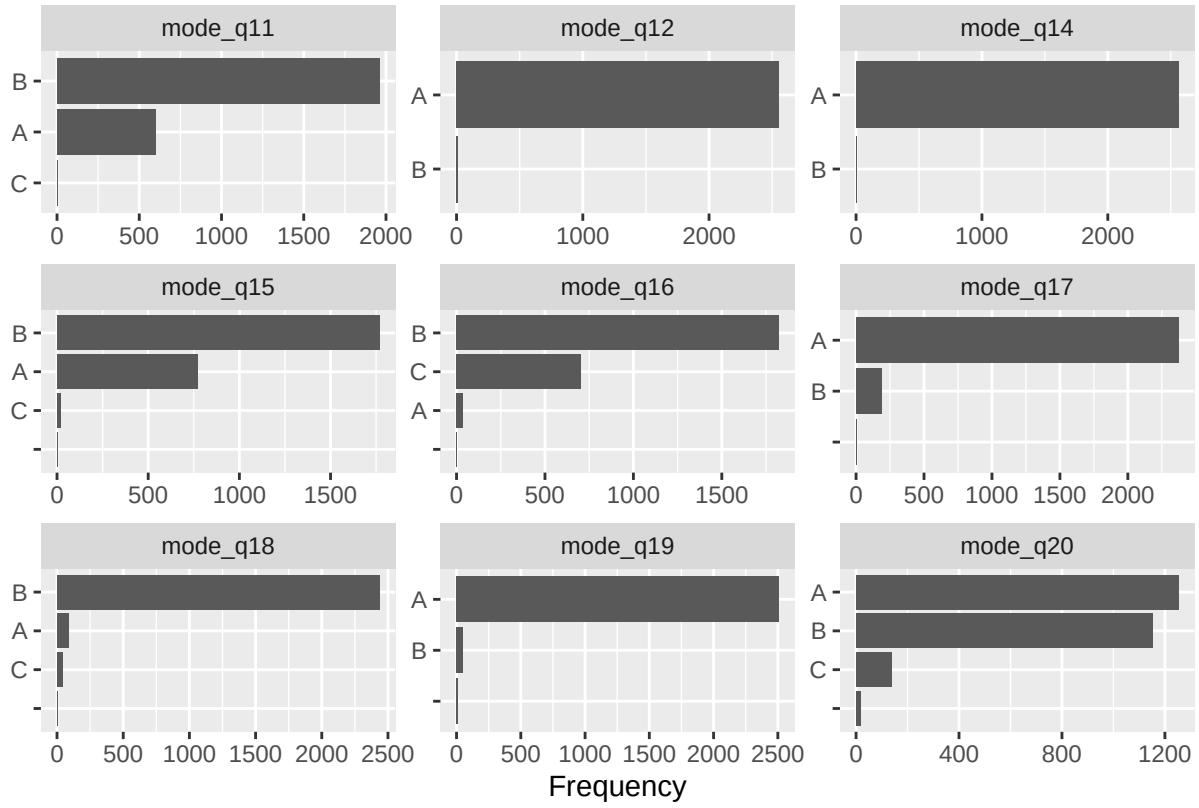
Missing

```
plot_missing(df)
```

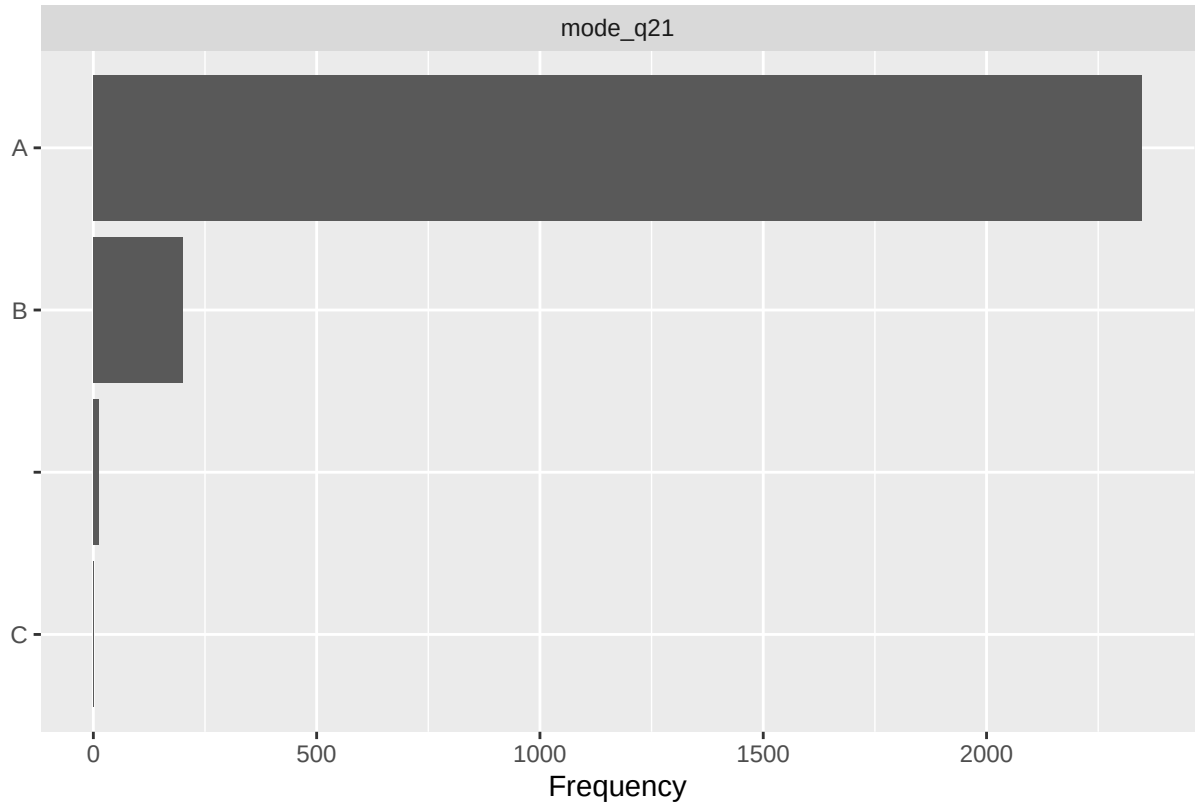


Volume

```
plot_bar(final_data)
```


mode_q11
B
A
C
0
500
1000
1500
2000
mode_q12
A
B
0
1000
2000
mode_q14
A
B
0
1000
2000
mode_q15
B
A
C
-
0
500
1000
1500
mode_q16
B
C
A
-
0
500
1000
1500
mode_q17
A
B
-
0
500
1000
1500
2000
mode_q18
B
A
C
-
0
500
1000
1500
2000
2500
mode_q19
A
B
-
0
500
1000
1500
2000
2500
mode_q20
A
B
C
-
0
400
800
1200
Frequency

Page 1

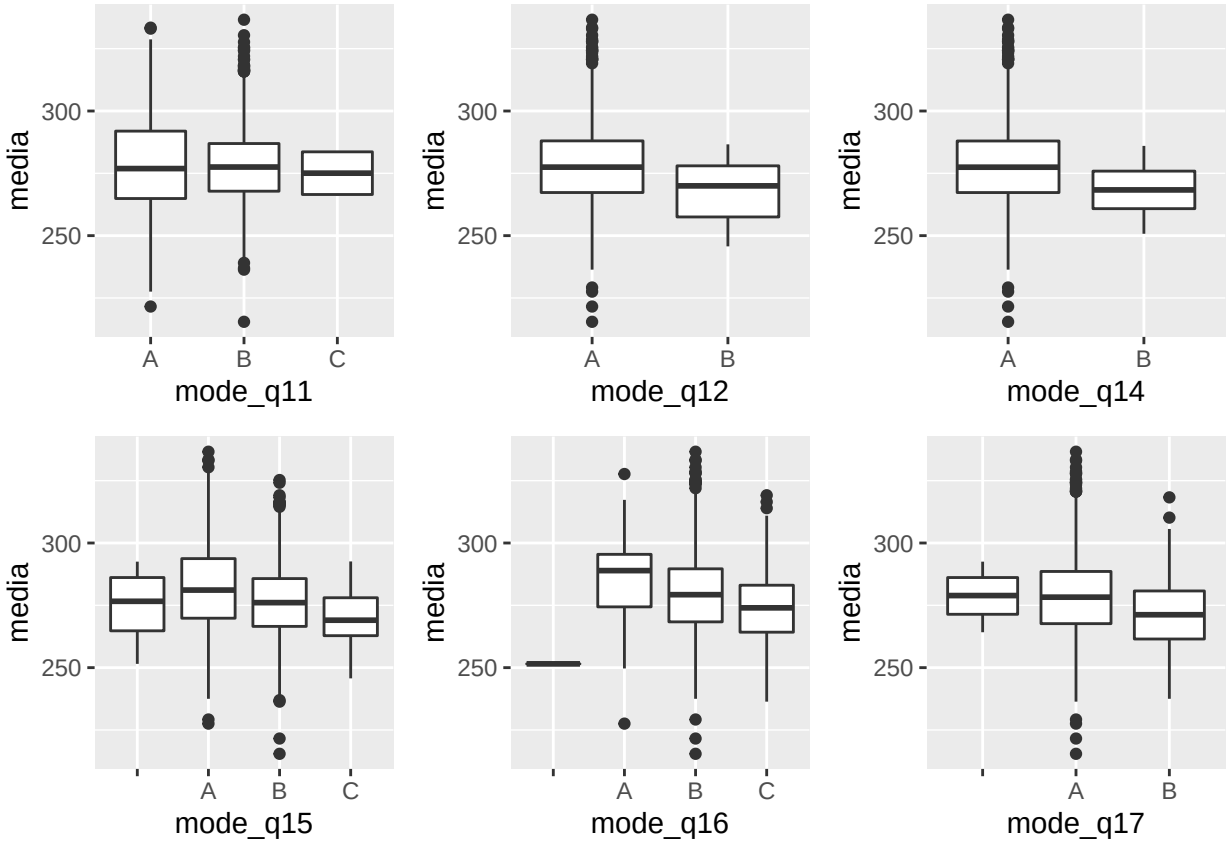


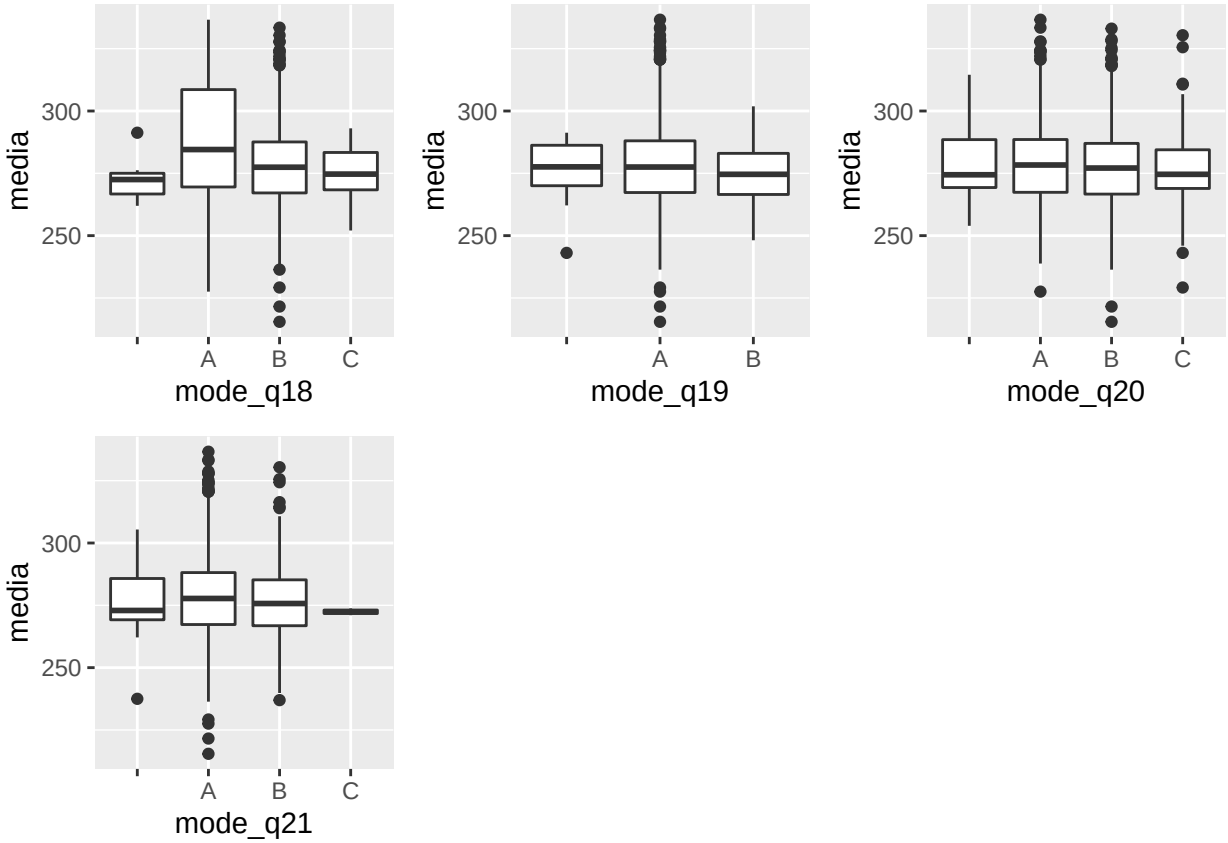
Boxplot

```
vars <- colnames(final_data)
vars <- vars[-c(1,2)]
plots <- list()
i <- 1
for (variable in vars) {
  #plots[[i]] <- plot_boxplot(final_data, by = variable)
  plots[[i]] <- ggplot(final_data, aes_string(variable, "media")) + geom_boxplot()
  i <- i + 1
}

n <- length(plots)

i <- 1
while (i <= n) {
  do.call("grid.arrange", c(plots[i:(min(i+5, n))], ncol=3, nrow = 2))
  i <- i + 6
}
```


media
300
250
A
B
C
mode_q11
media
300
250
A
B
mode_q12
media
300
250
A
B
mode_q14
media
300
250
A
B
C
mode_q15
media
300
250
A
B
C
mode_q16
media
300
250
A
B
mode_q17



Análise Univariada

```
vars <- colnames(final_data)
vars <- vars[-c(1,2)]
y_resp <- "media"

remove_cols <- nearZeroVar(df, names = TRUE)
final_cols <- setdiff(vars, remove_cols)
final_cols
```

```
## [1] "mode_q11" "mode_q15" "mode_q16" "mode_q17" "mode_q20" "mode_q21"
```

```
tb_r2 <- data.frame(var = final_cols)

rsquared <- c()
for (variable in final_cols) {
  lm_formula <- as.formula(str_glue("{y_resp} ~ {variable}"))
  model_lm <- lm(lm_formula, df)
  rsquared <- append(rsquared, summary(model_lm)$r.squared)
}

tb_r2$rsquared <- rsquared
tb_r2 %>% head(nrow(tb_r2))
```

```
##        var     rsquared
## 1 mode_q11 0.0014638002
## 2 mode_q15 0.0335452273
## 3 mode_q16 0.0296368439
## 4 mode_q17 0.0138935063
## 5 mode_q20 0.0010641542
## 6 mode_q21 0.0006030284
```

```
catcorrm <- function(vars, dat) sapply(vars, function(y) sapply(vars, function(x) assocstats(table(dat[

matriz <- catcorrm(final_cols, data_corr)

ggcorrplot(matriz, show.diag = F, type="lower", lab=TRUE, lab_size=14, show.legend = F)
```