

# Análise das variáveis Saeb - moda por escola

## Série 9EF

Livia Kobayashi

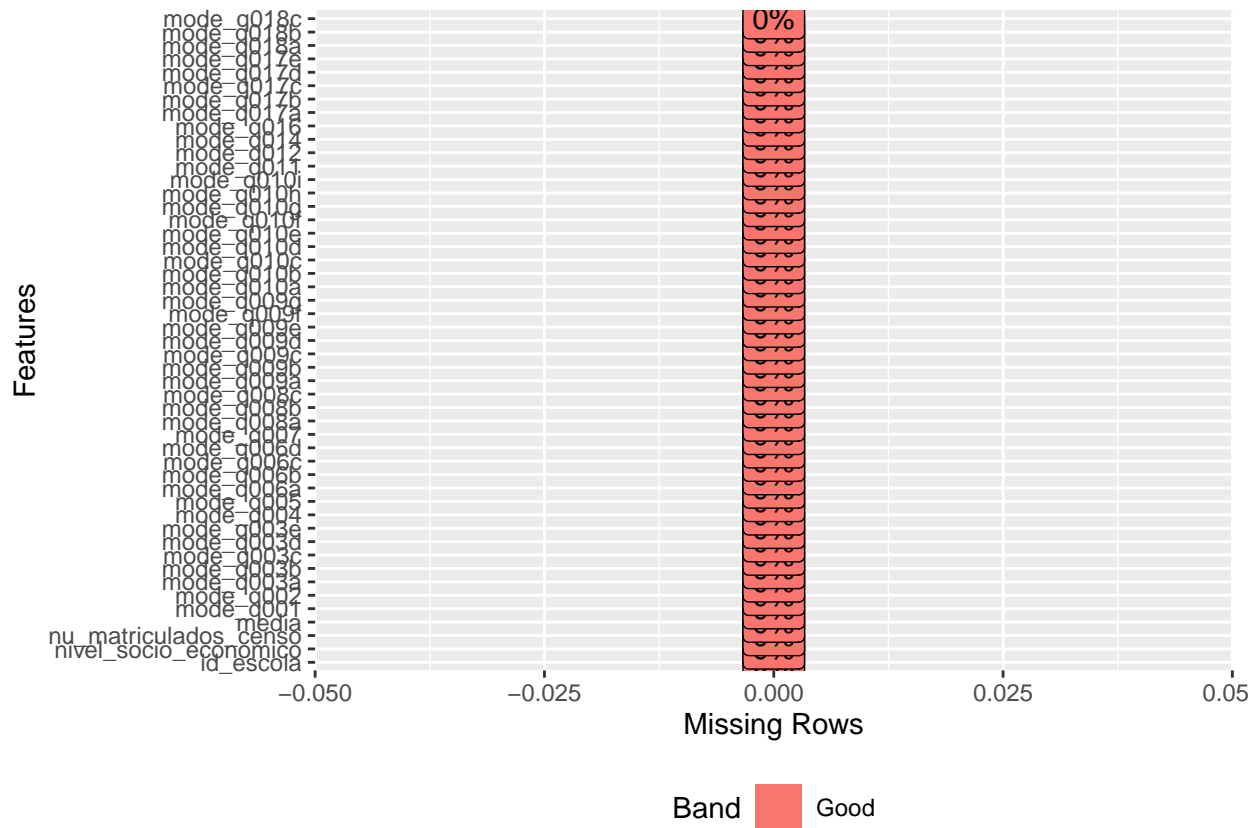
10 junho 2021

```
df_publico <- read.csv2("../books/df_publico.csv")  
  
book <- read.csv2(params$book)
```

```
## id_serie  
## 1      9EF
```

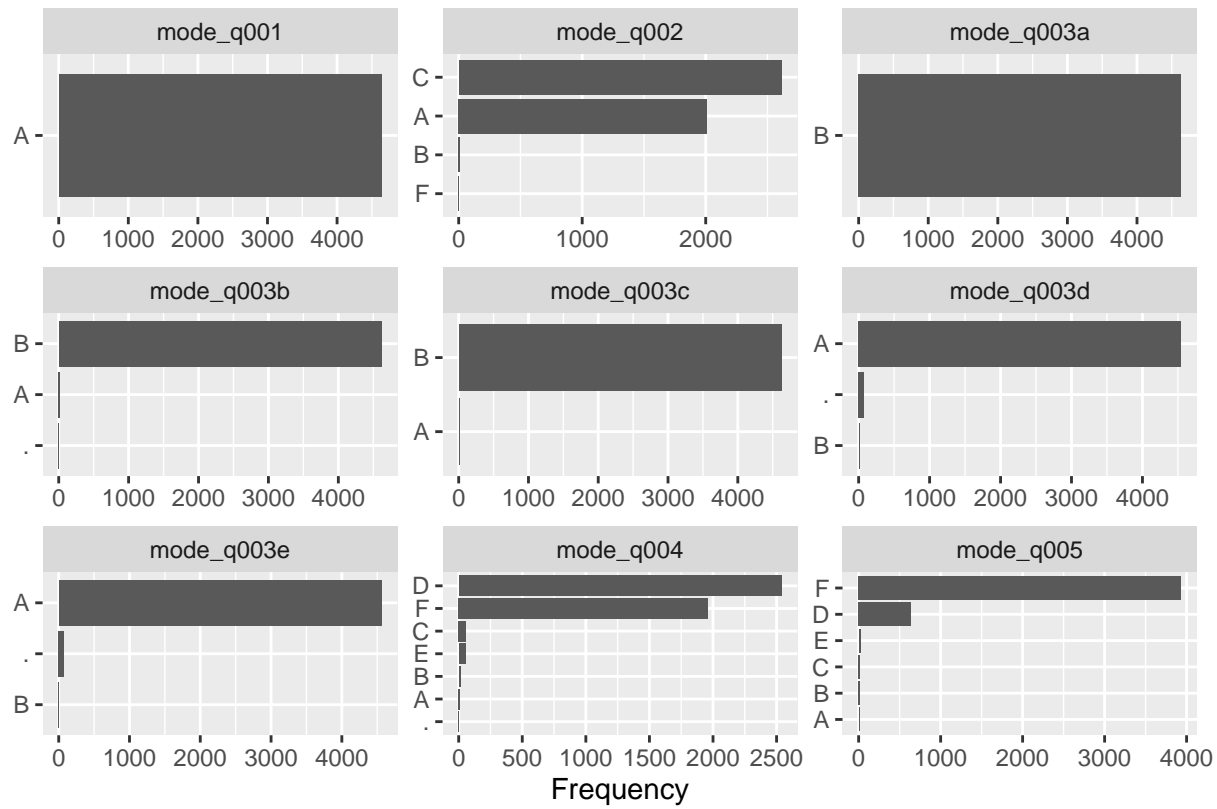
## Missing

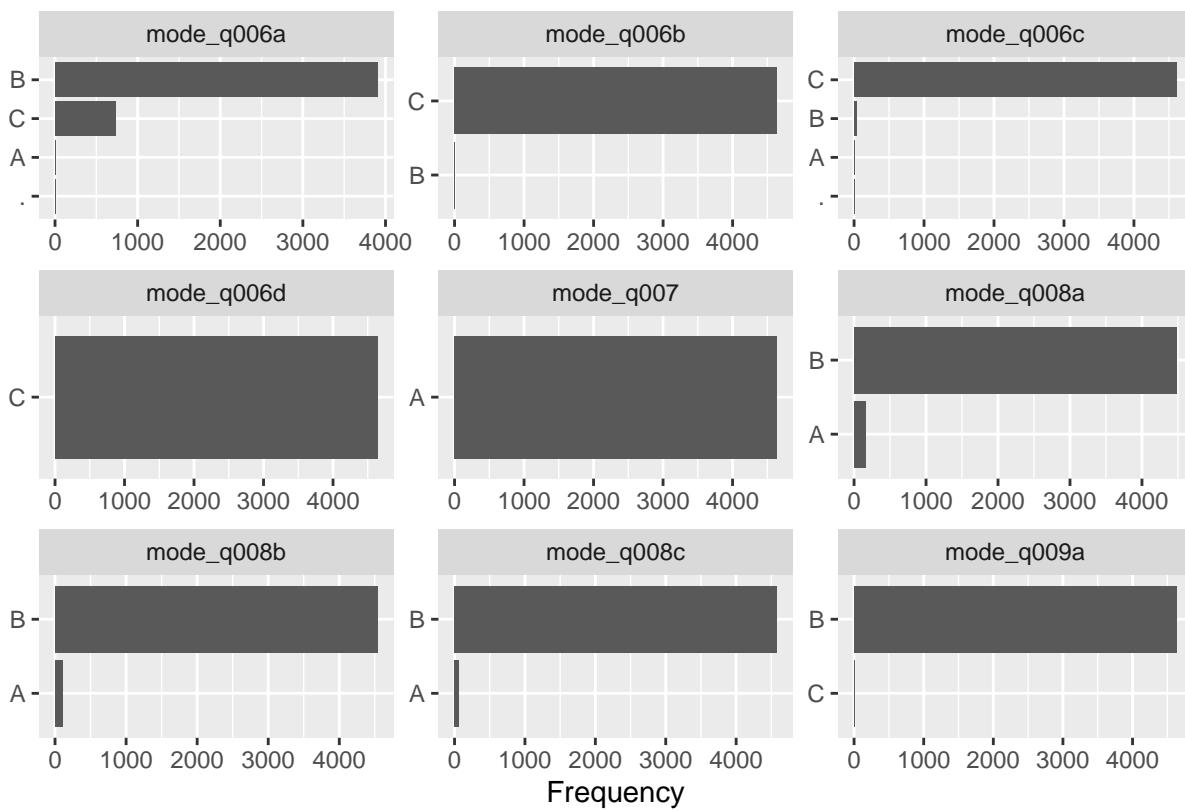
```
plot_missing(df)
```

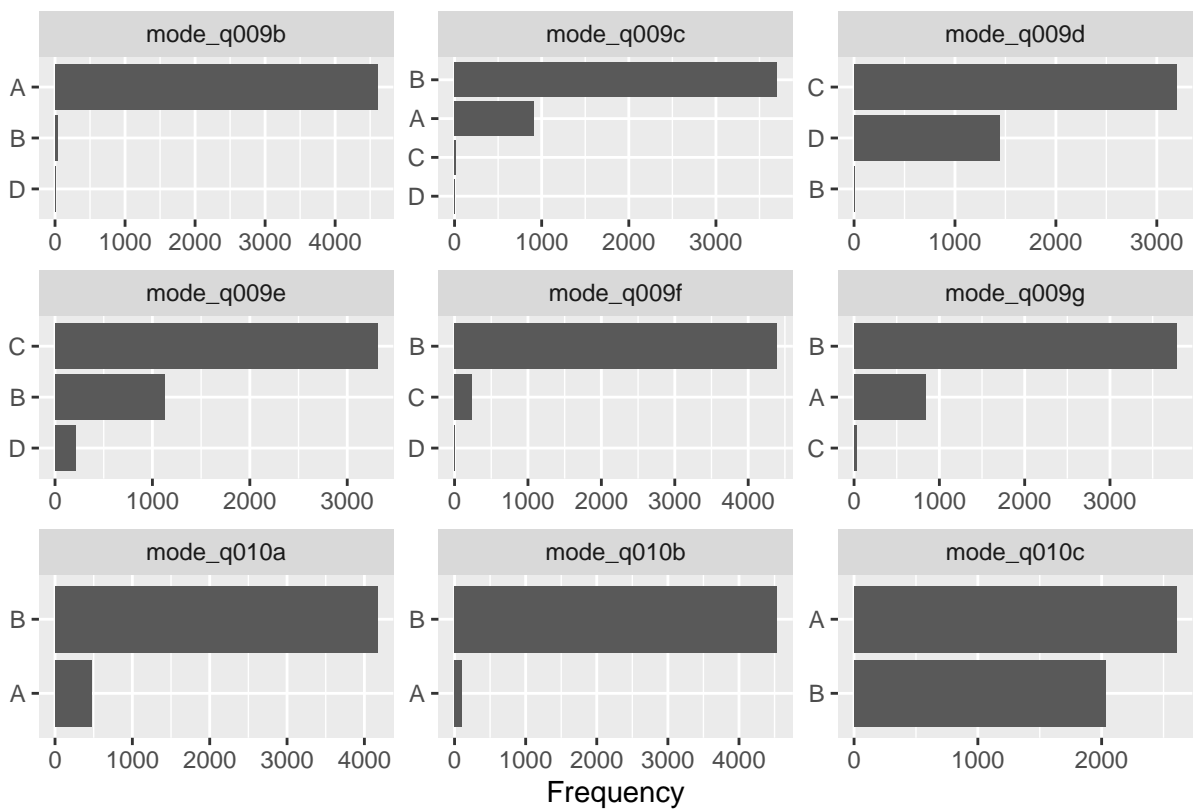


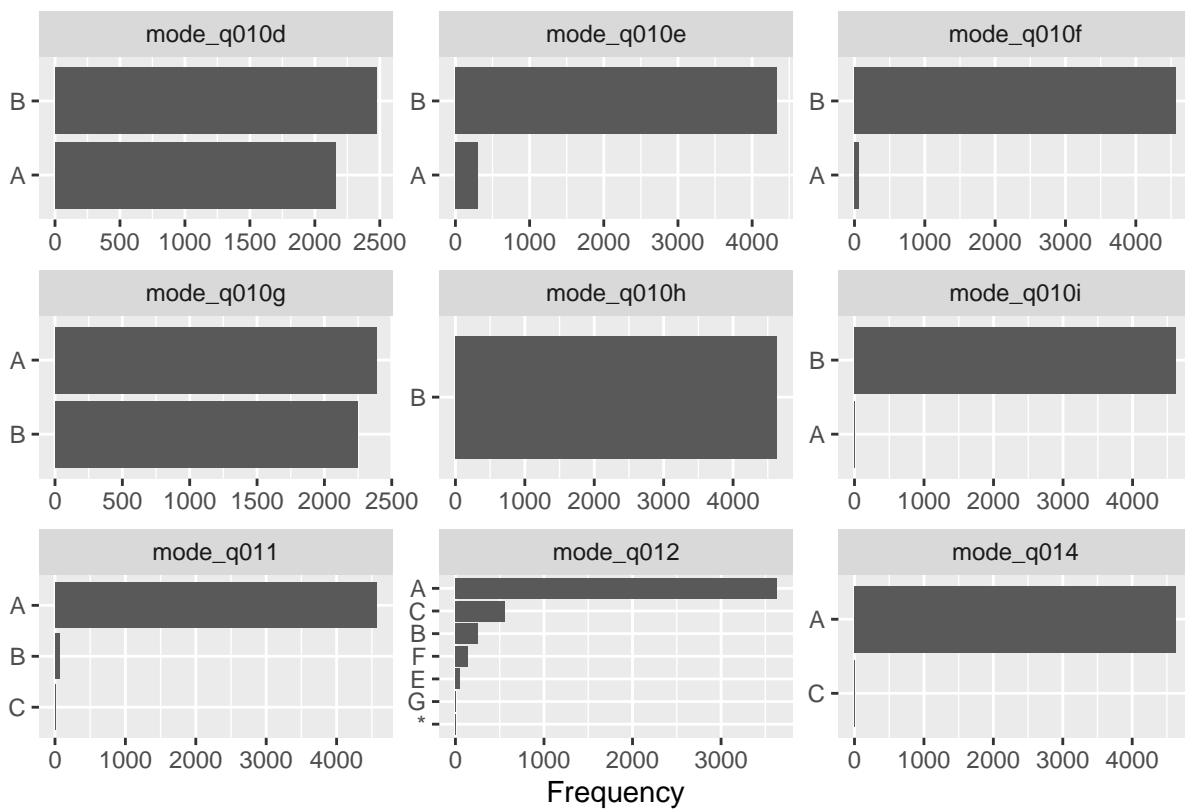
# Volume

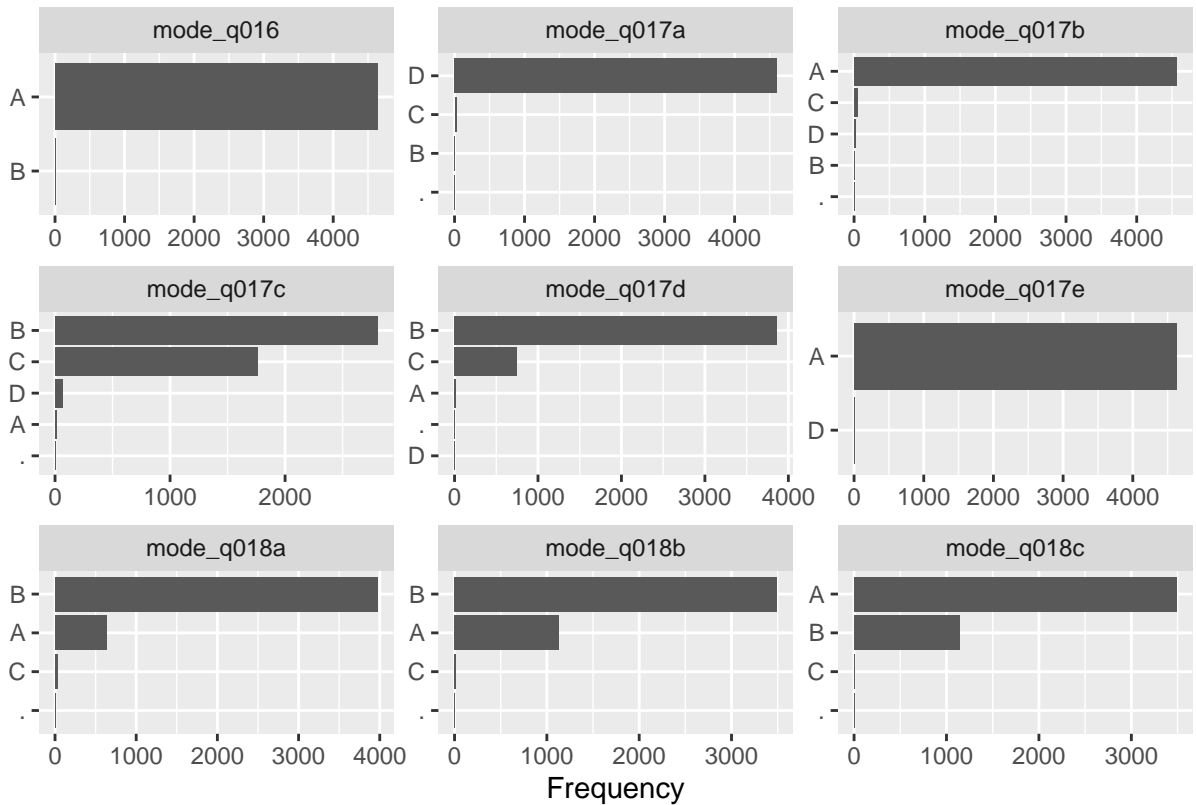
```
plot_bar(final_data)
```











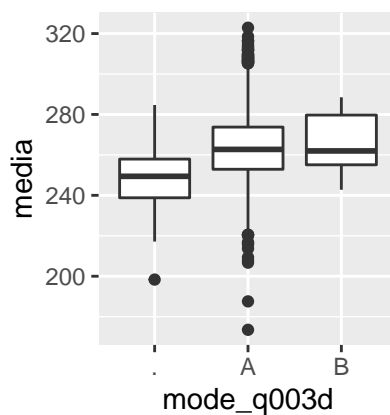
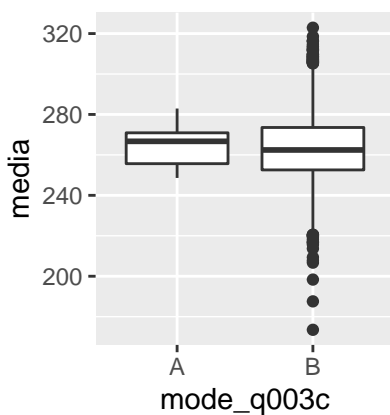
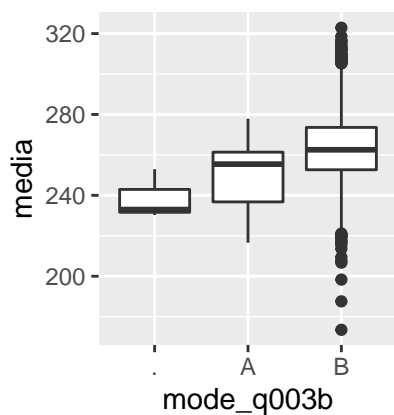
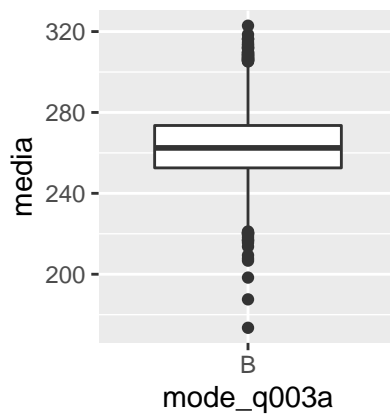
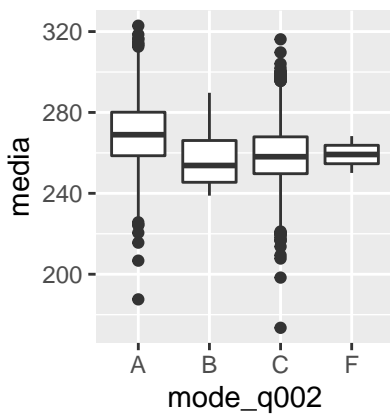
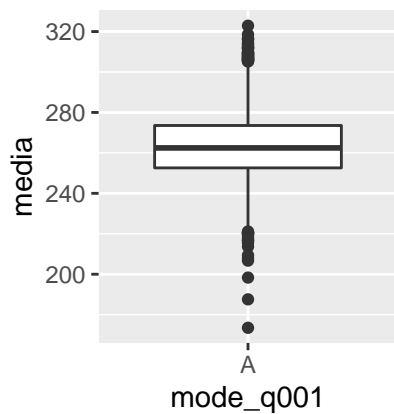
Page 5

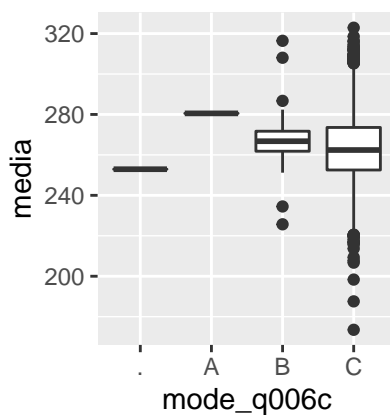
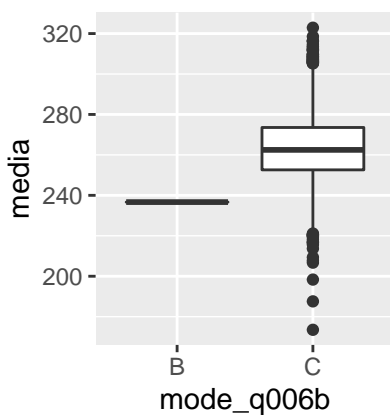
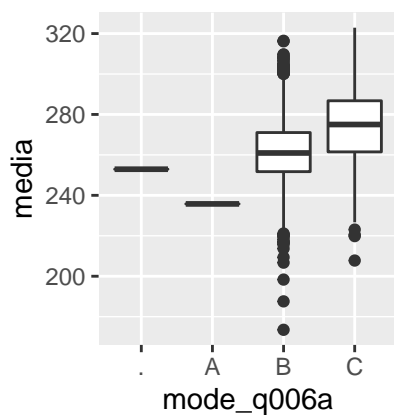
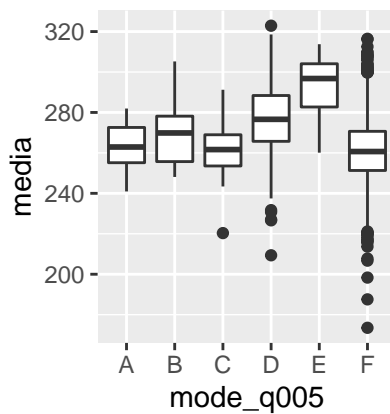
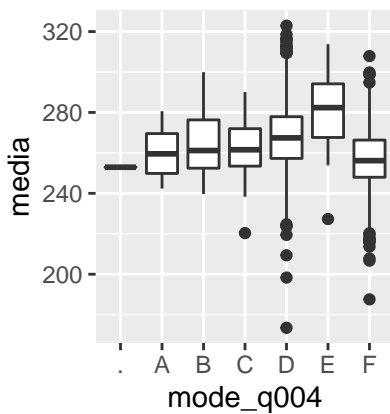
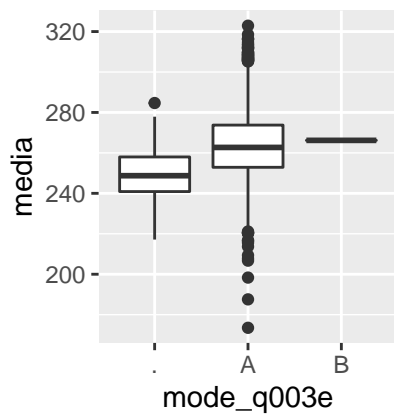
## Boxplot

```
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
}

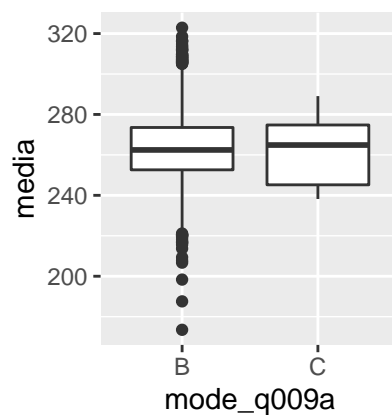
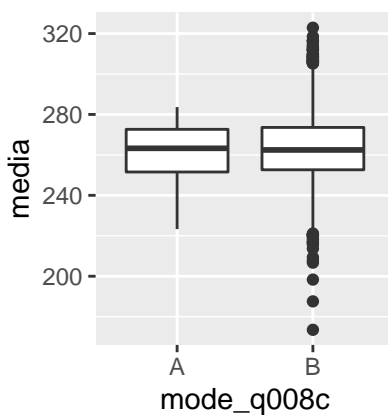
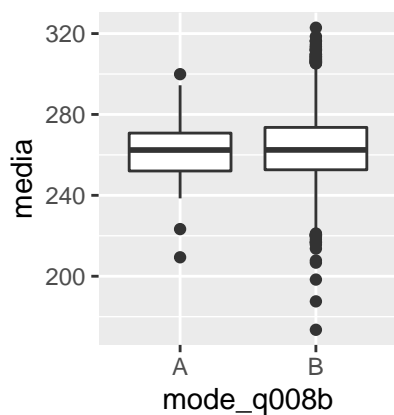
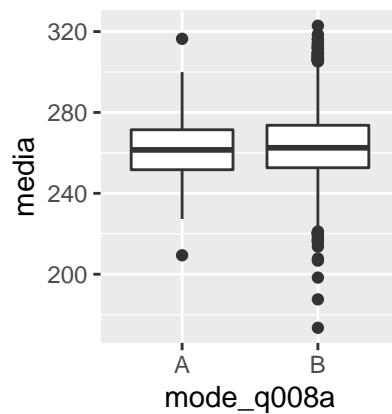
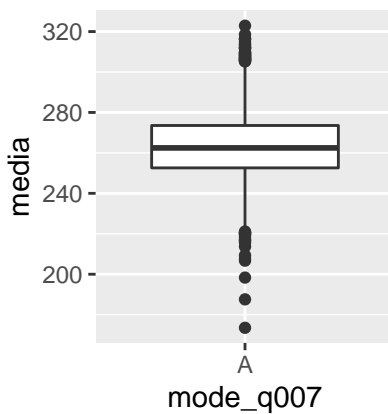
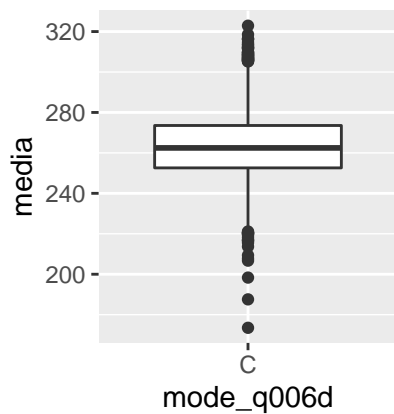
n <- length(plots)

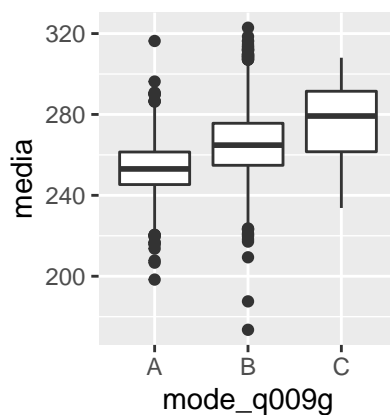
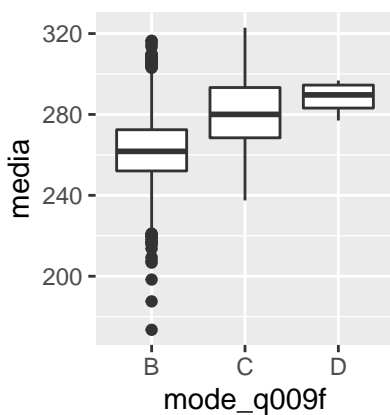
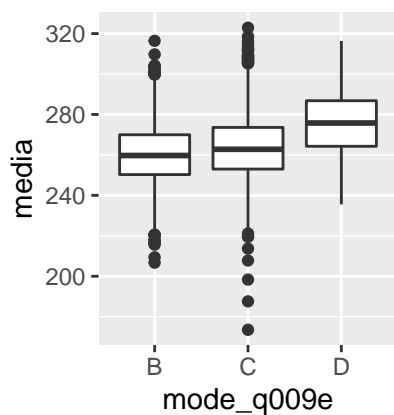
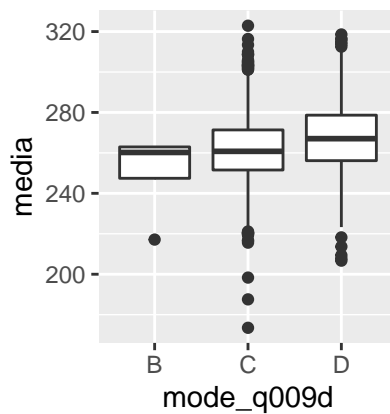
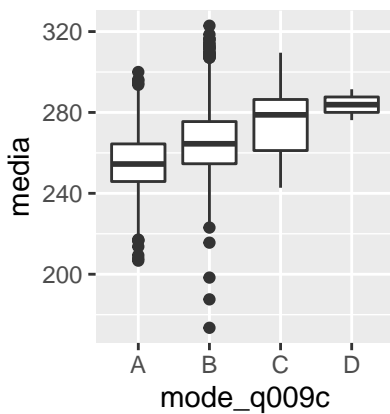
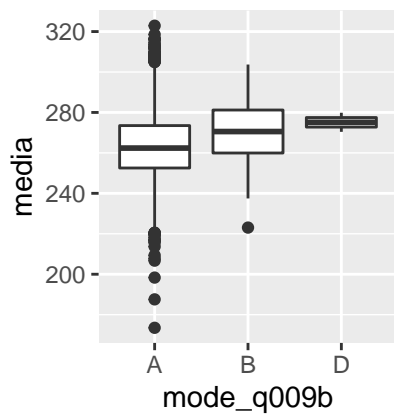
i <- 1
while (i <= n) {
  do.call("grid.arrange", c(plots[i:(min(i+5, n))], ncol=3, nrow = 2))
  i <- i + 6
}
```

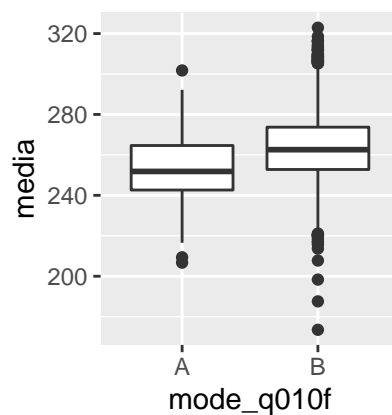
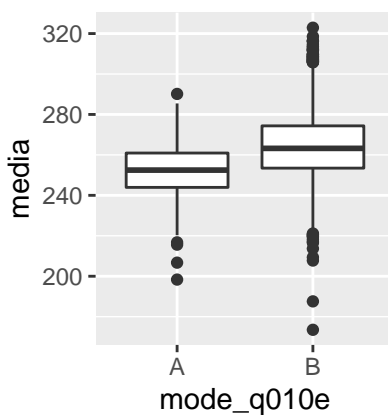
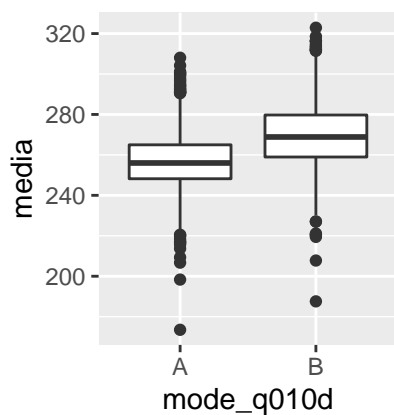
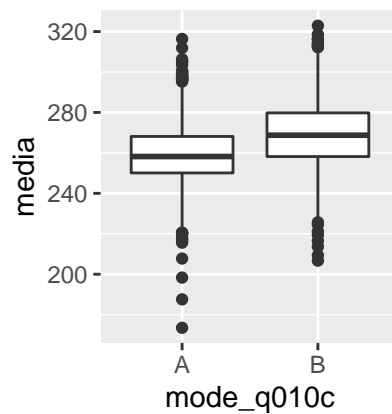
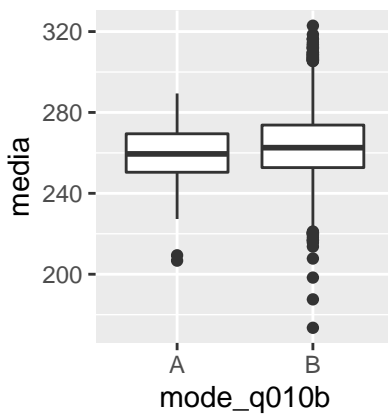
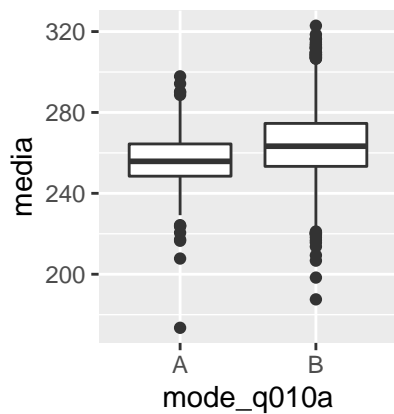


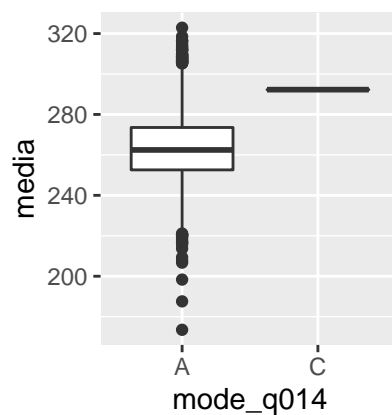
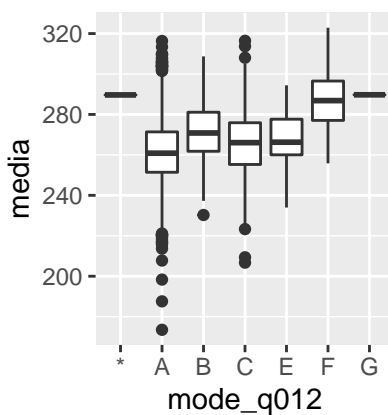
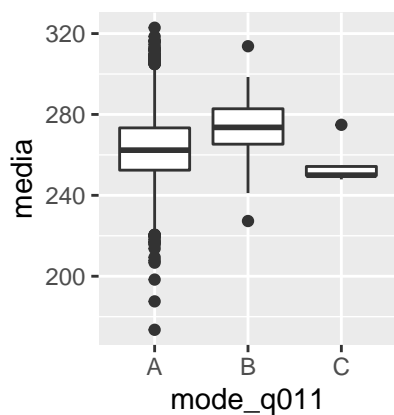
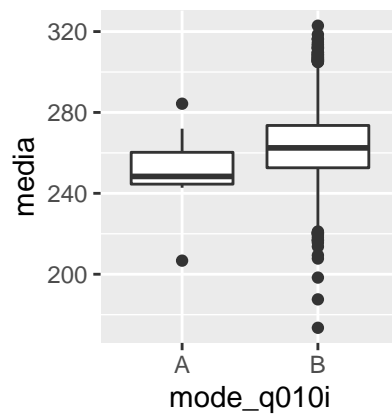
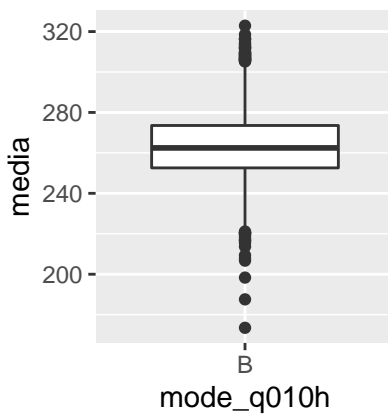
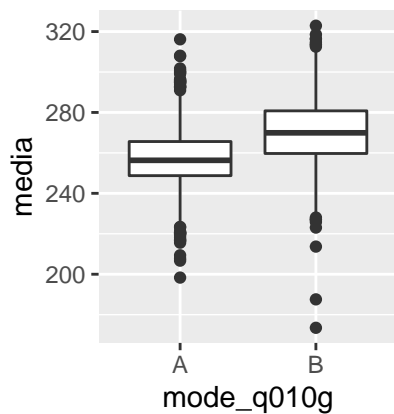


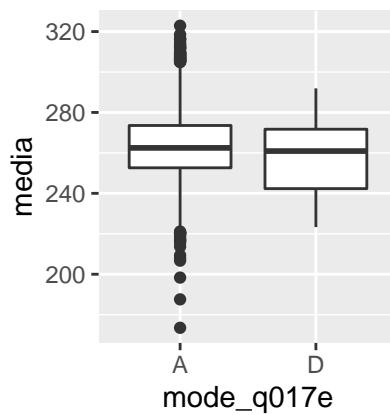
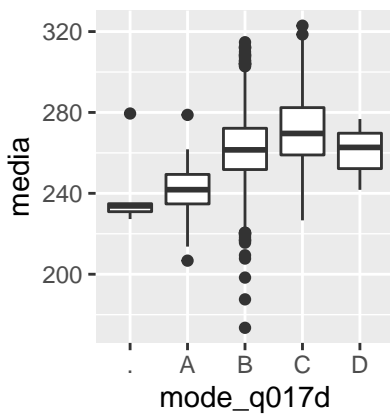
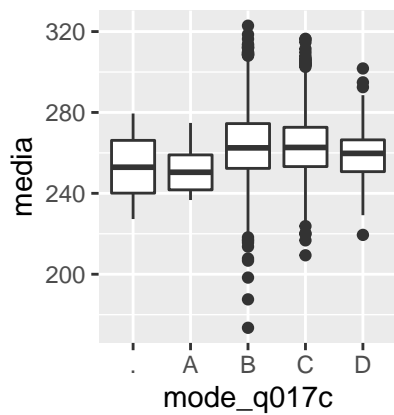
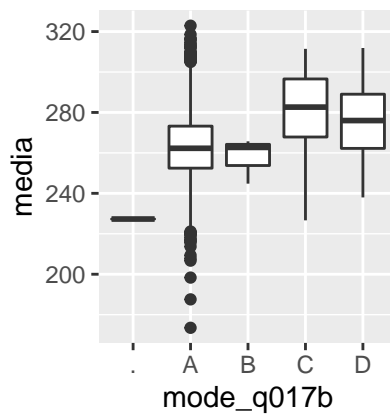
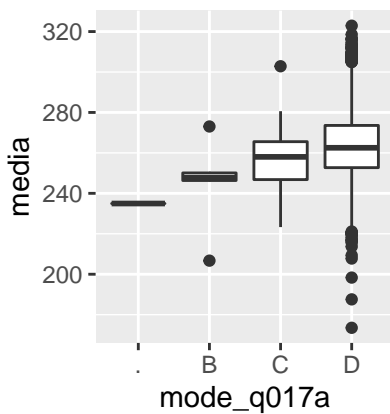
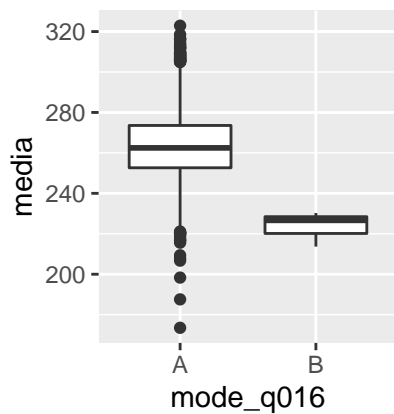


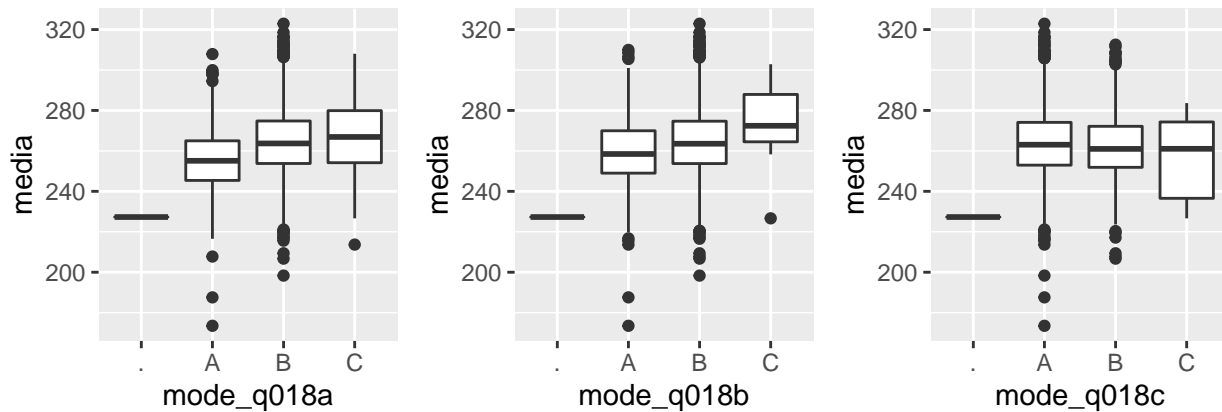












## Análise Univariada

```
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_q002" "mode_q004" "mode_q005" "mode_q006a" "mode_q009c"
## [6] "mode_q009d" "mode_q009e" "mode_q009f" "mode_q009g" "mode_q010a"
## [11] "mode_q010c" "mode_q010d" "mode_q010e" "mode_q010g" "mode_q012"
## [16] "mode_q017c" "mode_q017d" "mode_q018a" "mode_q018b" "mode_q018c"
```

```
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_q002 0.114479001  
## 2 mode_q004 0.128723265  
## 3 mode_q005 0.132719373  
## 4 mode_q006a 0.080089568  
## 5 mode_q009c 0.072262439  
## 6 mode_q009d 0.033173487  
## 7 mode_q009e 0.035991687  
## 8 mode_q009f 0.064590618  
## 9 mode_q009g 0.088286722  
## 10 mode_q010a 0.022290217  
## 11 mode_q010c 0.097323724  
## 12 mode_q010d 0.166898626  
## 13 mode_q010e 0.035068854  
## 14 mode_q010g 0.168636755  
## 15 mode_q012 0.096023586  
## 16 mode_q017c 0.002206833  
## 17 mode_q017d 0.055479615  
## 18 mode_q018a 0.036212310  
## 19 mode_q018b 0.021464346  
## 20 mode_q018c 0.003196556
```

```
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=6, show.legend = F)
```

