

Análise das variáveis Saresp Questionário pais - moda por escola

Série 9EF

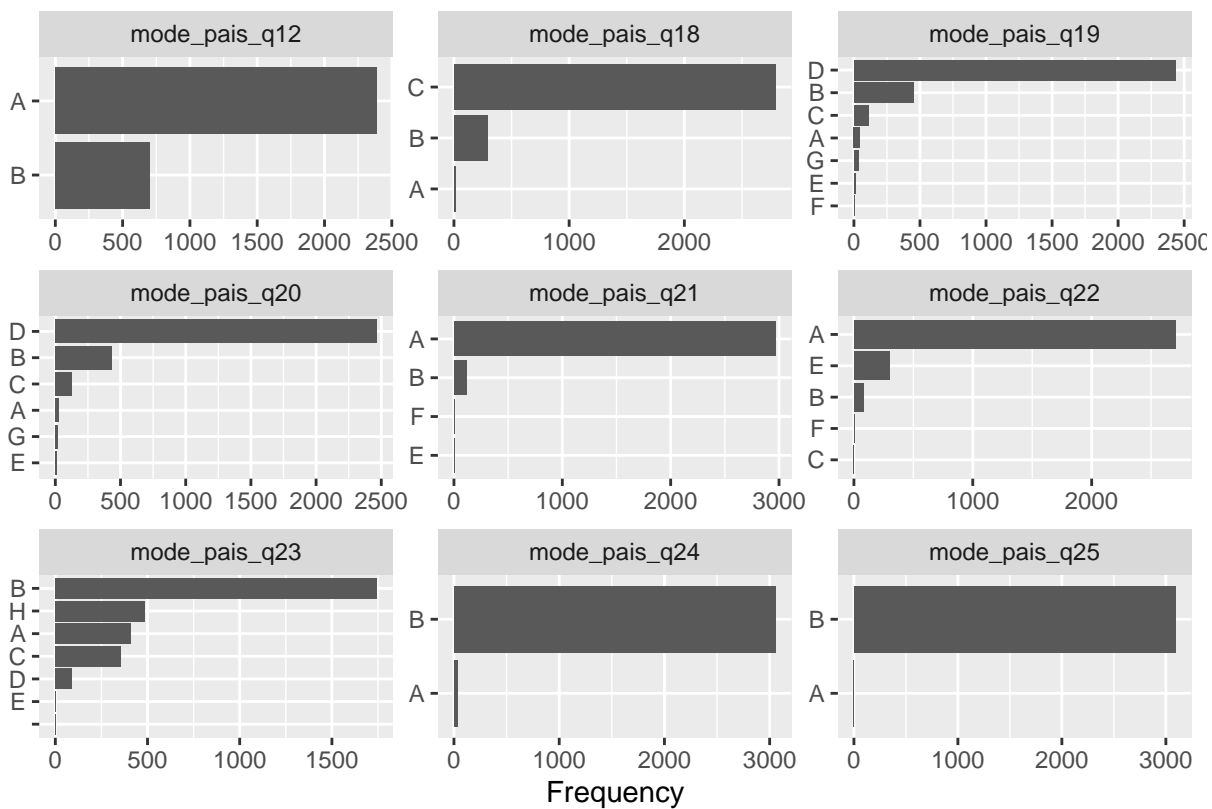
Livia Kobayashi

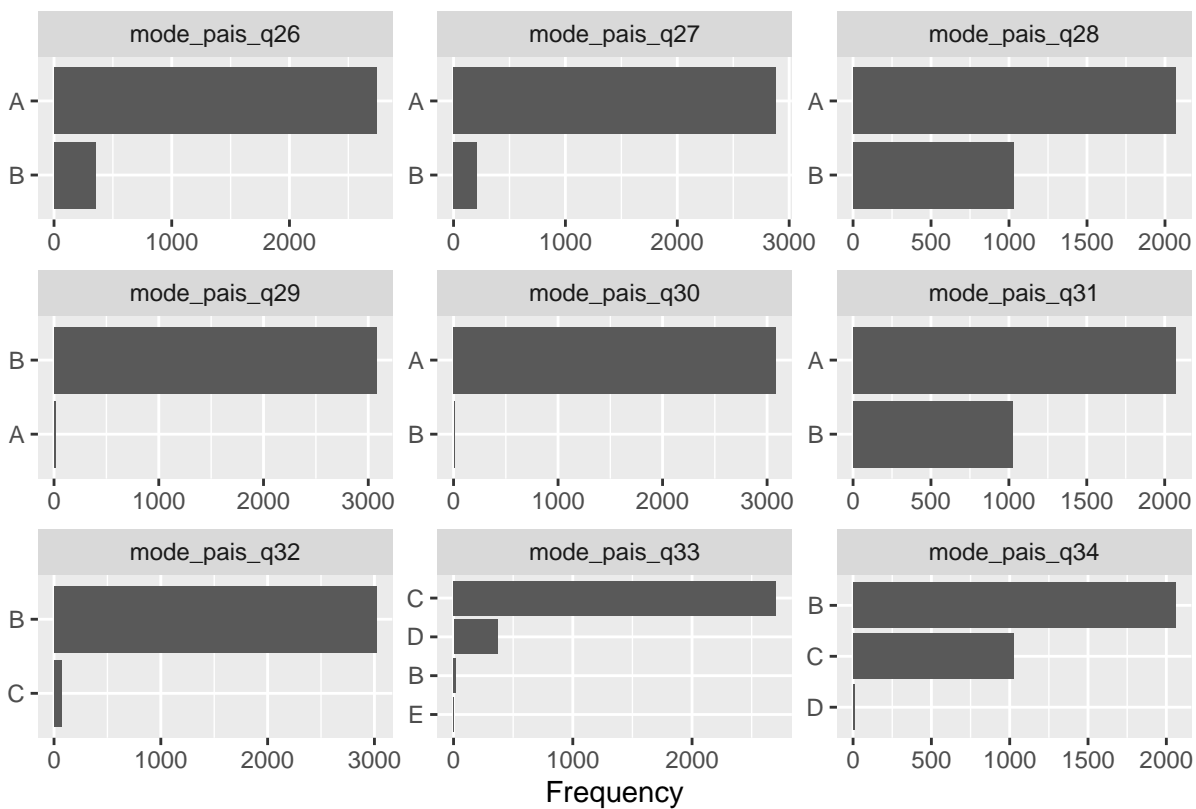
09 junho 2021

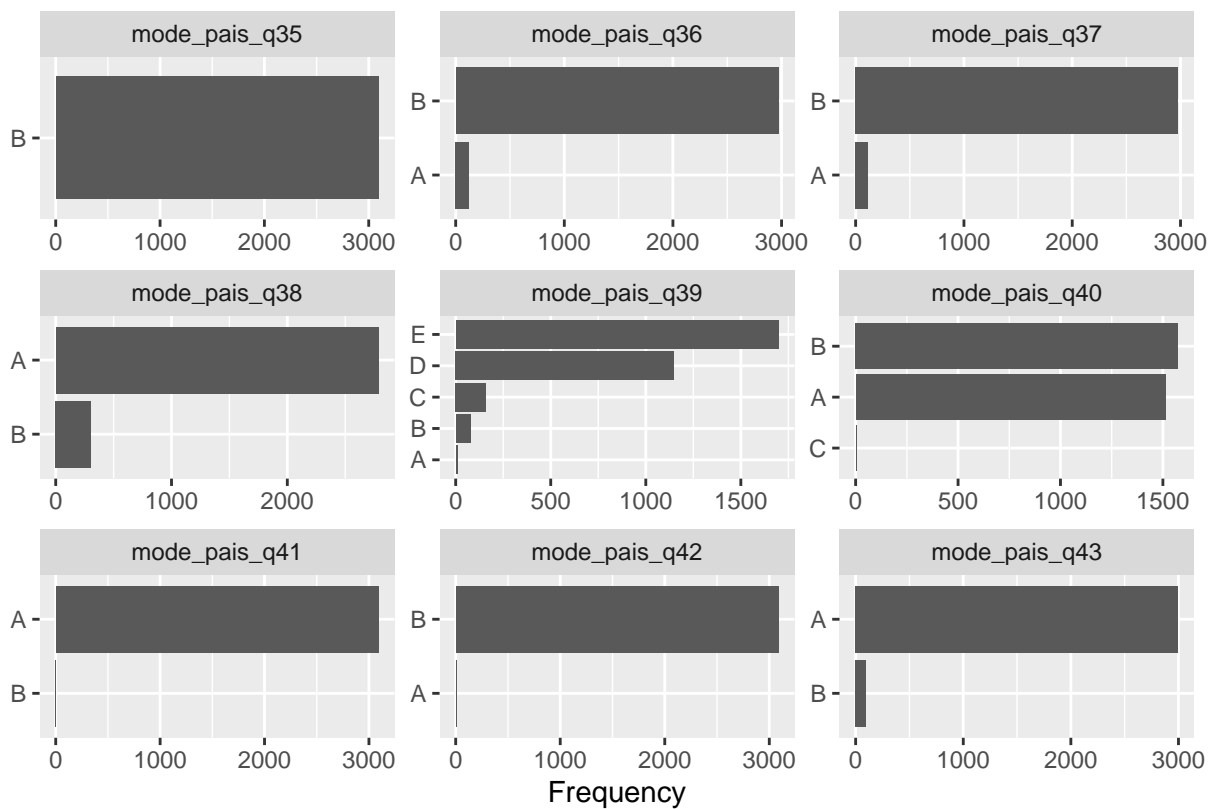
```
library(tidyverse)
library(DataExplorer)
library(gridExtra)
library(grid)
library(caret)
library(ggcorrplot)
library(vcd)
df_publico <- read.csv2("../books/df_publico.csv")
#book <- read.csv2("../books/saresp5ef_pais_mode.csv")
book <- read.csv2(params$book)
```

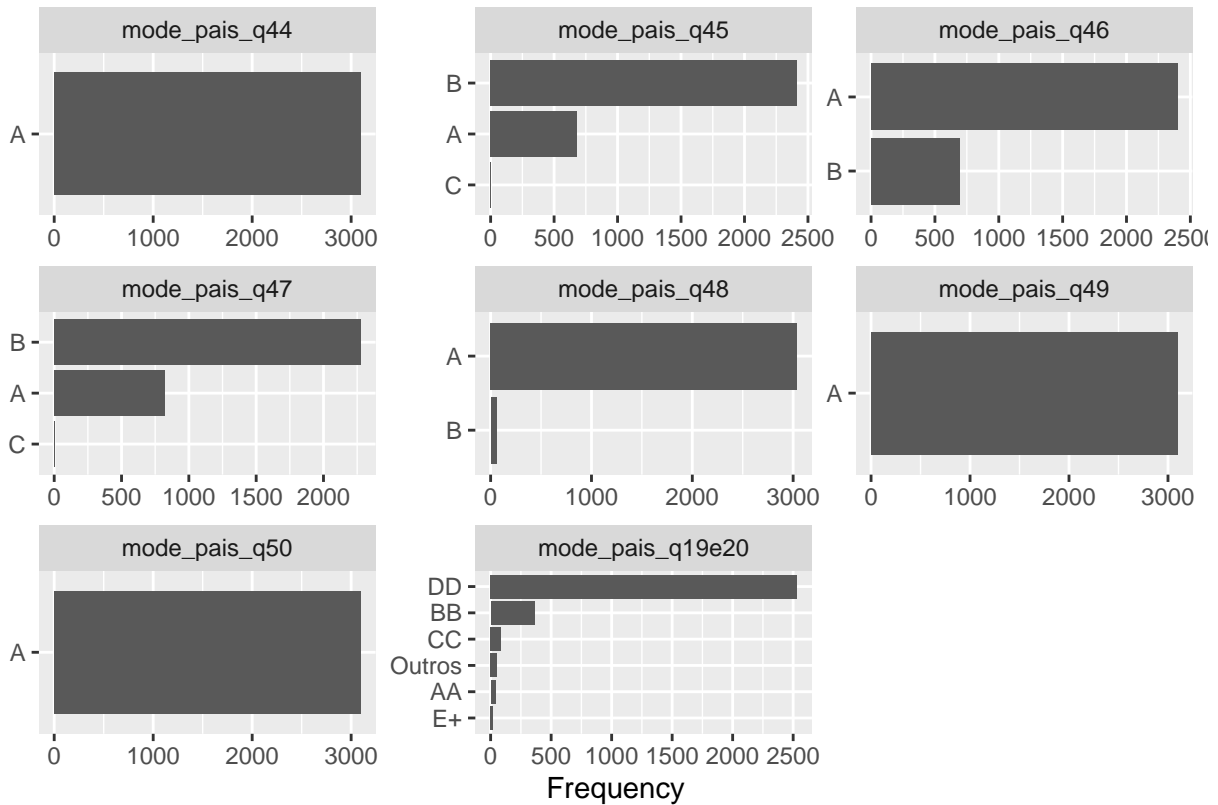
Missing

```
plot_missing(df)
```







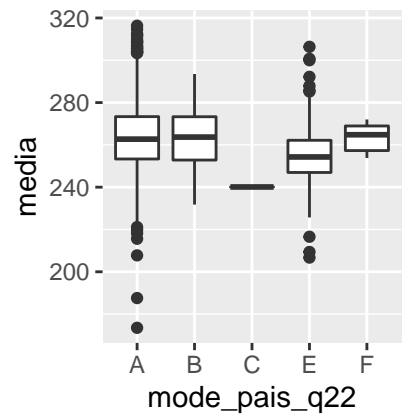
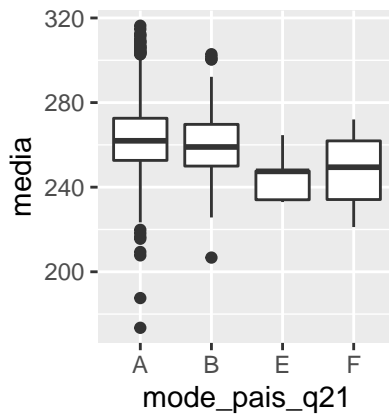
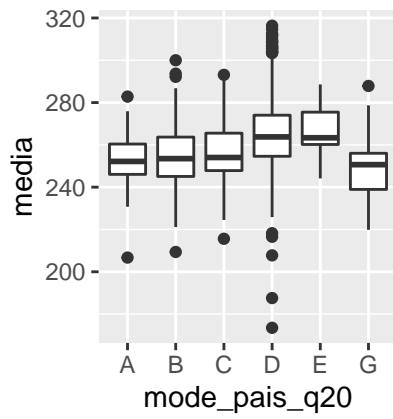
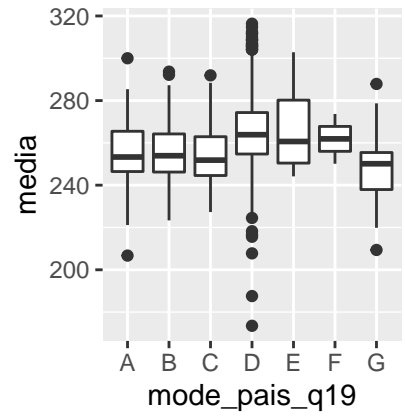
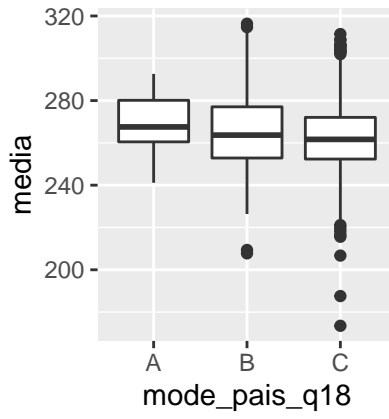
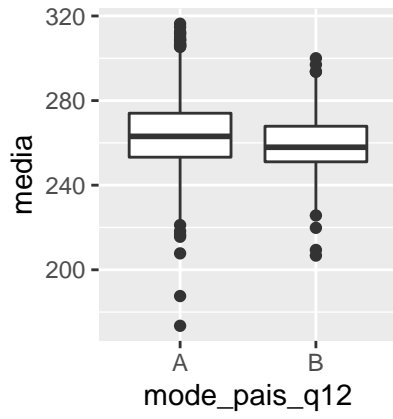
Page 4

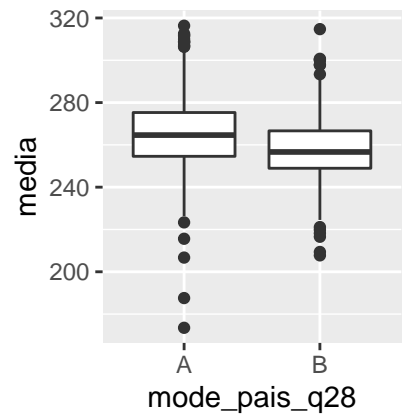
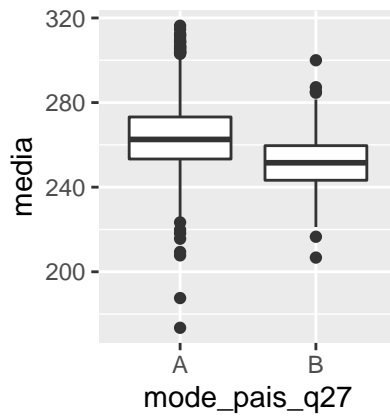
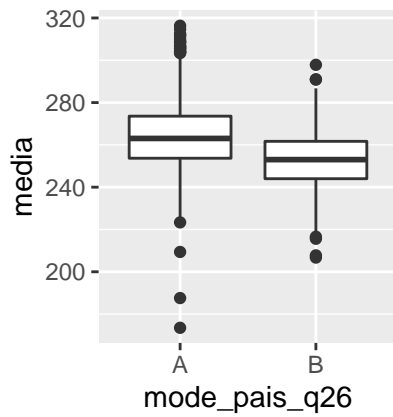
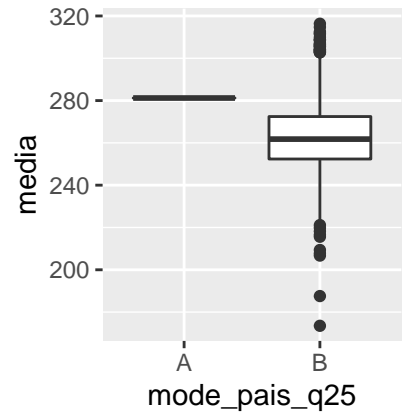
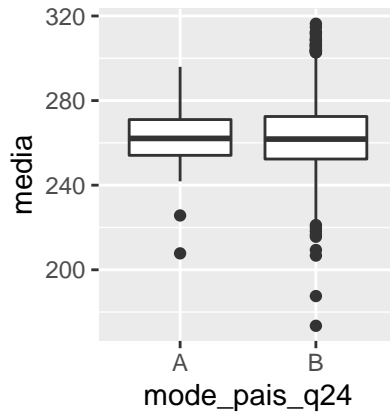
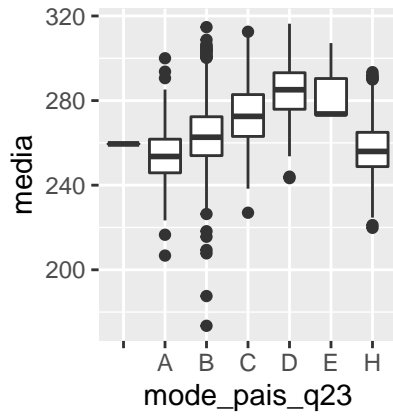
Boxplot

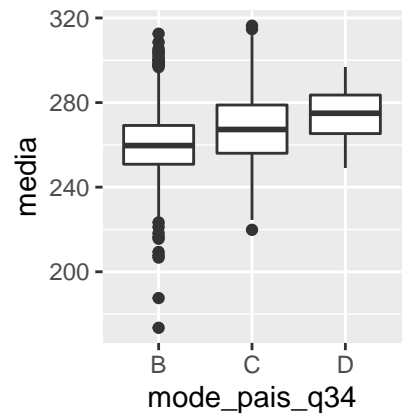
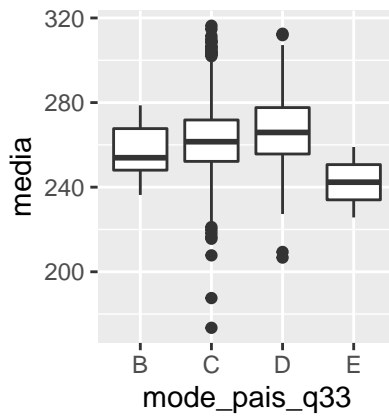
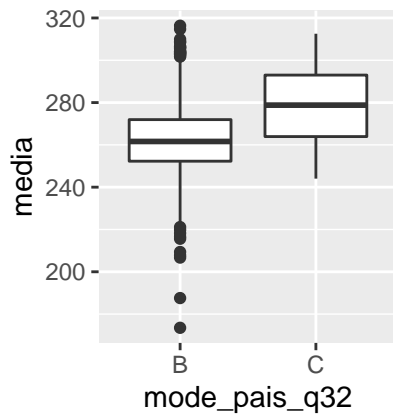
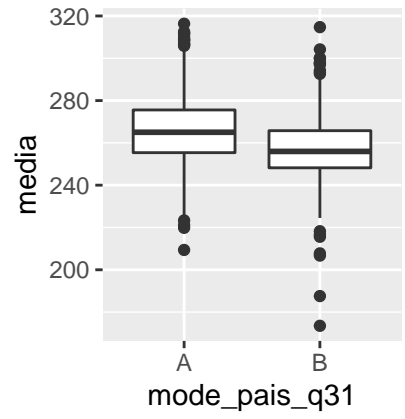
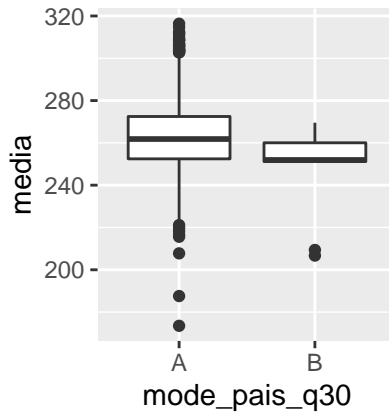
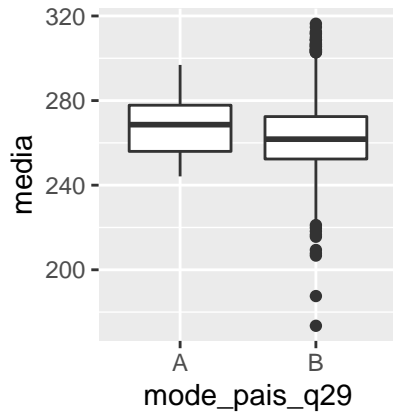
```
vars <- colnames(final_data)
vars <- vars[-c(1,2)]
plots <- list()
i <- 1
for (variable in vars) {
  #plots[[i]] <- plot_boxplot(final_data, by = variable)
  plots[[i]] <- ggplot(final_data, aes_string(variable, "media")) + geom_boxplot()
  i <- i + 1
}

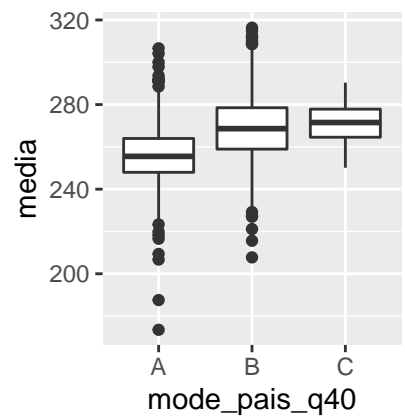
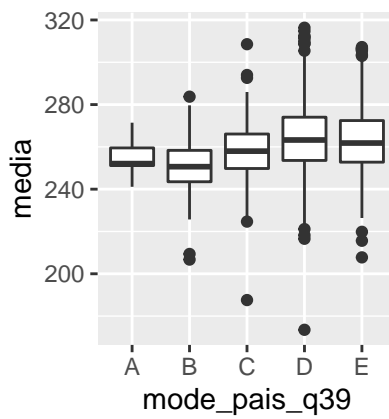
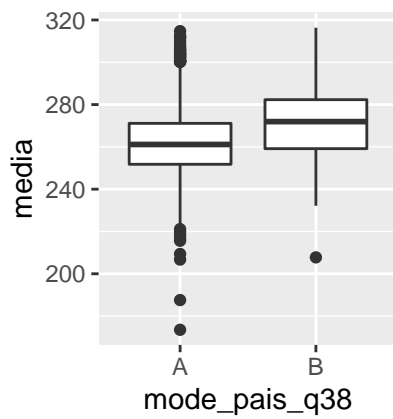
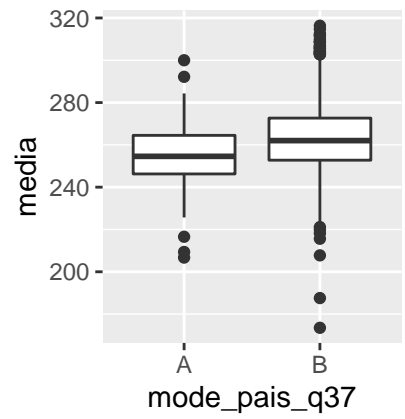
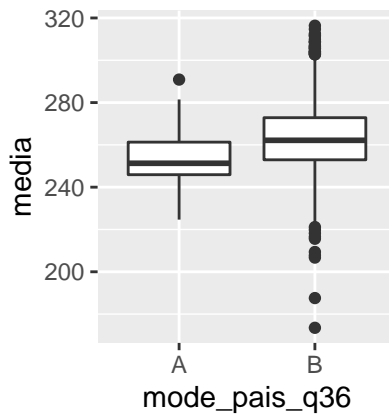
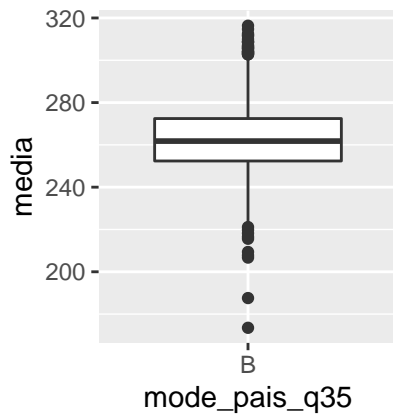
n <- length(plots)

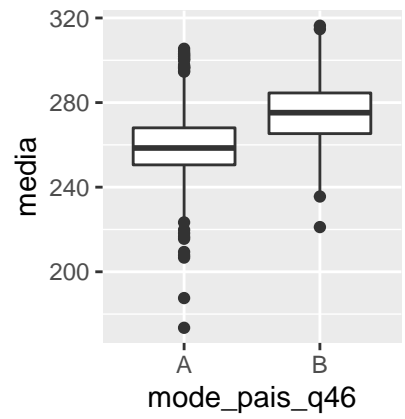
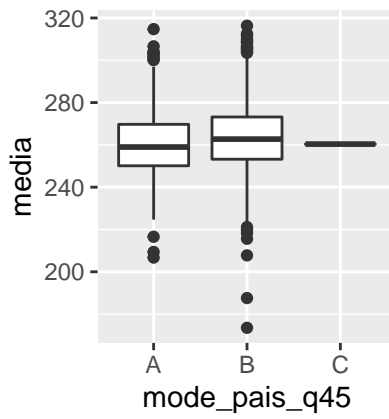
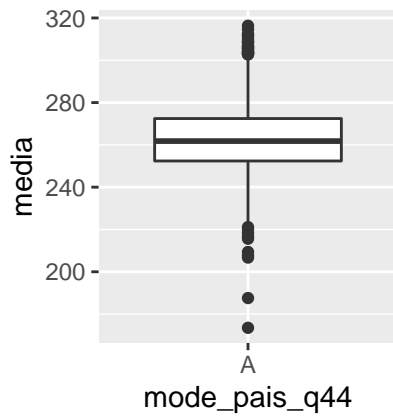
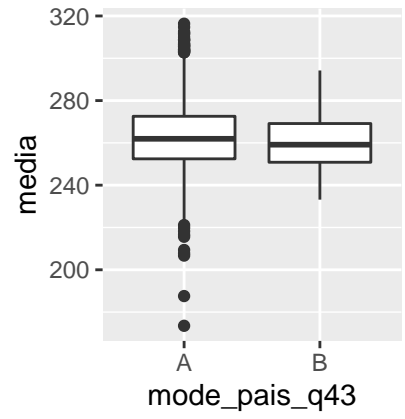
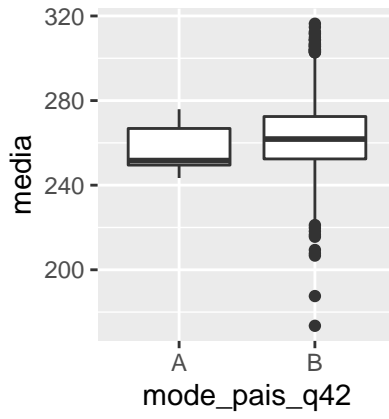
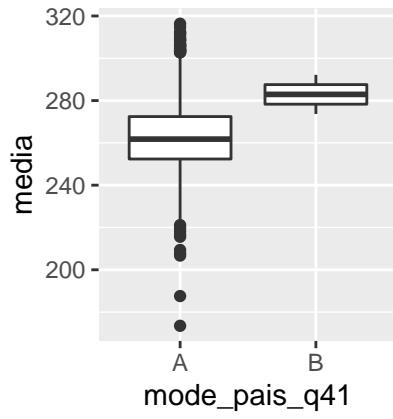
i <- 1
while (i <= n) {
  do.call("grid.arrange", c(plots[i:(min(i+5, n))], ncol=3, nrow = 2))
  i <- i + 6
}
```

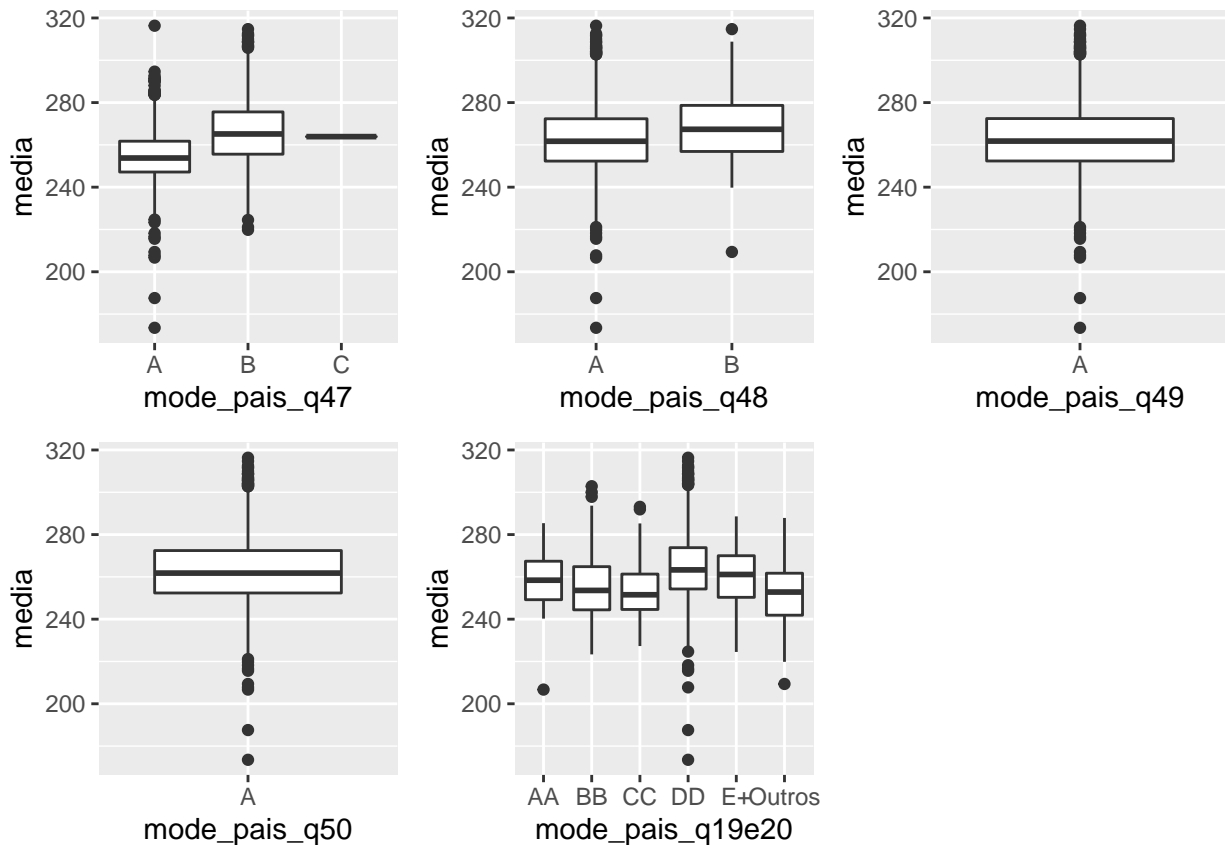












Análise Univariada

```
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_pais_q12" "mode_pais_q18" "mode_pais_q19" "mode_pais_q20"
## [5] "mode_pais_q22" "mode_pais_q23" "mode_pais_q26" "mode_pais_q27"
## [9] "mode_pais_q28" "mode_pais_q31" "mode_pais_q33" "mode_pais_q34"
## [13] "mode_pais_q38" "mode_pais_q39" "mode_pais_q40" "mode_pais_q45"
## [17] "mode_pais_q46" "mode_pais_q47" "mode_pais_q19e20"
```

```
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 0.018614633
## 2  mode_pais_q18 0.004551118
## 3  mode_pais_q19 0.078644949
## 4  mode_pais_q20 0.071754293
## 5  mode_pais_q22 0.031474338
## 6  mode_pais_q23 0.176849647
## 7  mode_pais_q26 0.049320062
## 8  mode_pais_q27 0.037100386
## 9  mode_pais_q28 0.054967895
## 10 mode_pais_q31 0.074907827
## 11 mode_pais_q33 0.010990986
## 12 mode_pais_q34 0.051084242
## 13 mode_pais_q38 0.031791165
## 14 mode_pais_q39 0.026458933
## 15 mode_pais_q40 0.191434824
## 16 mode_pais_q45 0.008219789
## 17 mode_pais_q46 0.186038043
## 18 mode_pais_q47 0.108420230
## 19 mode_pais_q19e20 0.057999523

```

```

catcorrmm <- function(vars, dat) sapply(vars, function(y) sapply(vars, function(x) assocstats(table(dat[,
matriz <- catcorrmm(final_cols, data_corr)

ggcorrplot(matriz, show.diag = F, type="lower", lab=TRUE, lab_size=2, show.legend = F)

```

