# Análise das variáveis Saresp Questionário - moda por escola Série 9EF

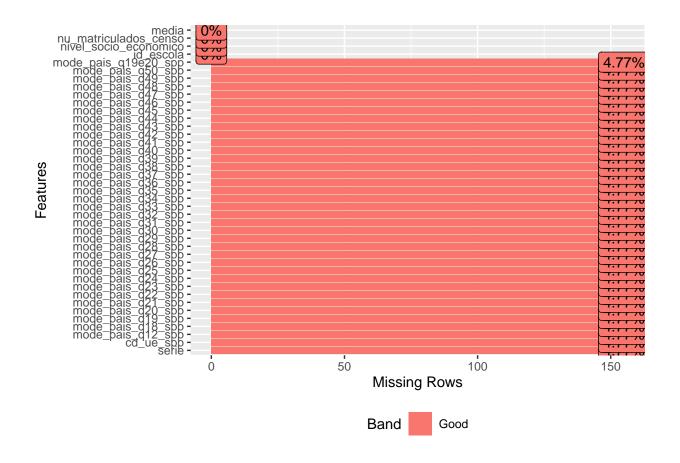
#### Livia Kobayashi

14 junho 2021

```
library(tidyverse)
library(DataExplorer)
library(gridExtra)
library(grid)
library(caret)
library(ggcorrplot)
library(vcd)
df_publico <- read.csv2("../output/books/df_publico.csv")
book <- read.csv2(params$book)</pre>
## id_serie
## 1 9EF
```

# Missing: 4,7% de dados faltantes

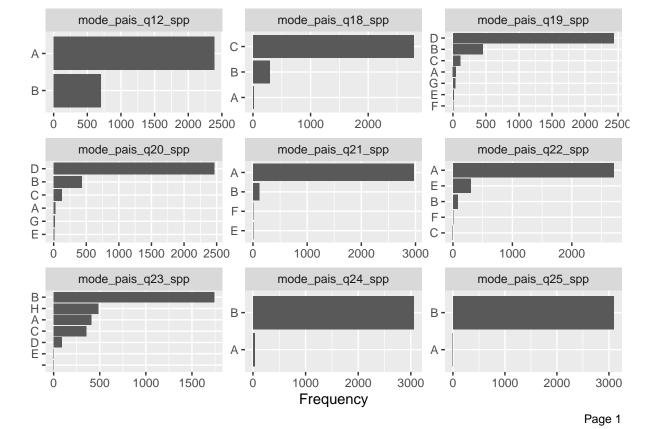
```
plot_missing(df)
```

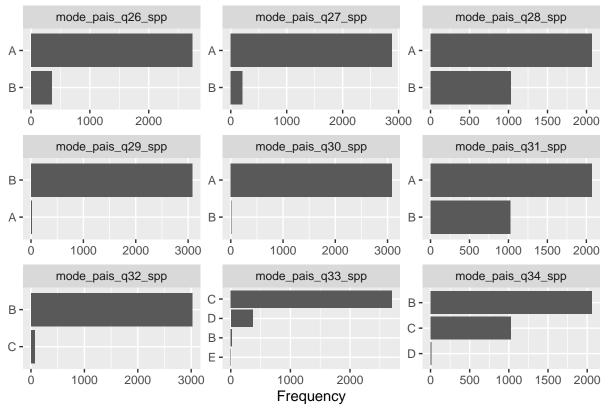


### Volume: VAriáveis com bom volume

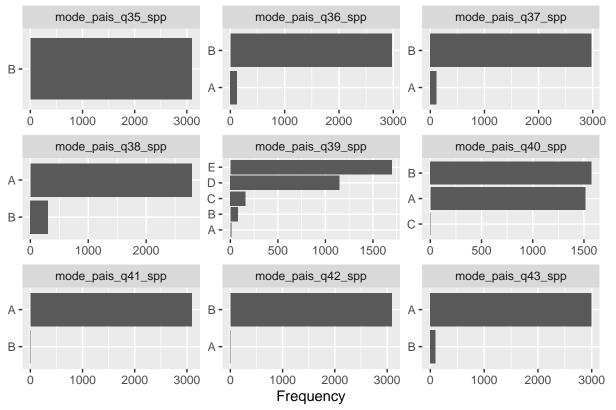
- $\bullet$  mode\_pais\_q12
- $mode_pais_q18$
- $mode_pais_q19$
- mode\_pais\_q20
- $mode_pais_q22$
- mode\_pais\_q23
- $mode\_pais\_q26$
- $mode\_pais\_q28$
- $mode\_pais\_q31$
- mode pais q33
- $mode_pais_q34$ • mode\_pais\_q38
- mode\_pais\_q39
- $mode\_pais\_q40$
- $mode\_pais\_q45$
- $\bullet$  mode\_pais\_q46
- $mode\_pais\_q47$

plot\_bar(final\_data)

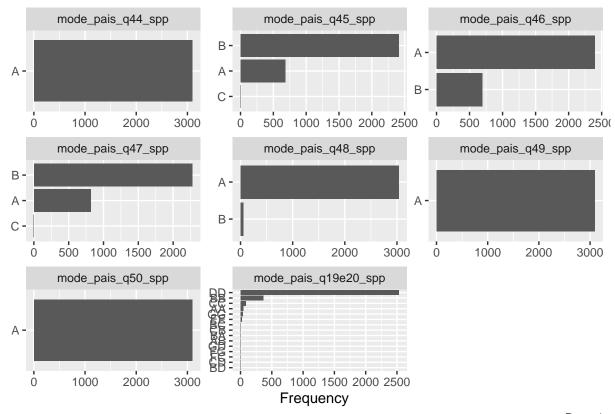




Page 2



Page 3



Page 4

#### Boxplot: Variáveis com bom volume e variância

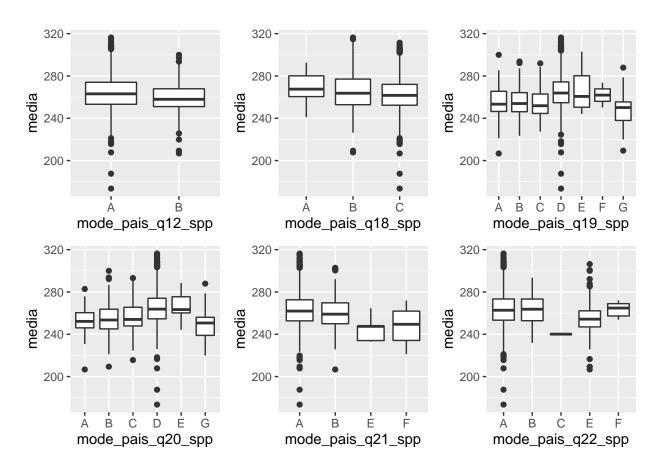
```
mode_pais_q19: graduação mãe > ordena (juntar A,B,C, G e E+F)
mode_pais_q20: graduação pai > ordena (juntar A,B,C, G)
mode_pais_q22: notas < para mães desempregadas</li>
mode_pais_q23: renda maior=> notas > (juntar A com H)
mode_pais_q31: TV assinatura => Notas >
mode_pais_q34: #TV => Notas > (B x C)
mode_pais_q38: Telefone => Notas >
```

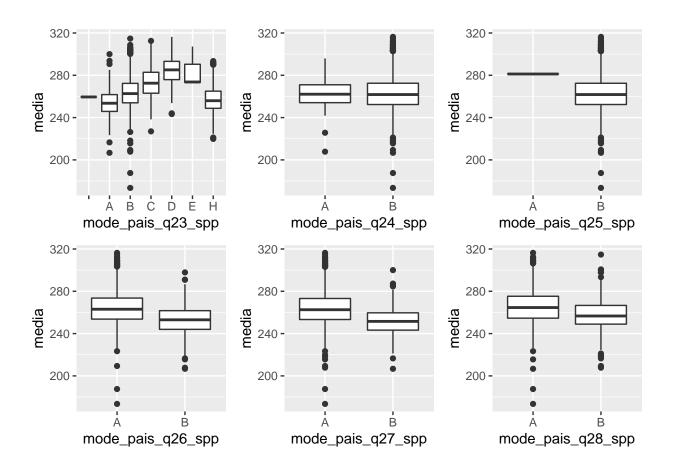
mode\_pais\_q40: COmputador => Notas > (A x B)
mode\_pais\_q46: Aspirador => Notas > (A x B)

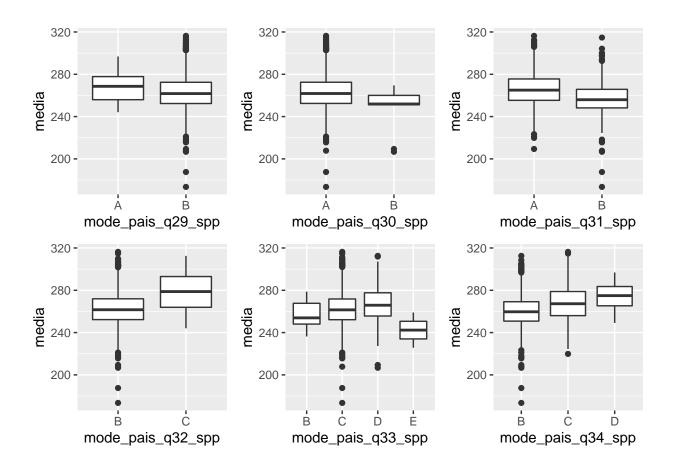
•  $mode_pais_q47$ : Carro => Notas > (A x B)

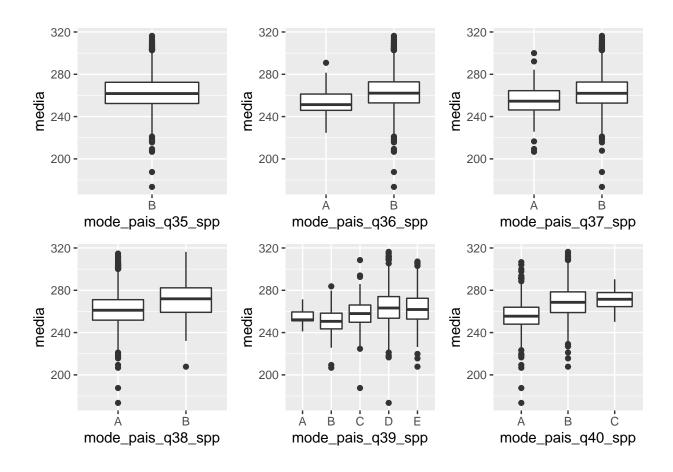
```
vars <- colnames(final_data)
vars <- vars[-c(1,2)]
plots <- list()
i <- 1
for (variable in vars) {
    #plots[[i]] <- plot_boxplot(final_data, by = variable)
    plots[[i]] <- ggplot(final_data, aes_string(variable, "media")) + geom_boxplot()
    i <- i + 1
}
n <- length(plots)</pre>
```

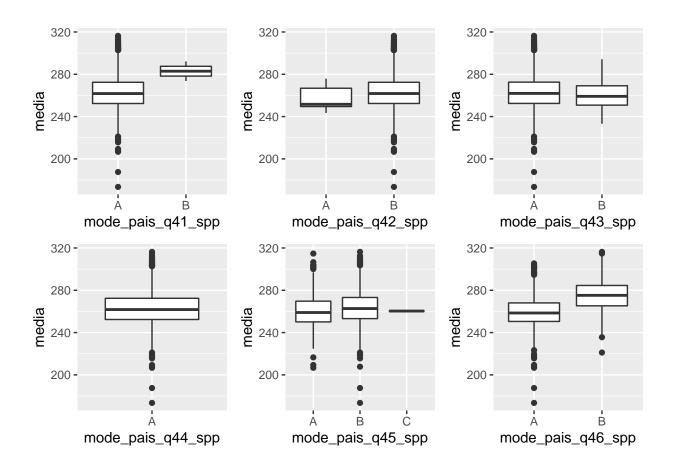
```
i <- 1
while (i <= n) {
   do.call("grid.arrange", c(plots[i:(min(i+5, n))], ncol=3, nrow = 2))
   i <- i + 6
}</pre>
```

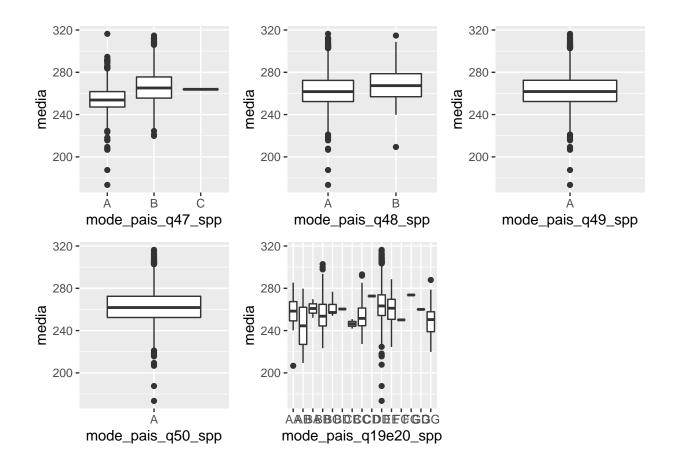












#### Análise Univariada

```
mode_pais_q40: COmputador => Notas > (A x B)
mode_pais_q23: renda maior=> notas >
mode_pais_q47: Carro => Notas > (A x B)
mode_pais_q46: Aspirador => Notas > (A x B)
```

• mode\_pais\_q31: TV assinatura => Notas >

```
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</pre>
```

```
##
    [1] "mode_pais_q12_spp"
                                "mode_pais_q18_spp"
                                                        "mode_pais_q19_spp"
##
    [4]
        "mode_pais_q20_spp"
                                "mode_pais_q22_spp"
                                                        "mode_pais_q23_spp"
       "mode_pais_q26_spp"
                                "mode_pais_q27_spp"
                                                        "mode_pais_q28_spp"
##
    [7]
## [10] "mode_pais_q31_spp"
                                "mode_pais_q33_spp"
                                                        "mode_pais_q34_spp"
   [13]
       "mode_pais_q38_spp"
                                "mode_pais_q39_spp"
                                                        "mode_pais_q40_spp"
   [16]
       "mode_pais_q45_spp"
                                "mode_pais_q46_spp"
                                                        "mode_pais_q47_spp"
## [19] "mode_pais_q19e20_spp"
```

```
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_pais_q12_spp 0.018614633
## 2
         mode_pais_q18_spp 0.004551118
## 3
         mode_pais_q19_spp 0.078644949
         mode_pais_q20_spp 0.071754293
## 5
         mode_pais_q22_spp 0.031474338
## 6
         mode_pais_q23_spp 0.176849647
         mode_pais_q26_spp 0.049320062
## 7
## 8
         mode_pais_q27_spp 0.037100386
## 9
         mode_pais_q28_spp 0.054967895
## 10
         mode_pais_q31_spp 0.074907827
## 11
         mode_pais_q33_spp 0.010990986
## 12
         mode_pais_q34_spp 0.051084242
## 13
         mode_pais_q38_spp 0.031791165
         mode_pais_q39_spp 0.026458933
## 14
## 15
         mode_pais_q40_spp 0.191434824
## 16
         mode_pais_q45_spp 0.008219789
## 17
         mode_pais_q46_spp 0.186038043
## 18
         mode_pais_q47_spp 0.108420230
## 19 mode_pais_q19e20_spp 0.060433159
```

## Matriz de correlação (60%)

• q19 x q20: Escolaridade da mãe e do pai: tentar combinar (apenas unir as duas não deu certo)

```
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=6, show.legend = F)</pre>
```

mode_pais_q47_spp																		0.06
mode_pais_q46_spp -																	0.29	0.16
mode_pais_q45_spp -																0.13	0.05	0.71
mode_pais_q40_spp -															0.12	0.38	0.17	0.55
mode_pais_q39_spp -														0.17	0.13	0.11	0.14	0.2
mode_pais_q38_spp -													0.08	0.26	0.07	0.22	0.03	0.13
mode_pais_q34_spp =												0.2	0.11	0.24	0.09	0.28	0.1	0.27
mode_pais_q33_spp -											0.04	0.07	0.12	0.08	0.05	0.06	0.13	0.18
mode_pais_q31_spp -										0.07	0.22	0.13	0.2	0.31	0.12	0.26	0.24	0.17
mode_pais_q28_spp -									0.13	0.14	0.18	0.19	0.16	0.32	0.19	0.19	0.02	0.25
mode_pais_q27_spp -								0.3	0.13	0.13	0.11	0.07	0.22	0.2	0.13	0.1	0.03	0.28
mode_pais_q26_spp -							0.43	0.34	0.08	0.09	0.11	0.1	0.14	0.25	0.14	0.13	0.03	0.28
mode_pais_q23_spp -						0.17	0.2	0.22	0.26	0.09	0.22	0.22	0.22	0.33	0.08	0.39	0.18	0.29
mode_pais_q22_spp -					0.14	0.15	0.17	0.15	0.17	0.08	0.11	0.08	0.14	0.16	0.05	0.12	0.08	0.13
mode_pais_q20_spp -				0.12	0.17	0.31	0.3	0.25	0.17	0.11	0.13	0.13	0.14	0.25	0.1	0.19	0.07	0.64
mode_pais_q19_spp -			0.47	0.11	0.23	0.29	0.3	0.27	0.16	0.13	0.21	0.14	0.13	0.56	0.08	0.19	0.06	0.71
mode_pais_q18_spp -		0.06	0.07	0.06	0.09	0.02	0.04	0.02	0.03	0.17	0.04	0.02	0.14	0.02	0.06	0.01	0.01	0.1
			0.22															
note pais d	note pie de pie														0 388			