

R Notebook

```
# install.packages("ggplot2")
# install.packages("data.table")
# install.packages("stargazer")
# install.packages("skimr")

library(ggplot2)
library(data.table)
library(stargazer)

##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(skimr)

##
## Attaching package: 'skimr'
## The following object is masked from 'package:stats':
##
## filter

library(readxl)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:data.table':
##
## between, first, last
## The following objects are masked from 'package:stats':
##
## filter, lag
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

dtAmostra <- fread("dadosBanco/Amostra_Modelo_Evasao_Correntistas_v3.csv")

dtAmostra$Evadiu <- as.factor(dtAmostra$Evadiu)
dtAmostra$Debito_Automatico <- as.factor(dtAmostra$Debito_Automatico)
dtAmostra$Credito_Salario <- as.factor(dtAmostra$Credito_Salario)
dtAmostra$Credenciamento <- as.factor(dtAmostra$Credenciamento)
dtAmostra$Caixa_Seguradora <- as.factor(dtAmostra$Caixa_Seguradora)
dtAmostra$Pediou_Portabilidade <- as.factor(dtAmostra$Pediou_Portabilidade)
dtAmostra$Abriu_Reclamacao <- as.factor(dtAmostra$Abriu_Reclamacao)
dtAmostra$Debito_Automatico_DIF <- as.factor(dtAmostra$Debito_Automatico_DIF)
dtAmostra$Credito_Salario_DIF <- as.factor(dtAmostra$Credito_Salario_DIF)
dtAmostra$Caixa_Seguradora_DIF <- as.factor(dtAmostra$Caixa_Seguradora_DIF)
```

```
str(dtAmostra)
```

```
## Classes 'data.table' and 'data.frame': 100000 obs. of 36 variables:
## $ Id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Evadiu : Factor w/ 2 levels "0","1": 1 2 2 1 2 1 1 1 2 2 ...
## $ Segmento : chr "GR" "GC" "GC" "GV" ...
## $ Debito_Automatico : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 1 1 ...
## $ Credito_Salario : Factor w/ 2 levels "0","1": 2 2 1 1 1 1 2 1 2 1 ...
## $ Credenciamento : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Caixa_Seguradora : Factor w/ 2 levels "0","1": 2 1 1 1 1 2 2 2 1 1 ...
## $ Pediu_Portabilidade : Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
## $ Abriu_Reclamacao : Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
## $ Debito_Automatico_DIF : Factor w/ 3 levels "-1","0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ Credito_Salario_DIF : Factor w/ 3 levels "-1","0","1": 3 3 2 2 2 2 2 2 2 2 ...
## $ Caixa_Seguradora_DIF : Factor w/ 3 levels "-1","0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ Produtos_Qualificados : int 6 0 5 4 5 6 11 8 3 3 ...
## $ Produtos_Qualificados_Anterior : int 8 1 6 3 5 5 8 10 3 3 ...
## $ Produtos_Qualificados_DIF : int -2 -1 -1 1 0 1 3 -2 0 0 ...
## $ Produtos_Qualificados_PERC : num -25 -100 -16.7 33.3 0 ...
## $ Produtos : int 6 6 5 4 6 8 14 9 5 3 ...
## $ Produtos_Anterior : int 9 1 6 3 6 7 11 11 4 3 ...
## $ Produtos_DIF : int -3 5 -1 1 0 1 3 -2 1 0 ...
## $ Produtos_PERC : num -33.3 500 -16.7 33.3 0 ...
## $ Movimentacoes : num 19.833 0.5 0.833 5.5 0.667 ...
## $ Movimentacoes_Anterior : num 1 0 1.17 6.5 3.5 ...
## $ Movimentacoes_DIF : num 18.833 0.5 -0.333 -1 -2.833 ...
## $ Movimentacoes_PERC : num 1883.3 0 -28.6 -15.4 -81 ...
## $ Aplicacao : num 13818.1 412.6 4268.3 10.1 85.6 ...
## $ Aplicacao_Anterior : num 47067.83 7.04 1353.43 0 689.75 ...
## $ Aplicacao_DIF : num -33249.7 405.6 2914.8 10.1 -604.1 ...
## $ Aplicacao_PERC : num -70.6 5760.7 215.4 0 -87.6 ...
## $ Credito : num 0 0 0 54380 37511 ...
## $ Credito_Anterior : num 261 0 173110 56118 37511 ...
## $ Credito_DIF : num -261 0 -173110 -1738 0 ...
## $ Credito_PERC : num -100 0 -100 -3.1 0 ...
## $ Rentabilidade : num 1010 1 5615 1664 650 ...
## $ Rentabilidade_Anterior : num 1009.29 2.88 5300.06 1233.58 2178.44 ...
## $ Rentabilidade_DIF : num 0.96 -1.88 315.24 430.46 -1528.05 ...
## $ Rentabilidade_PERC : num 0.0951 -65.2778 5.9479 34.8952 -70.1442 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
### A variável "Pediu_Portabilidade" possui somente o valor igual a 0
```

```
dtAmostra =
  dtAmostra %>%
  mutate(Evadiu_formatado = ifelse(Evadiu == 1, "Sim", "Não")) %>%
  mutate(winsored_Rentabilidade_PERC = psych::winsor(Rentabilidade_PERC, 0.1)) %>%
  mutate(winsored_Produtos = psych::winsor(Produtos, 0.01)) %>%
  mutate(winsored_Produtos_Qualificados_PERC = psych::winsor(Produtos_Qualificados_PERC, 0.01)) %>%
  mutate(winsored_Produtos_PERC = psych::winsor(Produtos_PERC, 0.01)) %>%
  mutate(winsored_Movimentacoes = psych::winsor(Movimentacoes, 0.1)) %>%
  mutate(winsored_Movimentacoes_PERC = psych::winsor(Movimentacoes_PERC, 0.1)) %>%
  mutate(winsored_Aplicacao = psych::winsor(Aplicacao, 0.1)) %>%
  mutate(winsored_Aplicacao_Anterior = psych::winsor(Aplicacao_Anterior, 0.1)) %>%
```

```

mutate(winsored_Aplicacao_DIF = psych::winsor(Aplicacao_DIF, 0.1)) %>%
mutate(winsored_Aplicacao_PERC = psych::winsor(Aplicacao_PERC, 0.1)) %>%
mutate(winsored_Credito = psych::winsor(Credito, 0.1)) %>%
mutate(winsored_Credito_Anterior = psych::winsor(Credito_Anterior, 0.1)) %>%
mutate(winsored_Credito_DIF = psych::winsor(Credito_DIF, 0.1)) %>%
mutate(winsored_Credito_PERC = psych::winsor(Credito_PERC, 0.1)) %>%
mutate(winsored_Rentabilidade = psych::winsor(Rentabilidade, 0.1)) %>%
mutate(winsored_Rentabilidade_Anterior = psych::winsor(Rentabilidade_Anterior, 0.1)) %>%
mutate(winsored_Rentabilidade_DIF = psych::winsor(Rentabilidade_DIF, 0.1)) %>%
mutate(winsored_Rentabilidade_PERC = psych::winsor(Rentabilidade_PERC, 0.1))

# ggplot(data = dtAmostra, aes(Evadiu_formatado, winsored_Produtos)) +
#   # geom_hline(yintercept = 0, linetype = "dashed", alpha = 0.2) +
#   geom_boxplot() +
#   # aes(fill = Segmento)
#   labs(title = "Box Plot",
#         subtitle = "",
#         x = "",
#         y = "") +
#   theme_bw() +
#   facet_wrap(~ Segmento)

variaveisSubgrupo <- c("Segmento", "Credenciamento", "Caixa_Seguradora", "Caixa_Seguradora_DIF", "Credi

variaveis1 <- c("Abriu_Reclamacao",
               "Produtos_Qualificados",
               "Produtos_Qualificados_Anterior",
               "Produtos_Qualificados_DIF",

               ## USAR WINSOR
               "Produtos_Qualificados_PERC",

               #NÃO TROUXE RESULTADOS BONS
               "Produtos",

               ##NÃO TROUXE RESULTADOS BONS
               "Produtos_Anterior",
               "Produtos_DIF",

               ## USAR WINSORED
               "Produtos_PERC",
               "Movimentacoes",
               "Movimentacoes_Anterior",

               ##NÃO TROUXE RESULTADOS BONS
               "Movimentacoes_DIF",

               ## USAR WINSOR
               "Movimentacoes_PERC",

               ## USAR WINSOR
               "Aplicacao",

```

```

## USAR WINSOR
"Aplicacao_Anterior",

## USAR WINSOR
"Aplicacao_DIF",

## USAR WINSOR
"Aplicacao_PERC",

## USAR WINSOR
"Credito",

## USAR WINSOR
"Credito_Anterior",

## USAR WINSOR
"Credito_DIF",

## USAR WINSOR
"Credito_PERC",

## USAR WINSOR
"Rentabilidade",

## USAR WINSOR
"Rentabilidade_Anterior",

## USAR WINSOR
"Rentabilidade_DIF",

## USAR WINSOR
"Rentabilidade_PERC")

copiaVariaveis1Winsored <- c("Produtos_Qualificados",
                             "winsored_Produtos_Qualificados_PERC",
                             "Produtos_DIF",
                             "winsored_Produtos_PERC",
                             "winsored_Movimentacoes",
                             "winsored_Movimentacoes_PERC",
                             "winsored_Credito",
                             "winsored_Credito_Anterior",
                             "winsored_Rentabilidade",
                             "winsored_Rentabilidade_Anterior")

variaveis2 <- c('winsored_Aplicacao',
                'winsored_Aplicacao_Anterior',
                'winsored_Aplicacao_DIF',
                'winsored_Aplicacao_PERC',
                'winsored_Credito',
                'winsored_Credito_Anterior',
                'winsored_Credito_DIF',
                'winsored_Credito_PERC',
                'winsored_Rentabilidade',

```

```

        'winsored_Rentabilidade_Anterior',
        'winsored_Rentabilidade_DIF',
        'winsored_Rentabilidade_PERC')

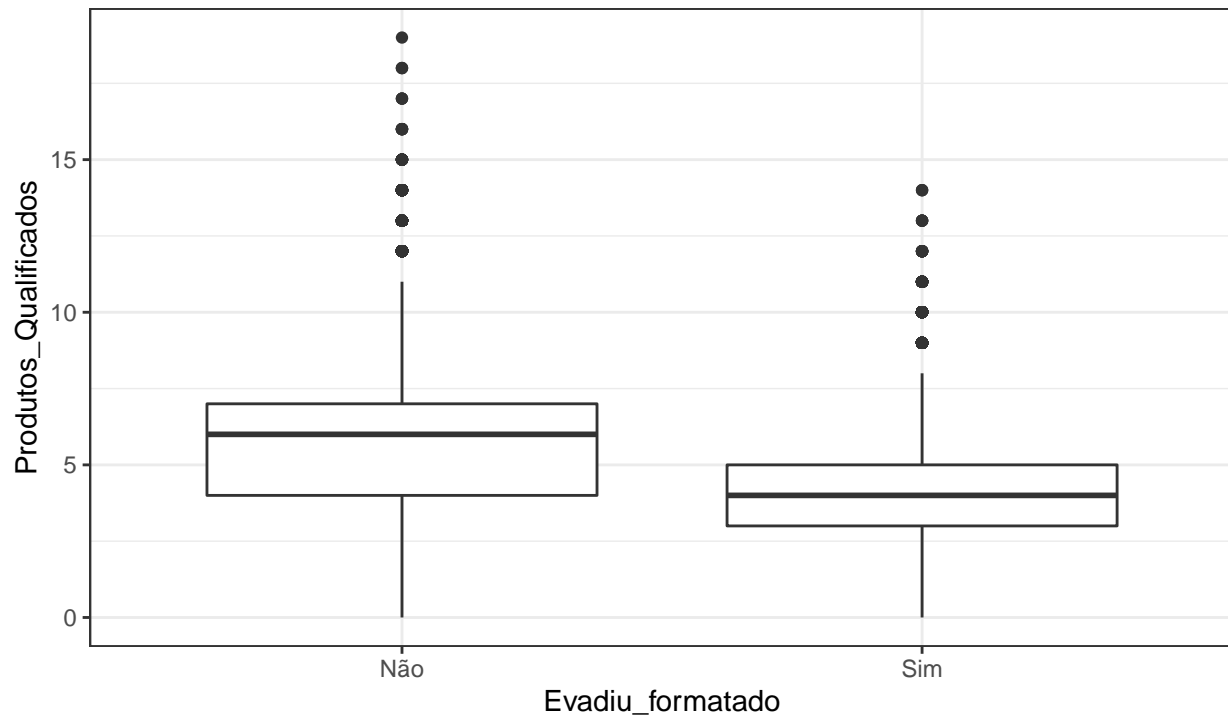
box_plot = function(data, atributos, atributosSubgrupo, variavelDependente = 'Evadiu_formatado'){
  for (atributo in atributos) {
    if(length(atributosSubgrupo) > 0){
      for (atributo2 in atributosSubgrupo){
        result = ggplot(data = dtAmostra, aes_string(variavelDependente, atributo)) +
          geom_boxplot(aes_string(fill = atributo2)) +
          # geom_boxplot() +
          labs(x = variavelDependente,
               y = atributo,
               title = "Boxplot",
               subtitle = paste(atributo, variavelDependente, sep = " vs "),
               caption = "") +
          theme_bw()
          # facet_wrap(~ atributo2)
        print(result)
      }
    }
    else{
      result = ggplot(data = dtAmostra, aes_string(variavelDependente, atributo)) +
        geom_boxplot() +
        labs(x = variavelDependente,
             y = atributo,
             title = "Boxplot",
             subtitle = paste(atributo, variavelDependente, sep = " vs "),
             caption = "") +
        theme_bw()
        # facet_wrap(~ atributo2)
      print(result)
    }
  }
}

box_plot(credit, copiaVariaveis1Winsored, NULL)

```

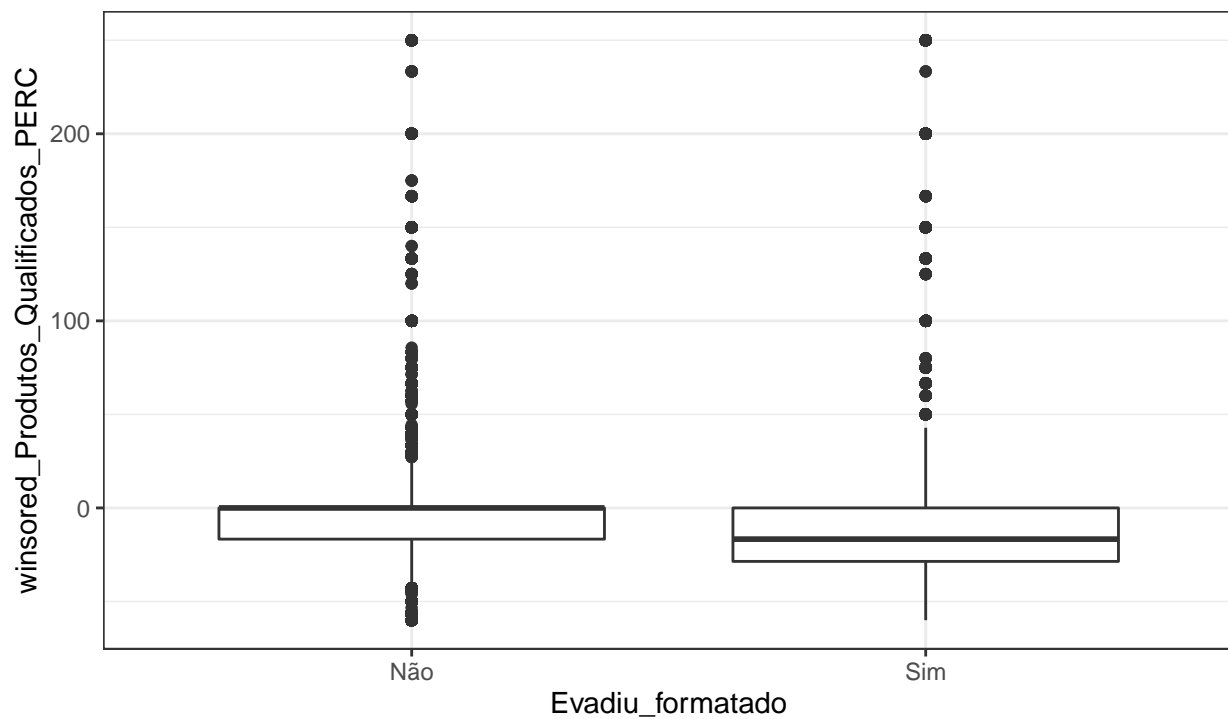
Boxplot

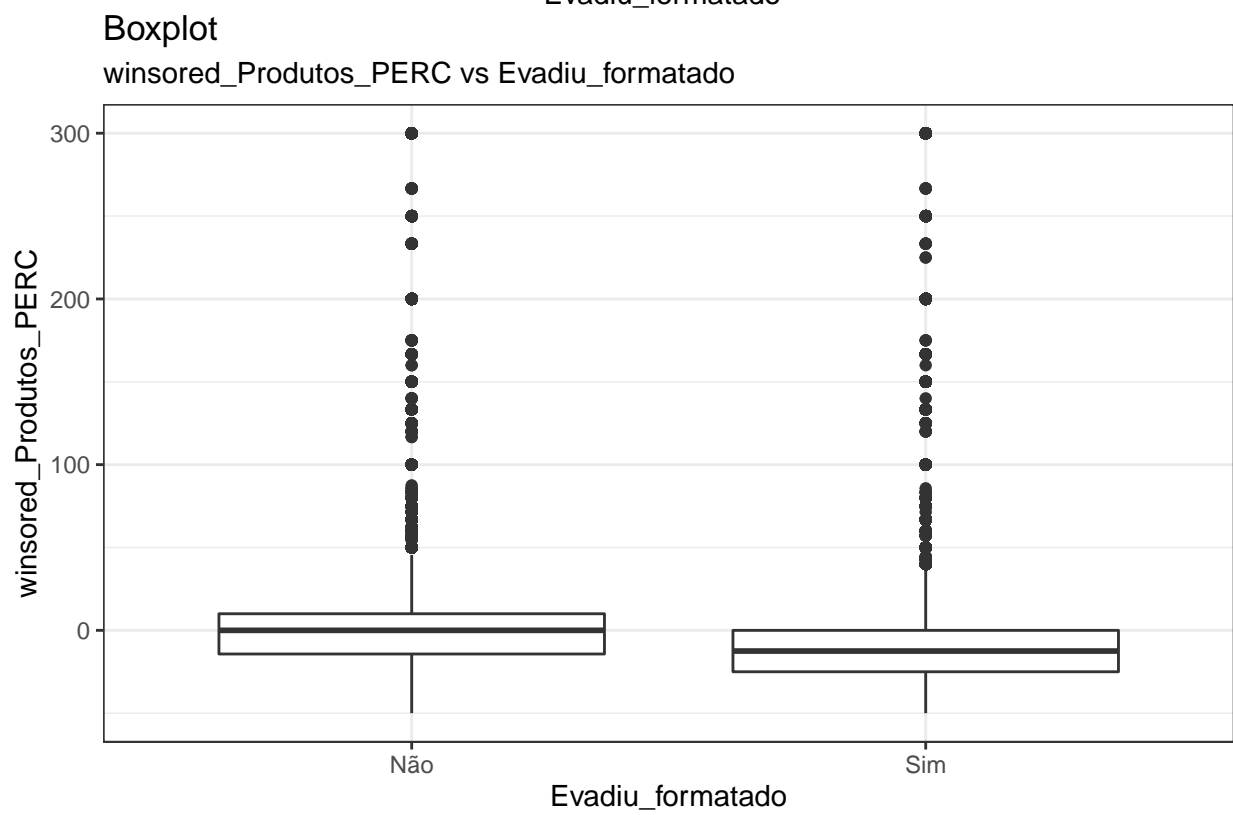
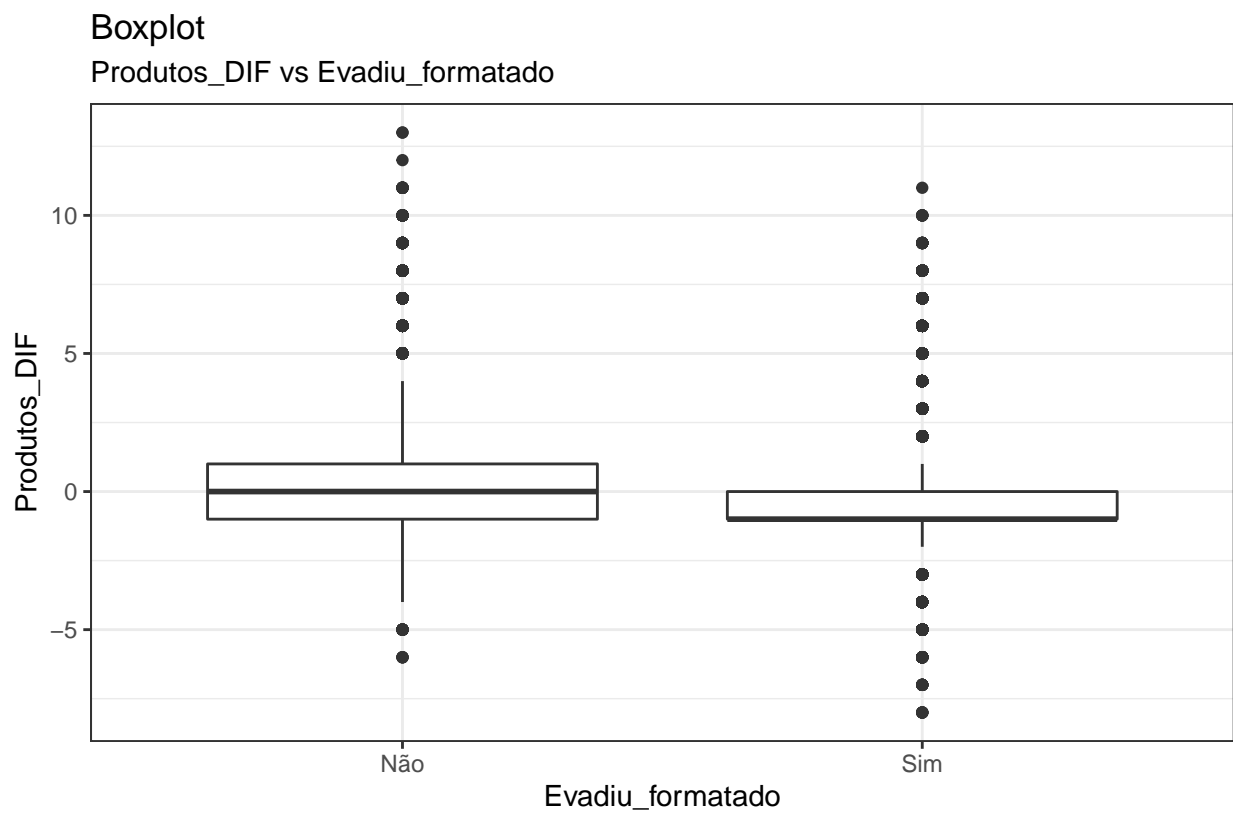
Produtos_Qualificados vs Evadiu_formatado



Boxplot

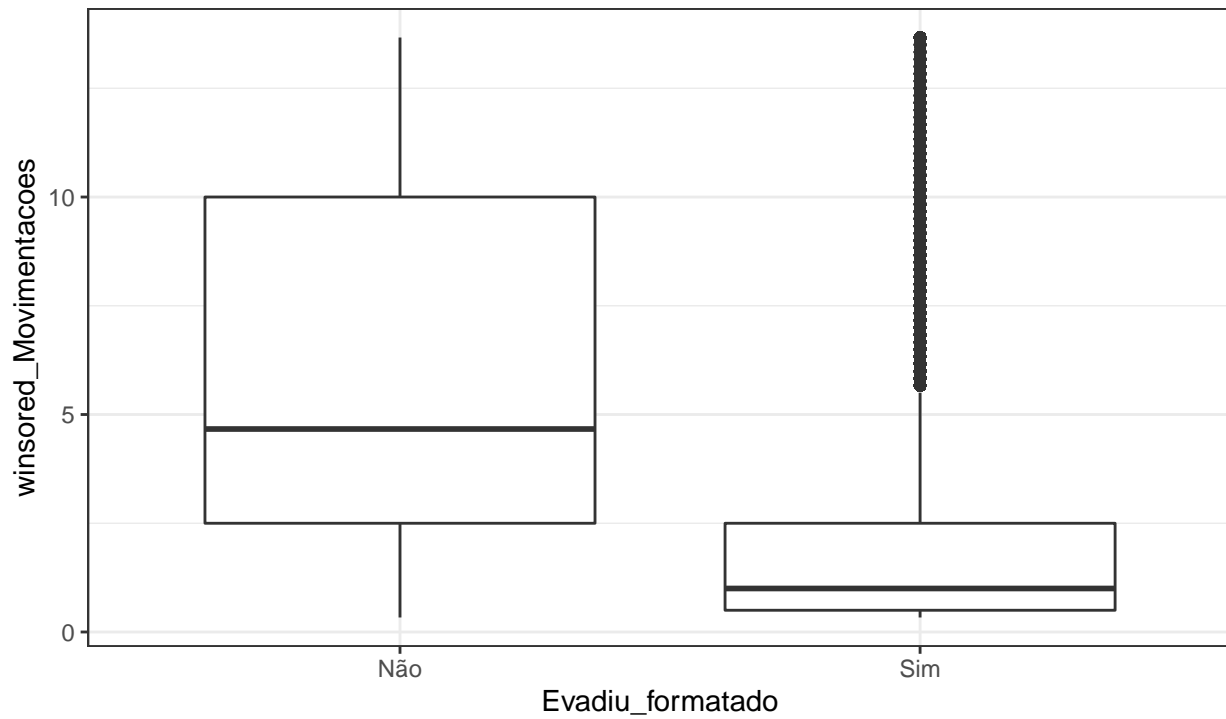
winsored_Produtos_Qualificados_PERC vs Evadiu_formatado





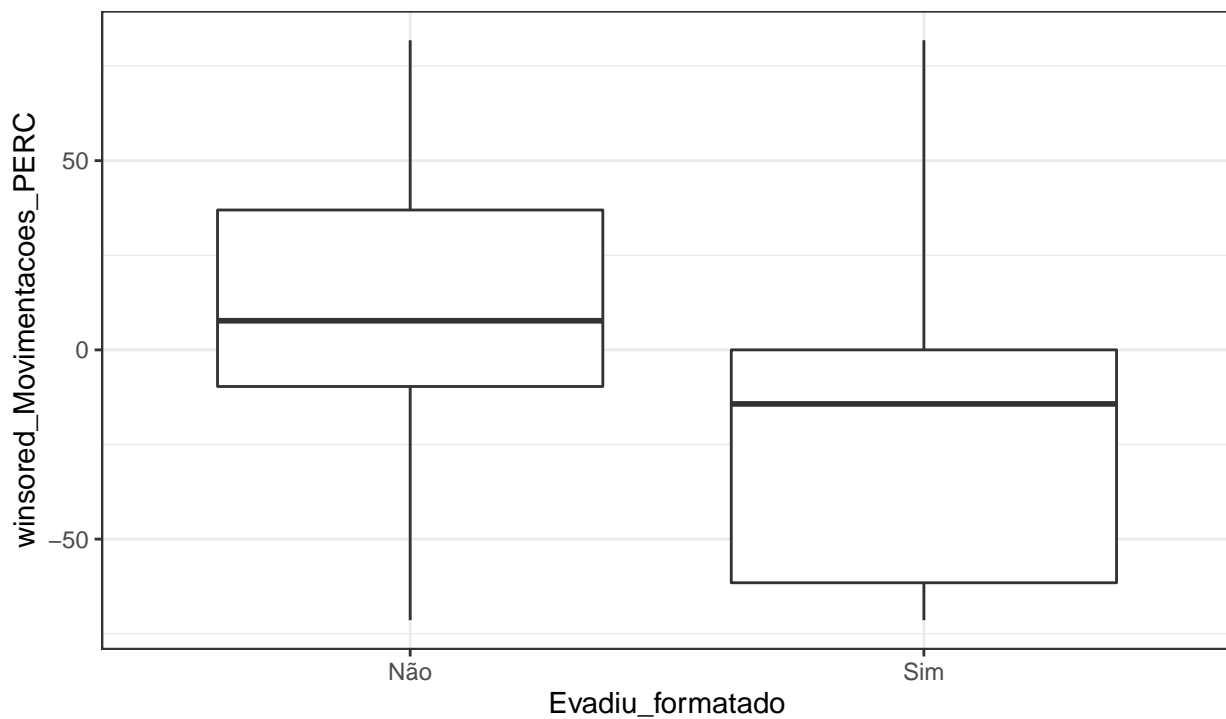
Boxplot

winsored_Movimentacoes vs Evadiu_formatado



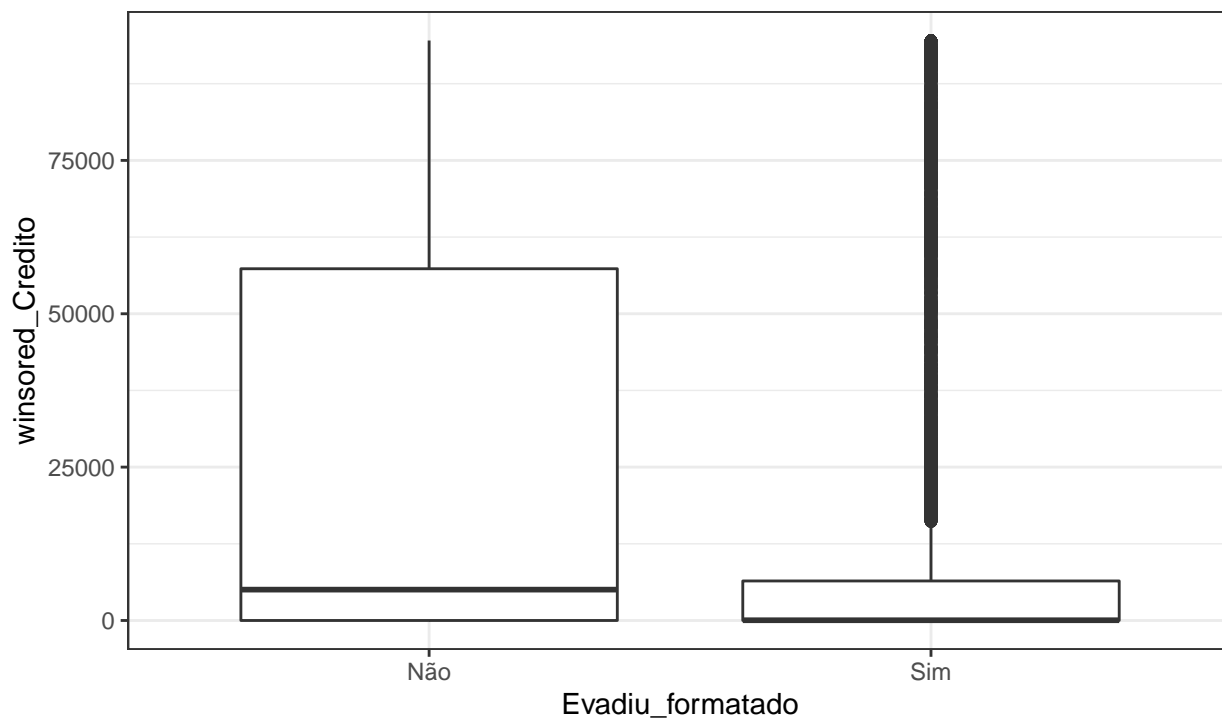
Boxplot

winsored_Movimentacoes_PERC vs Evadiu_formatado



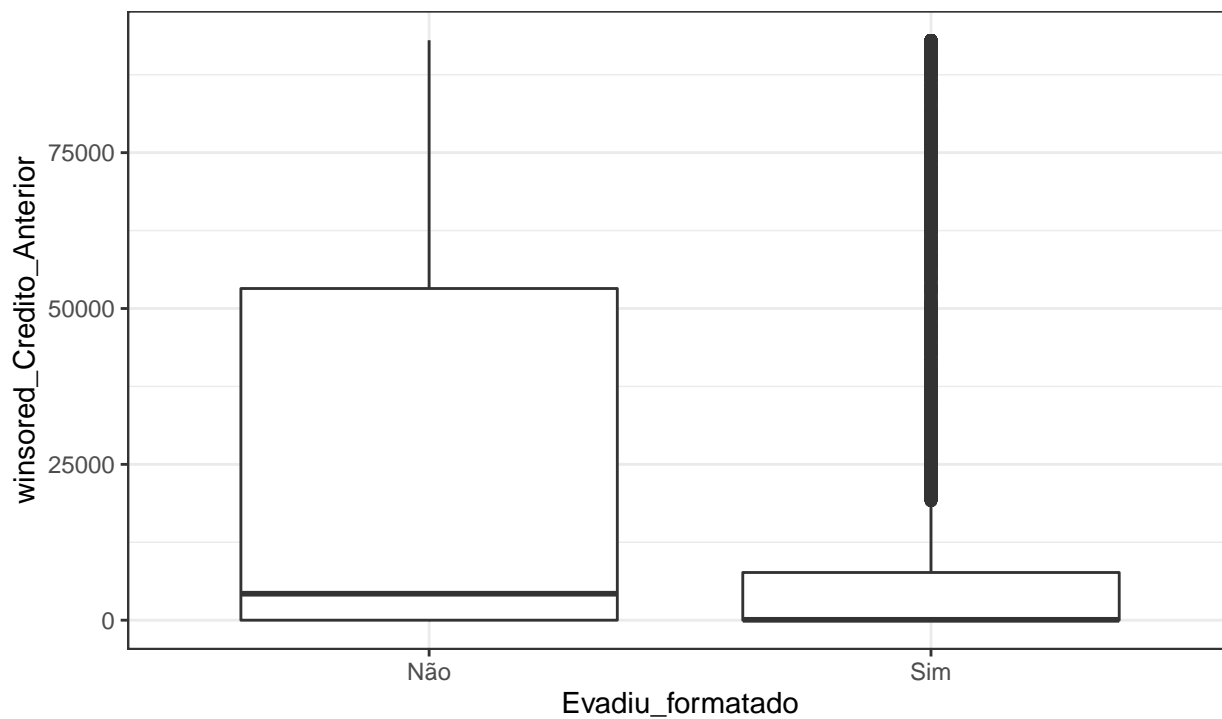
Boxplot

winsored_Credito vs Evadiu_formatado



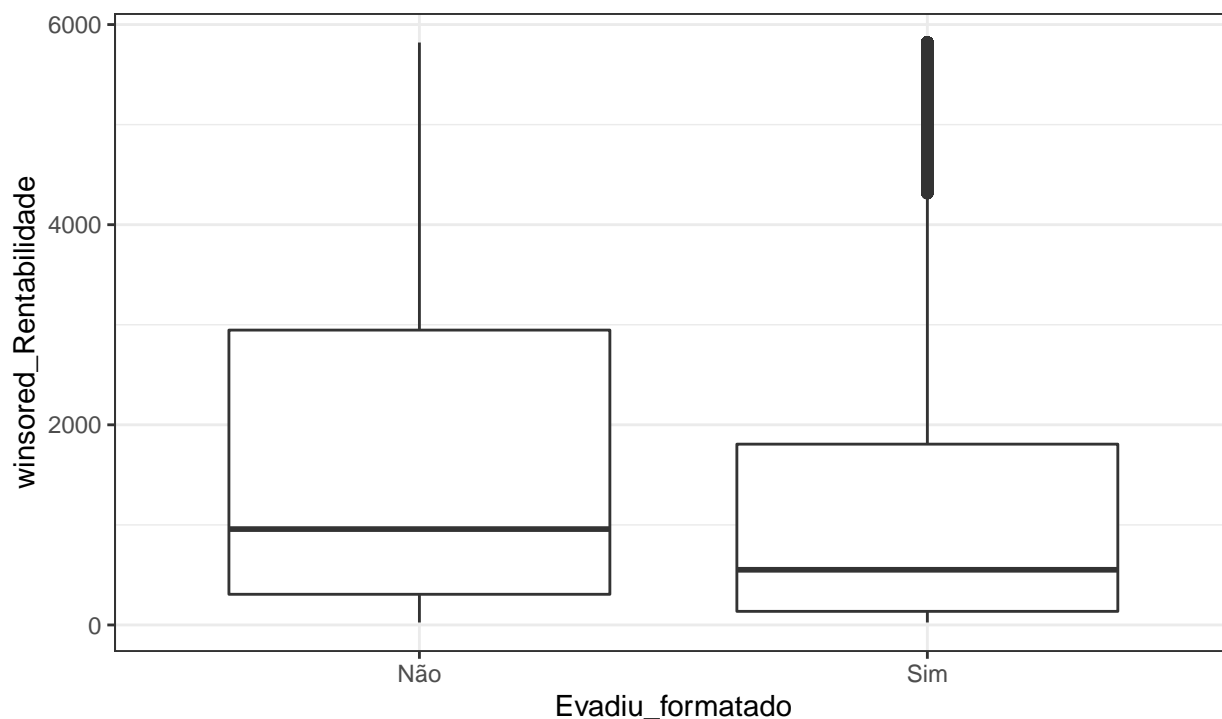
Boxplot

winsored_Credito_Anterior vs Evadiu_formatado



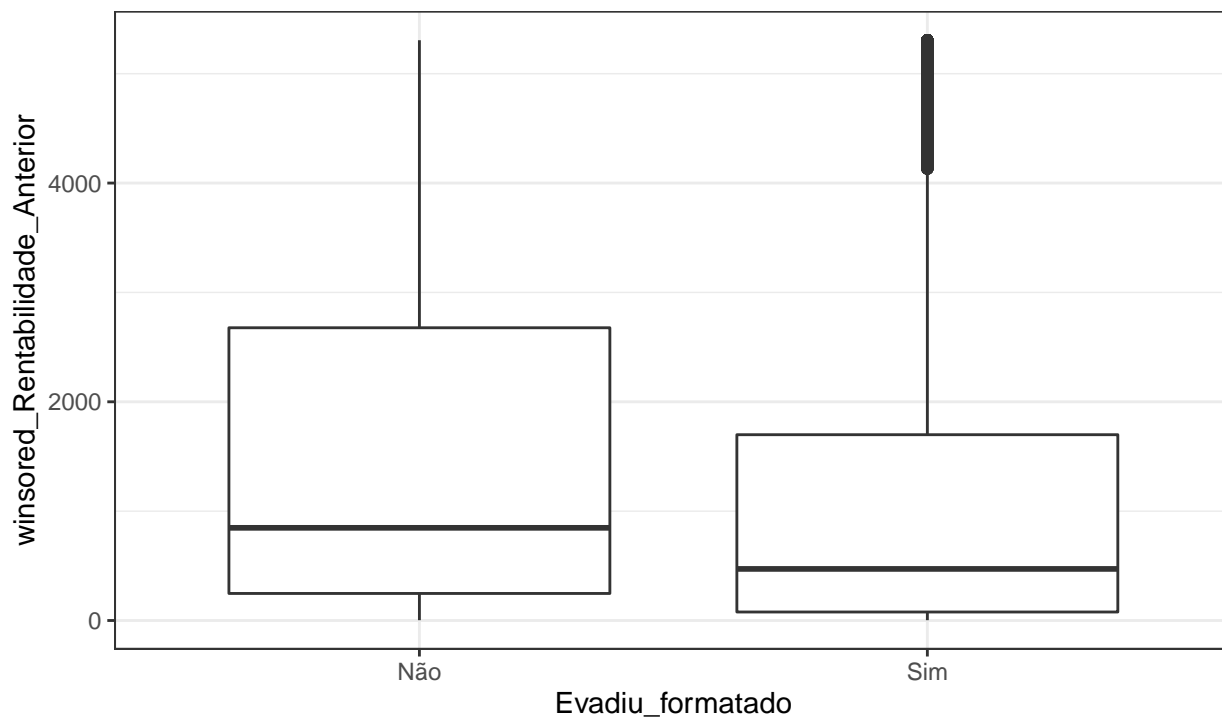
Boxplot

winsored_Rentabilidade vs Evadiu_formatado

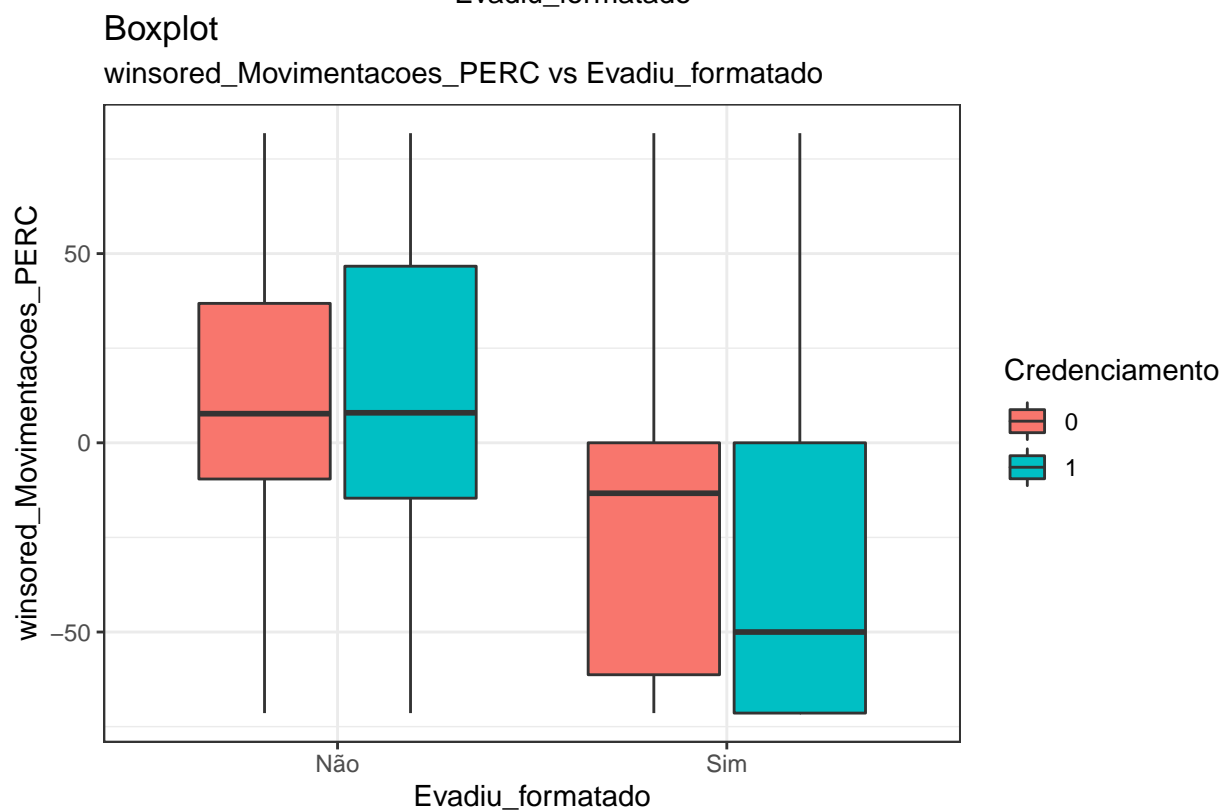
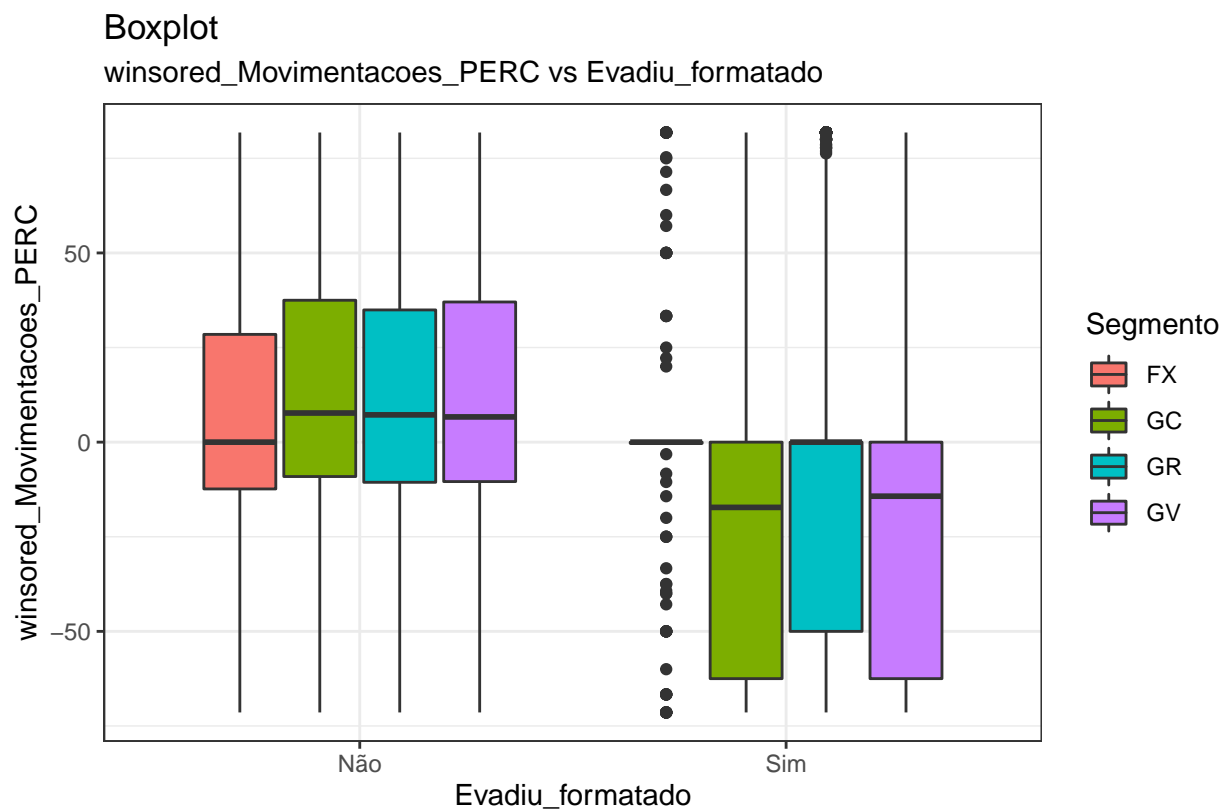


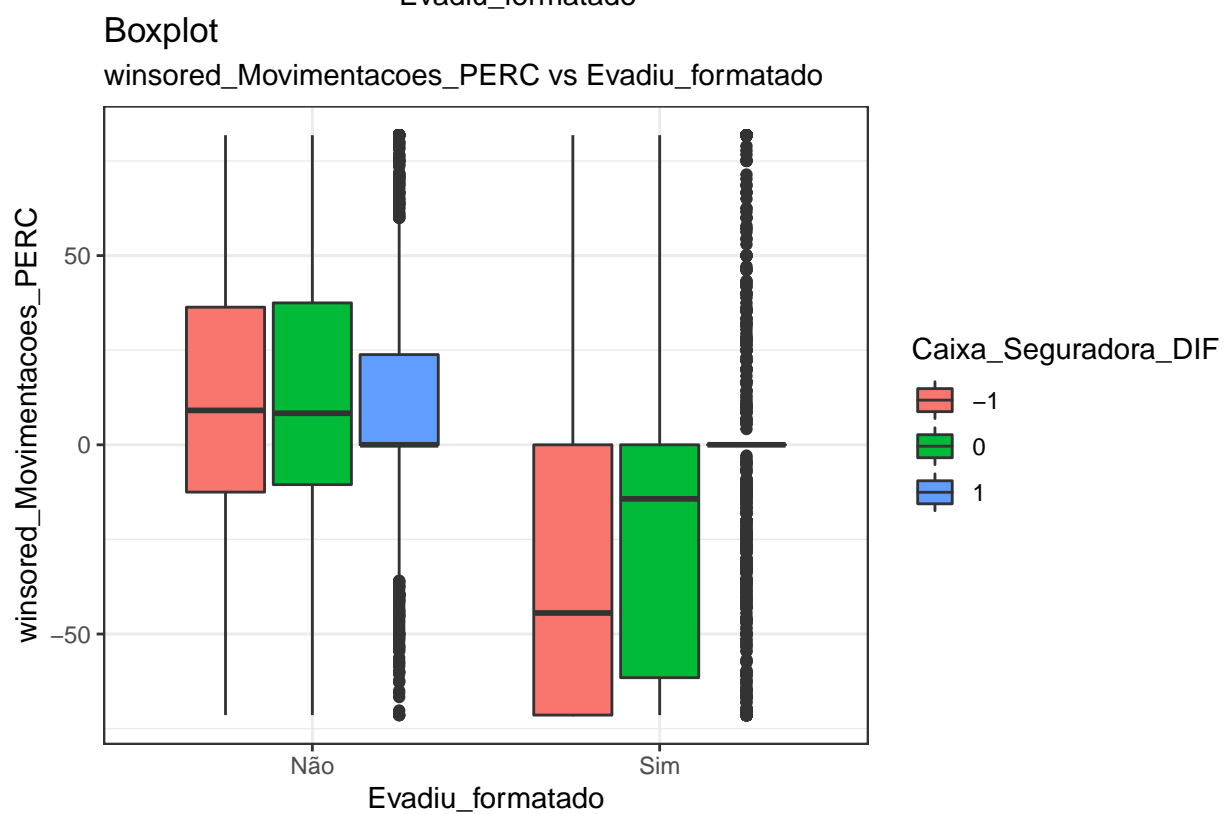
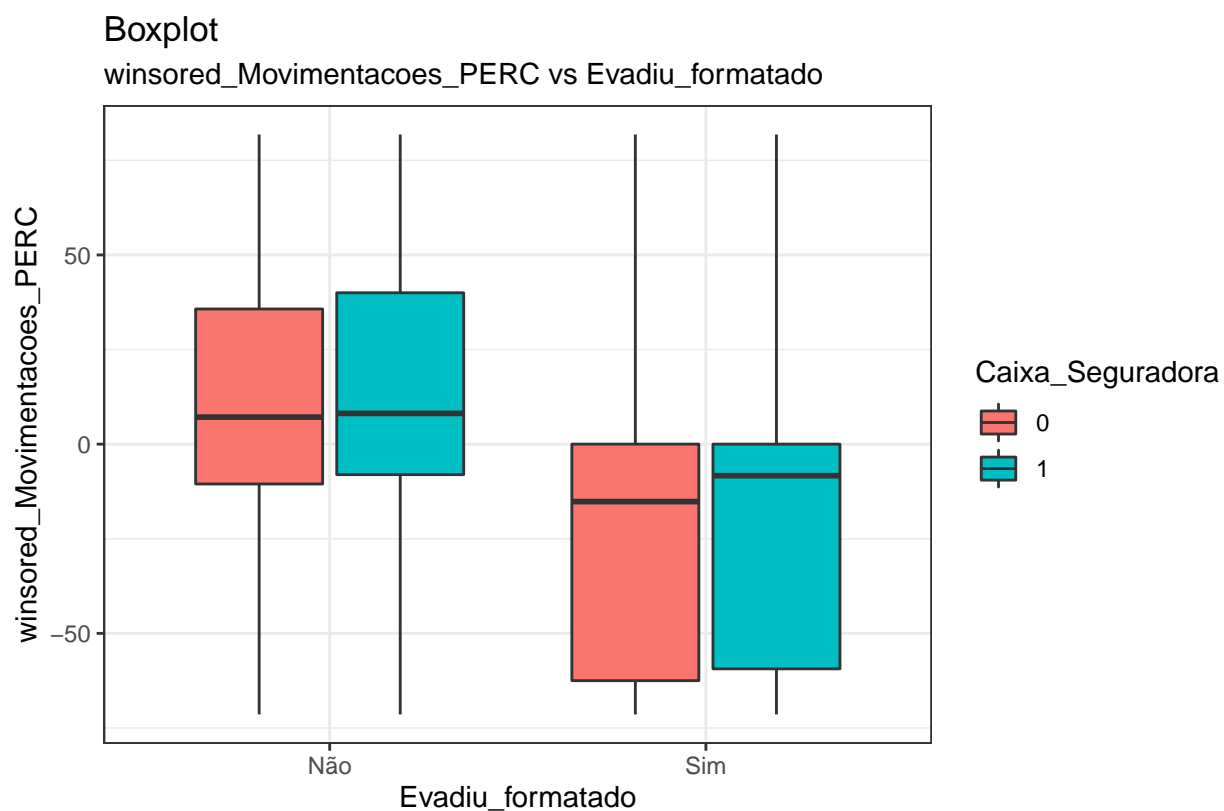
Boxplot

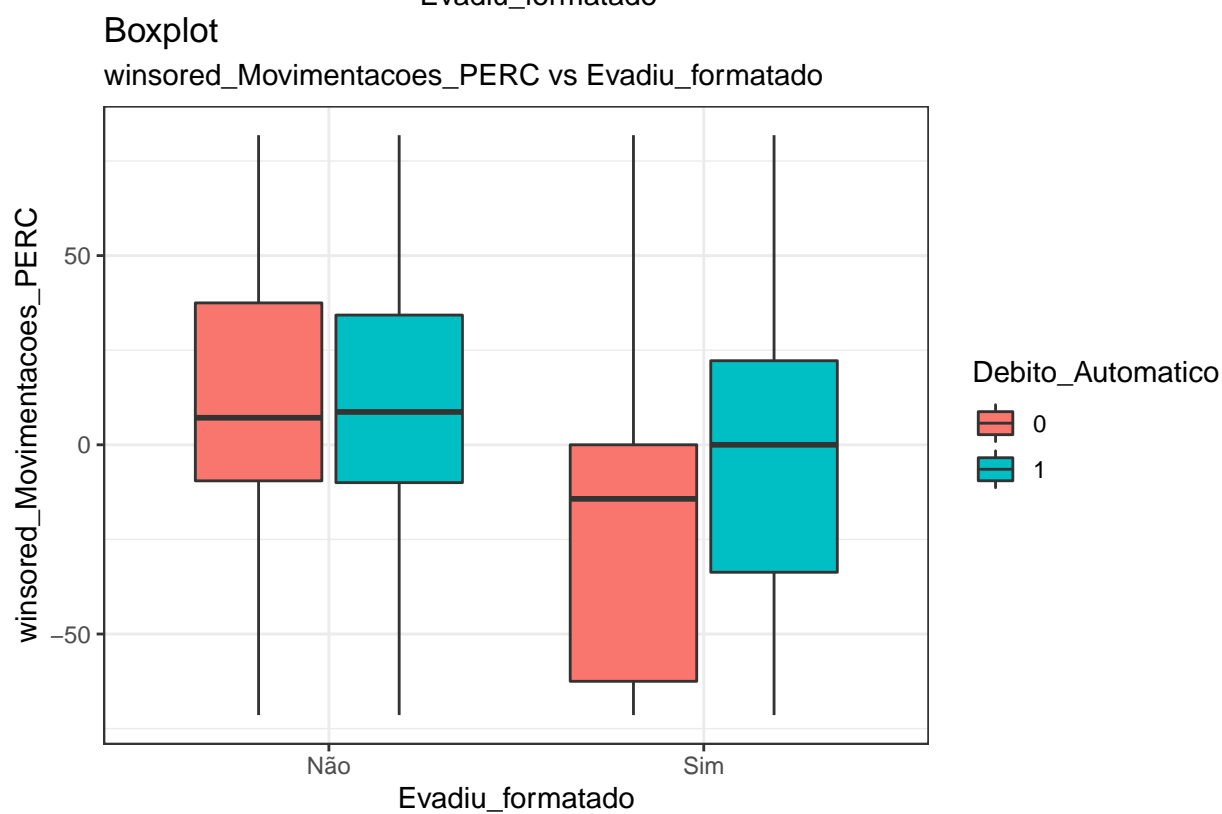
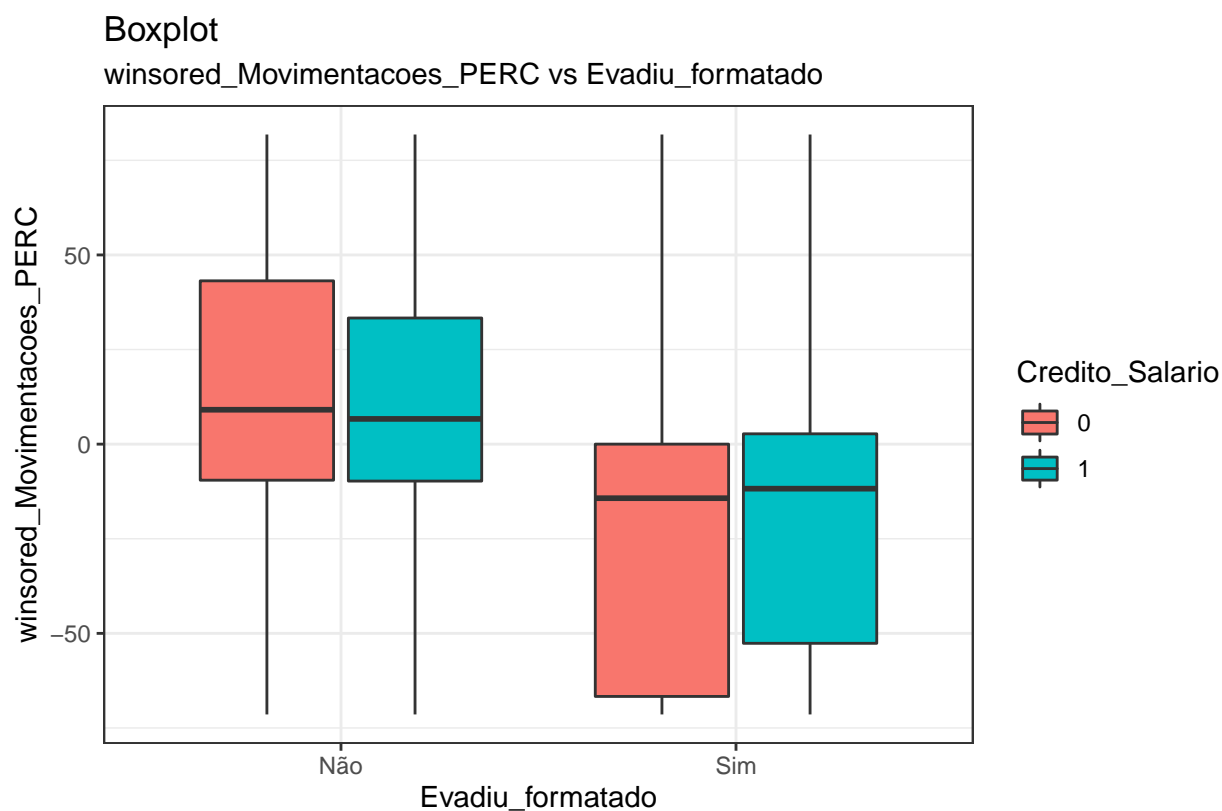
winsored_Rentabilidade_Anterior vs Evadiu_formatado



```
box_plot(credit, c("winsored_Movimentacoes_PERC"), variaveisSubgrupo)
```

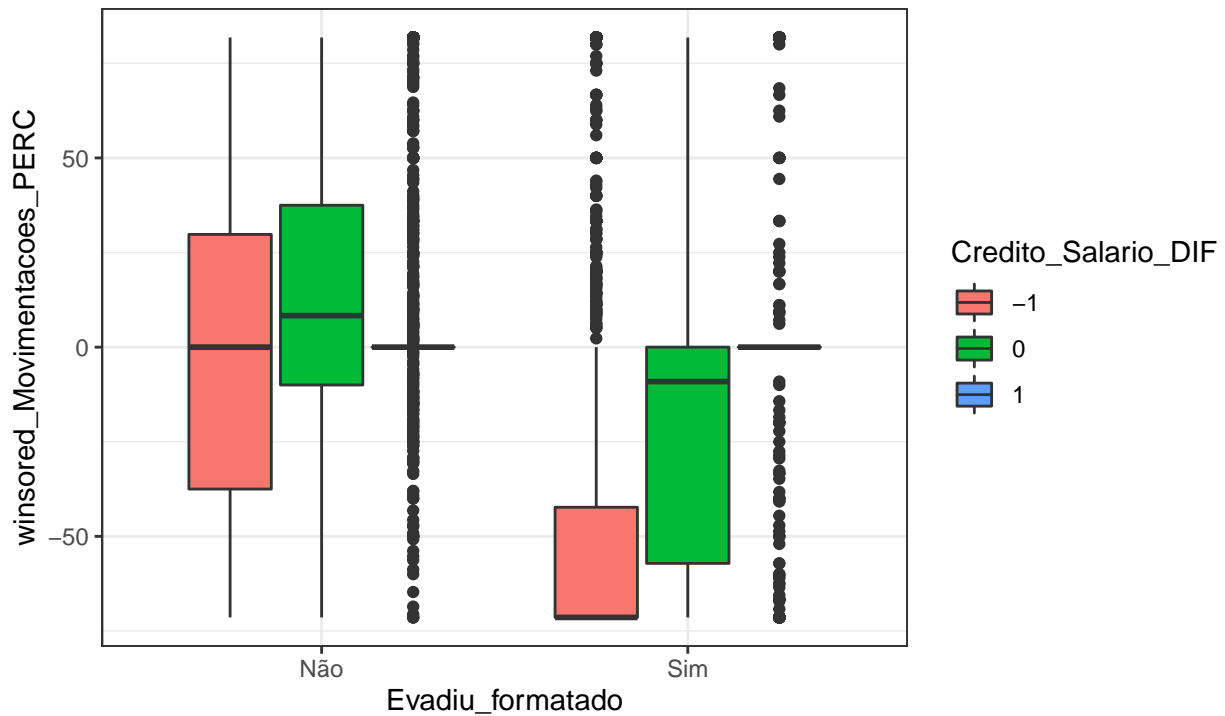






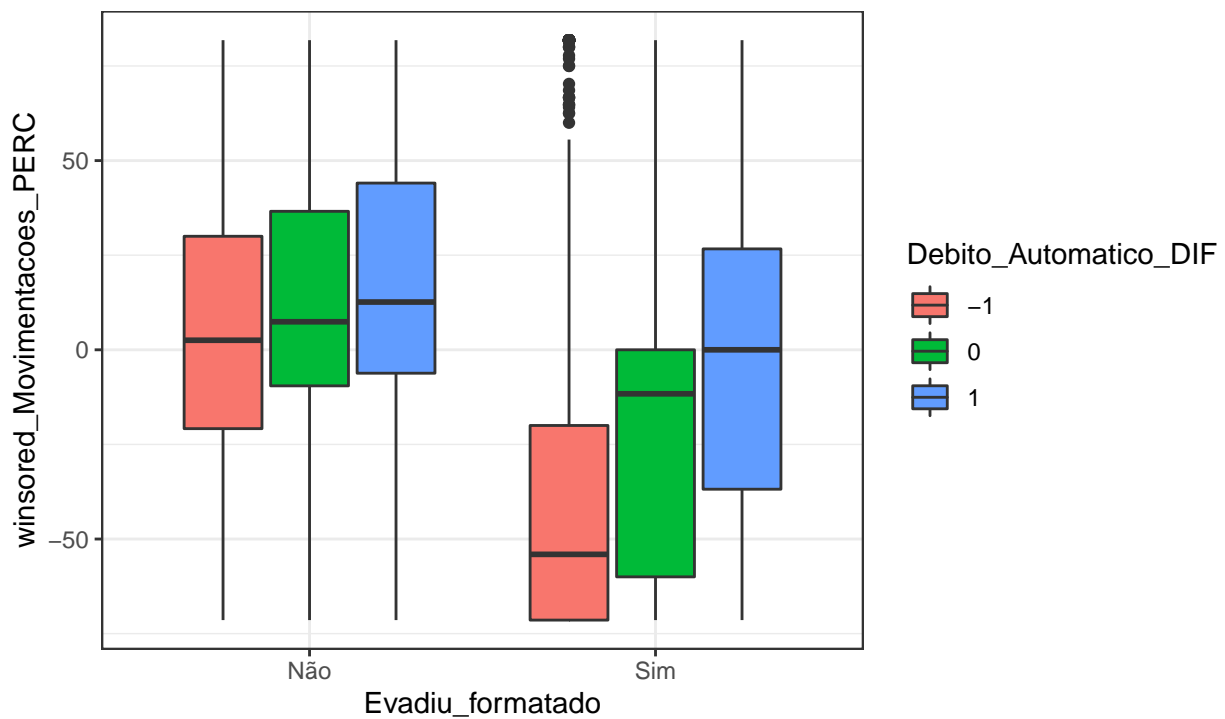
Boxplot

winsored_Movimentacoes_PERC vs Evadiu_formatado

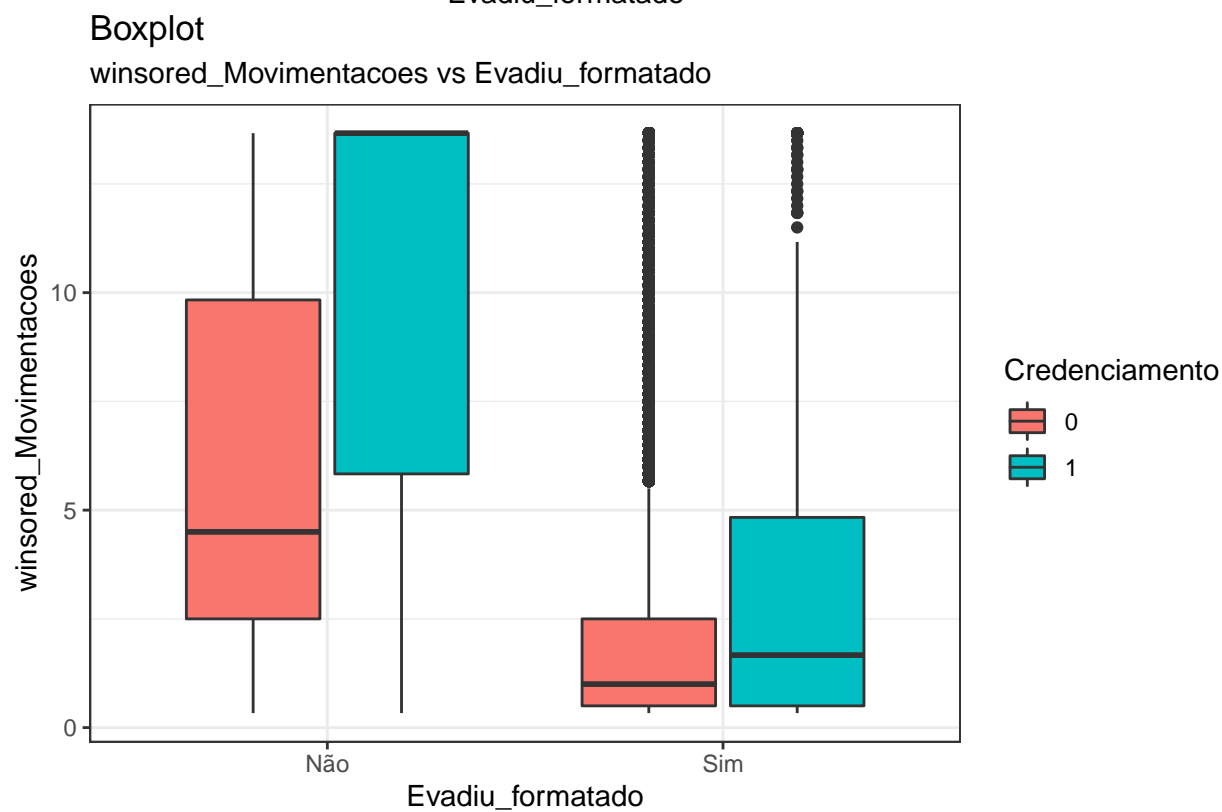
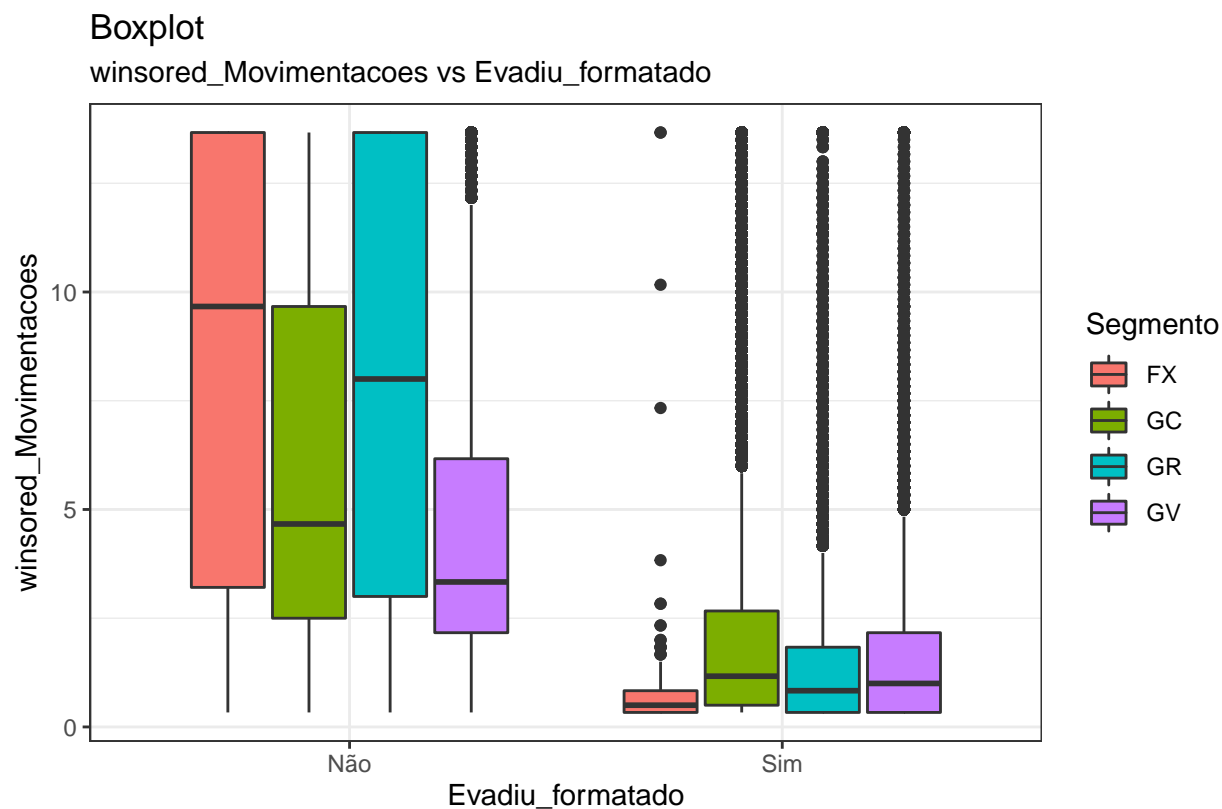


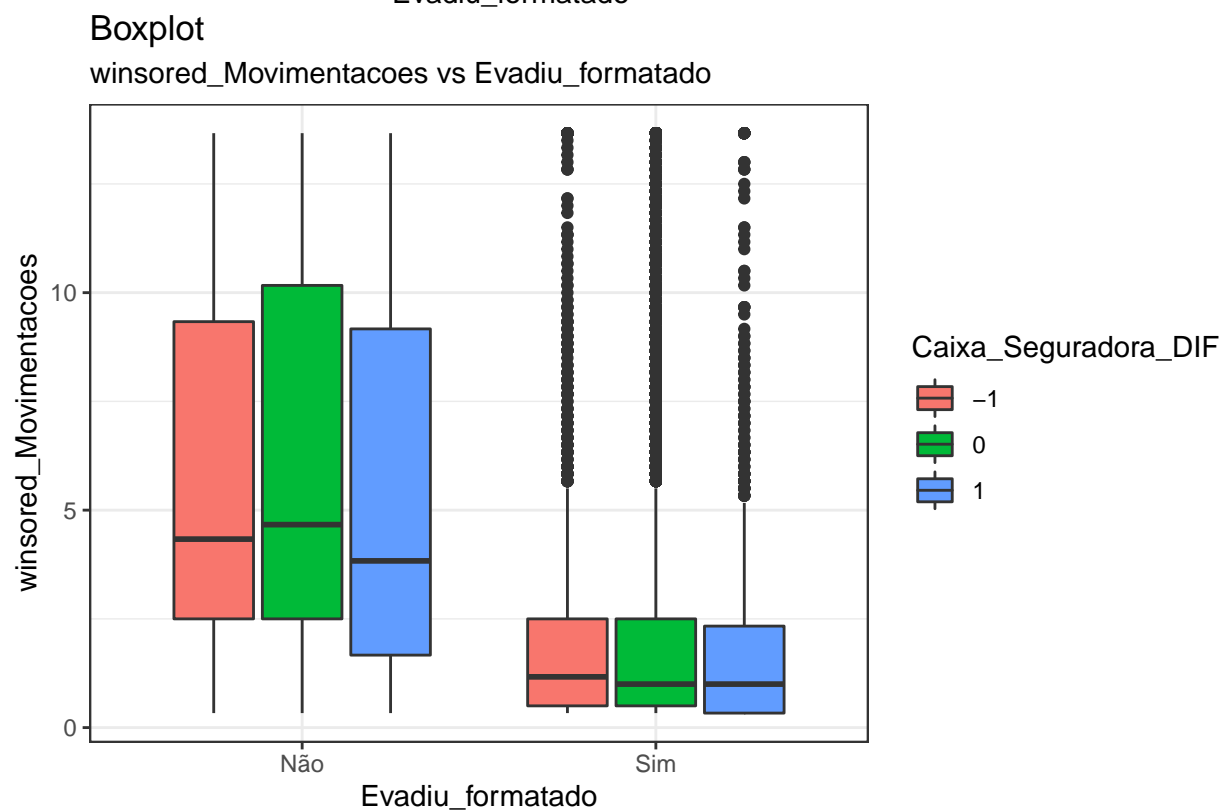
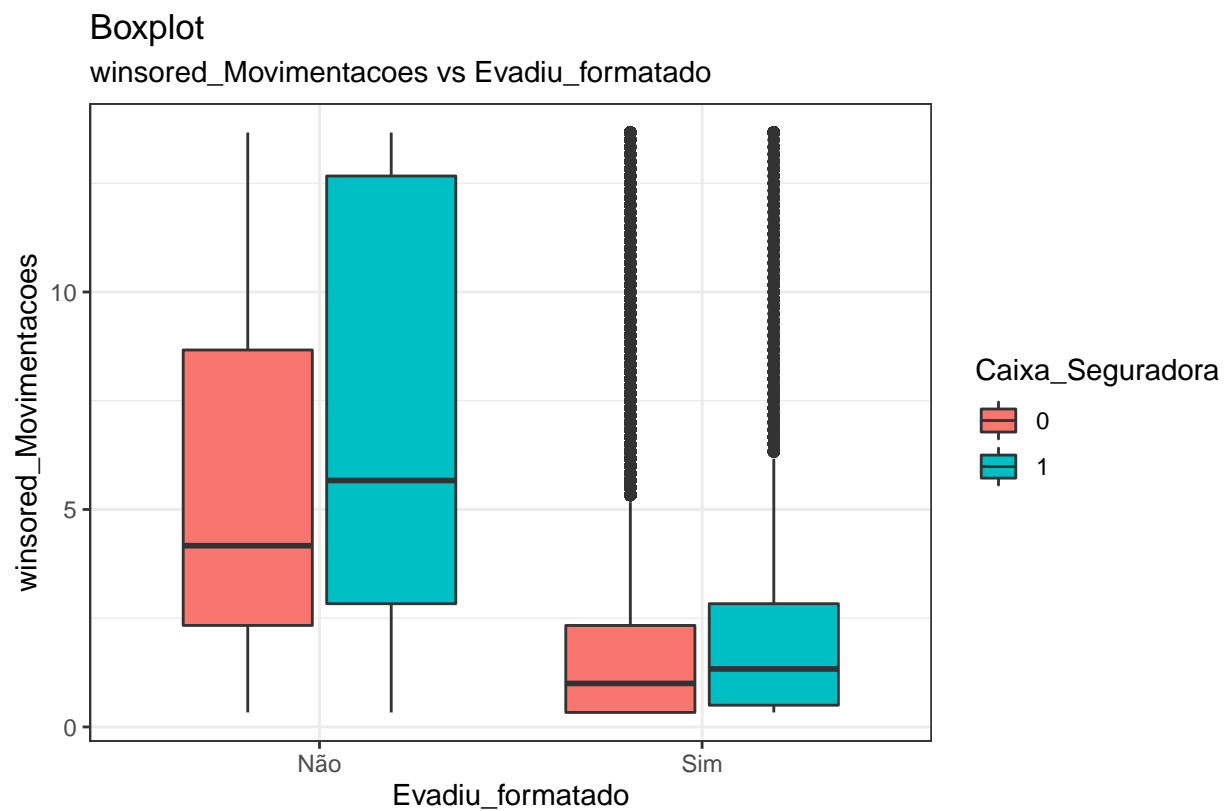
Boxplot

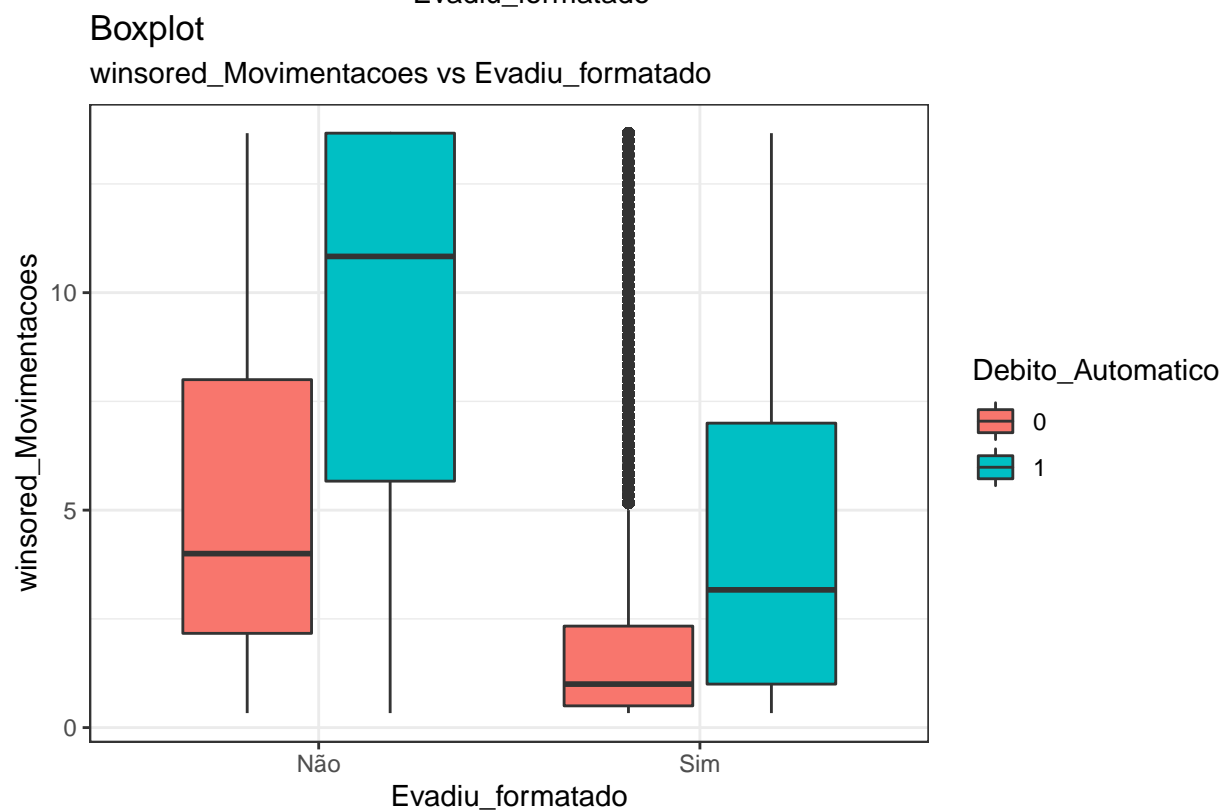
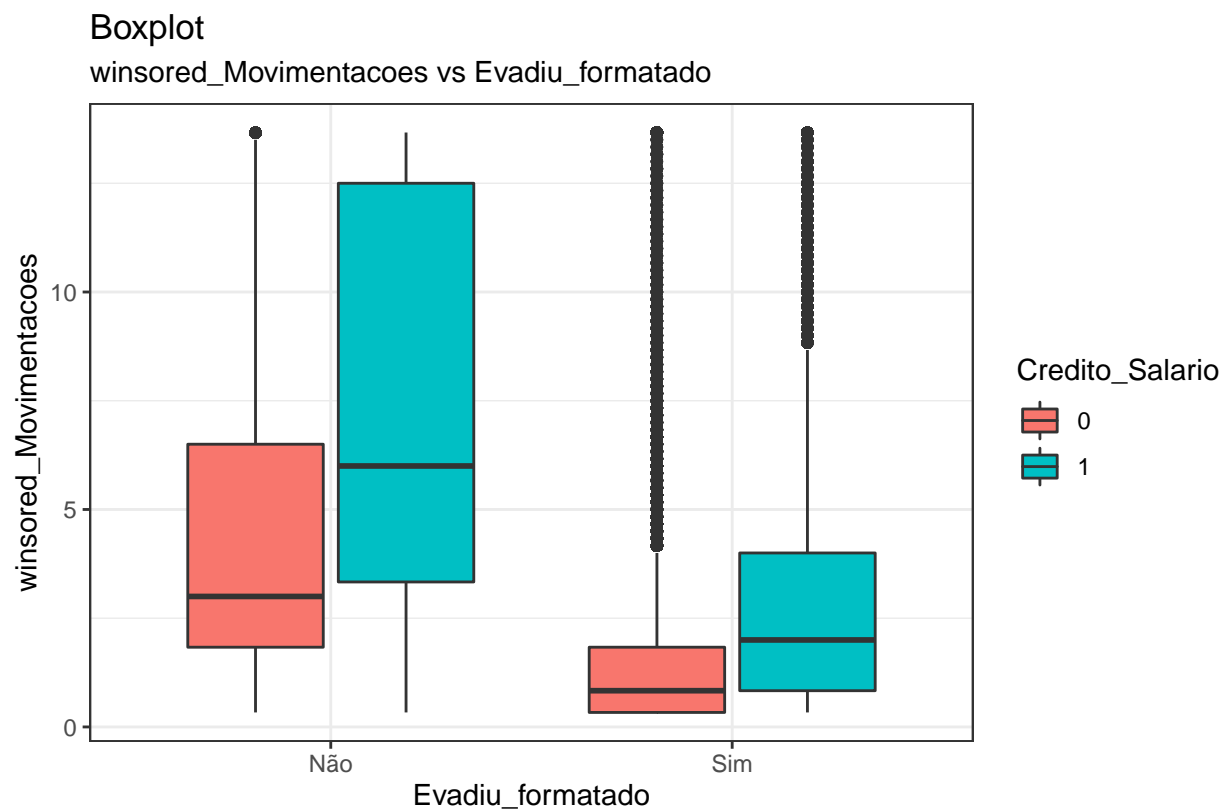
winsored_Movimentacoes_PERC vs Evadiu_formatado



```
box_plot(credit, c("winsored_Movimentacoes"), variaveisSubgrupo)
```

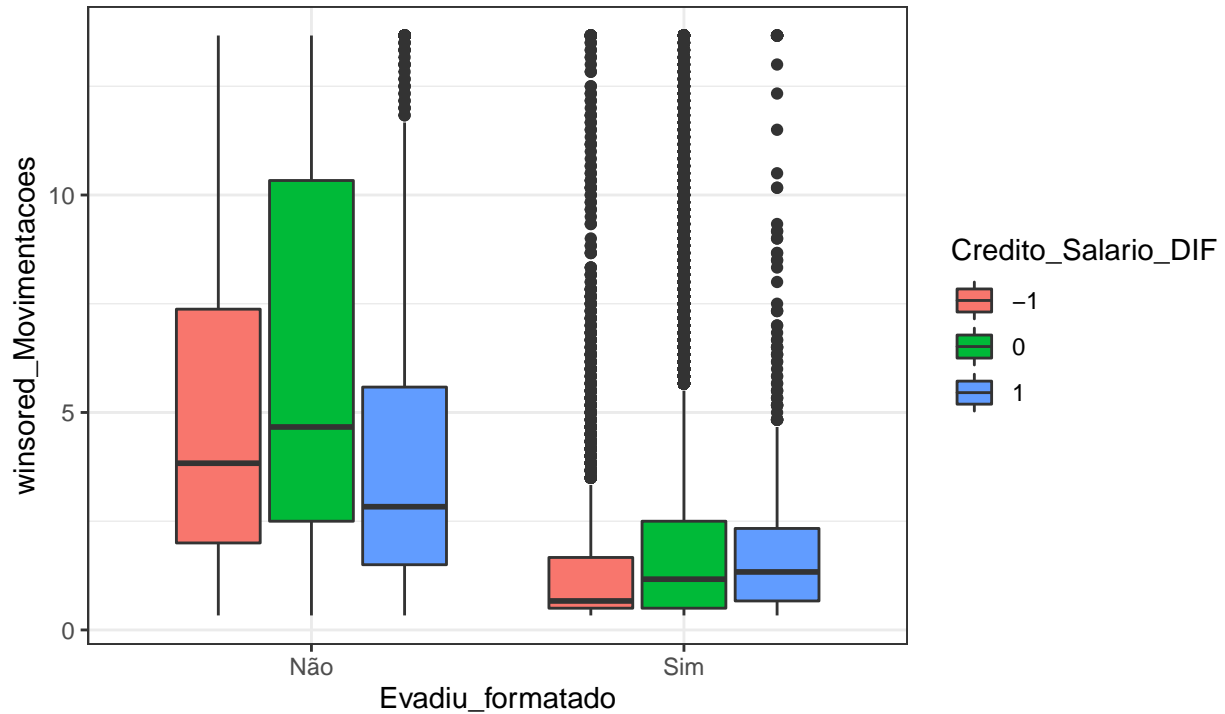






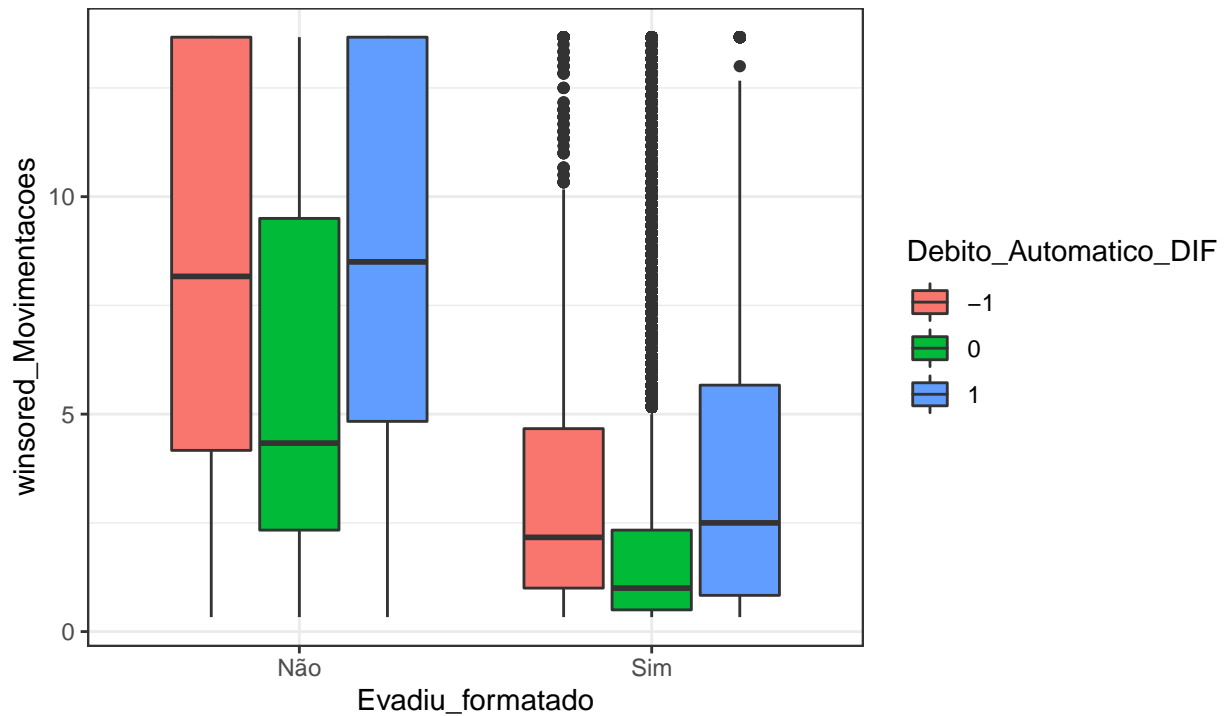
Boxplot

winsored_Movimentacoes vs Evadiu_formatado

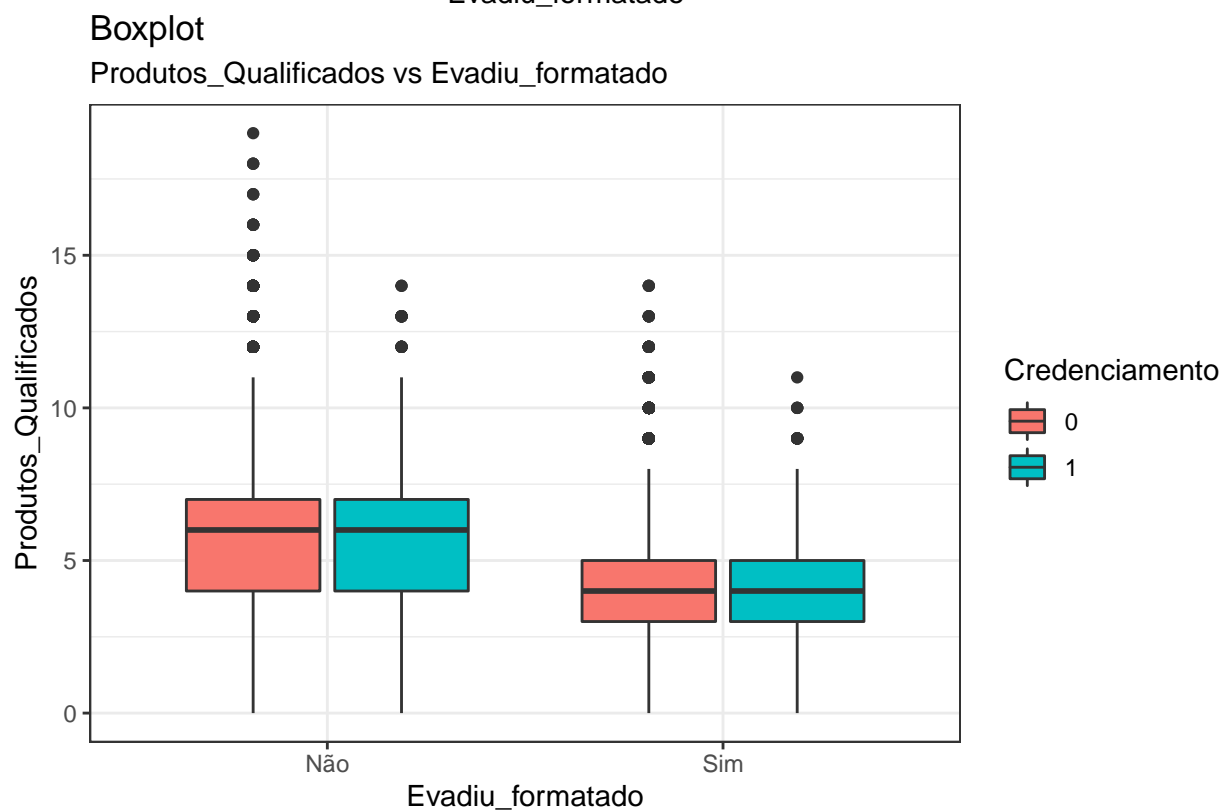
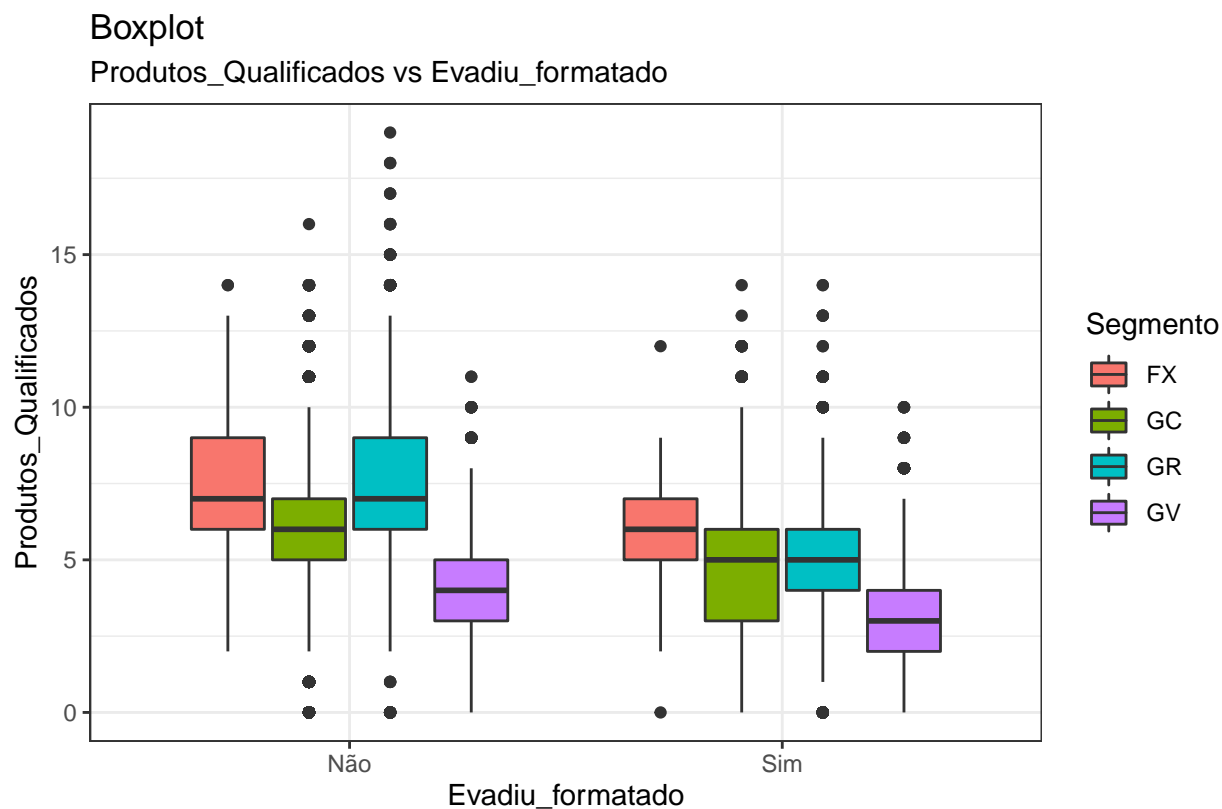


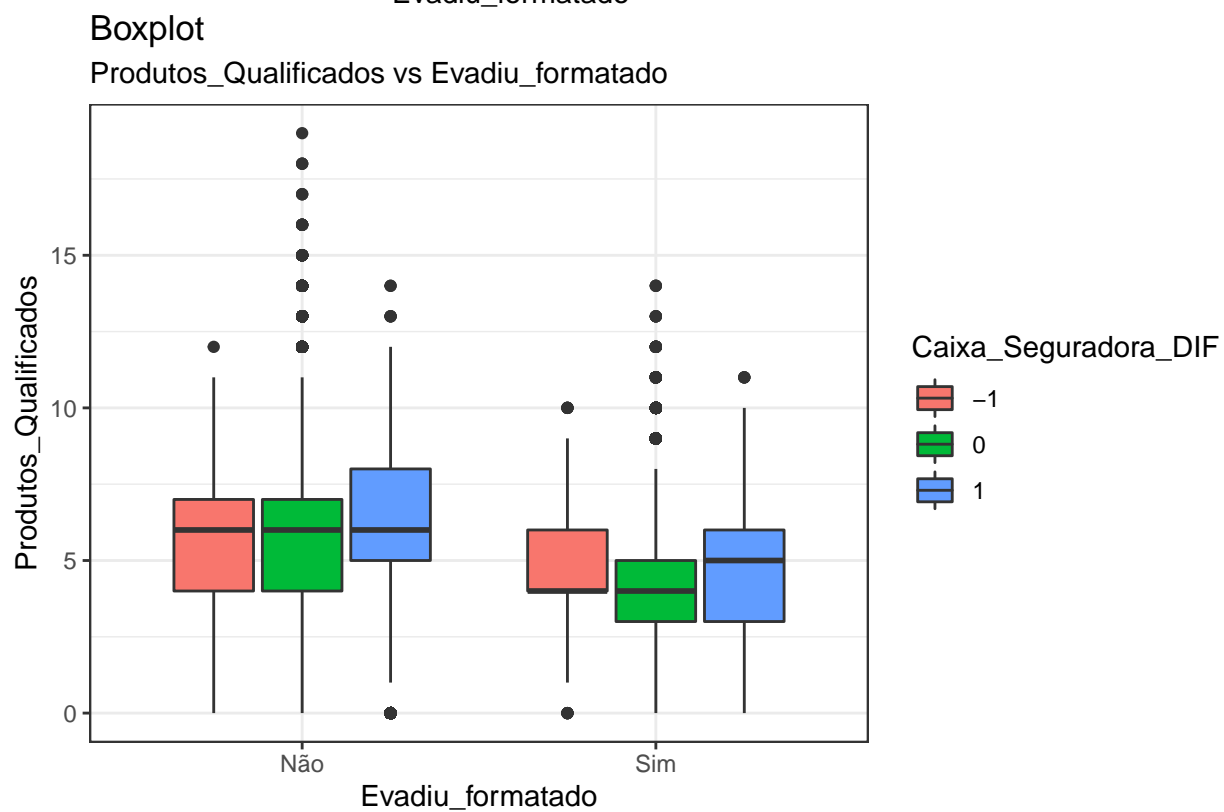
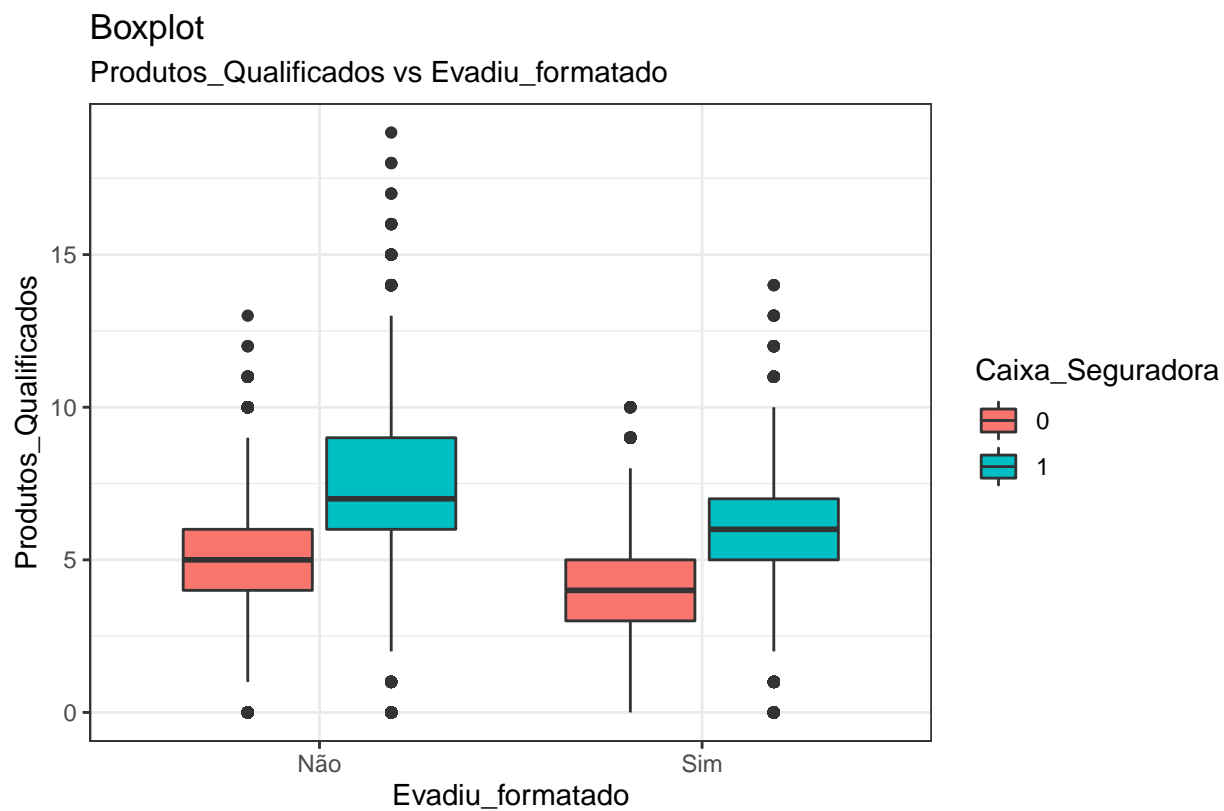
Boxplot

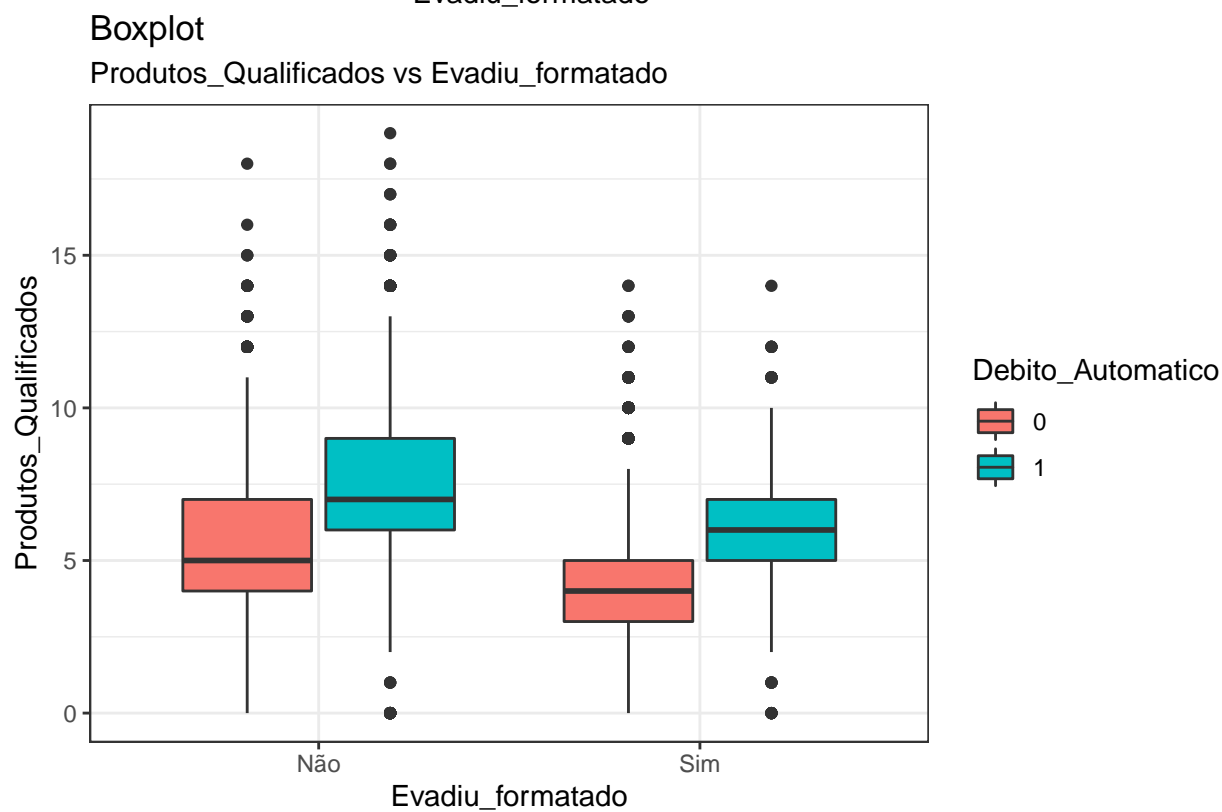
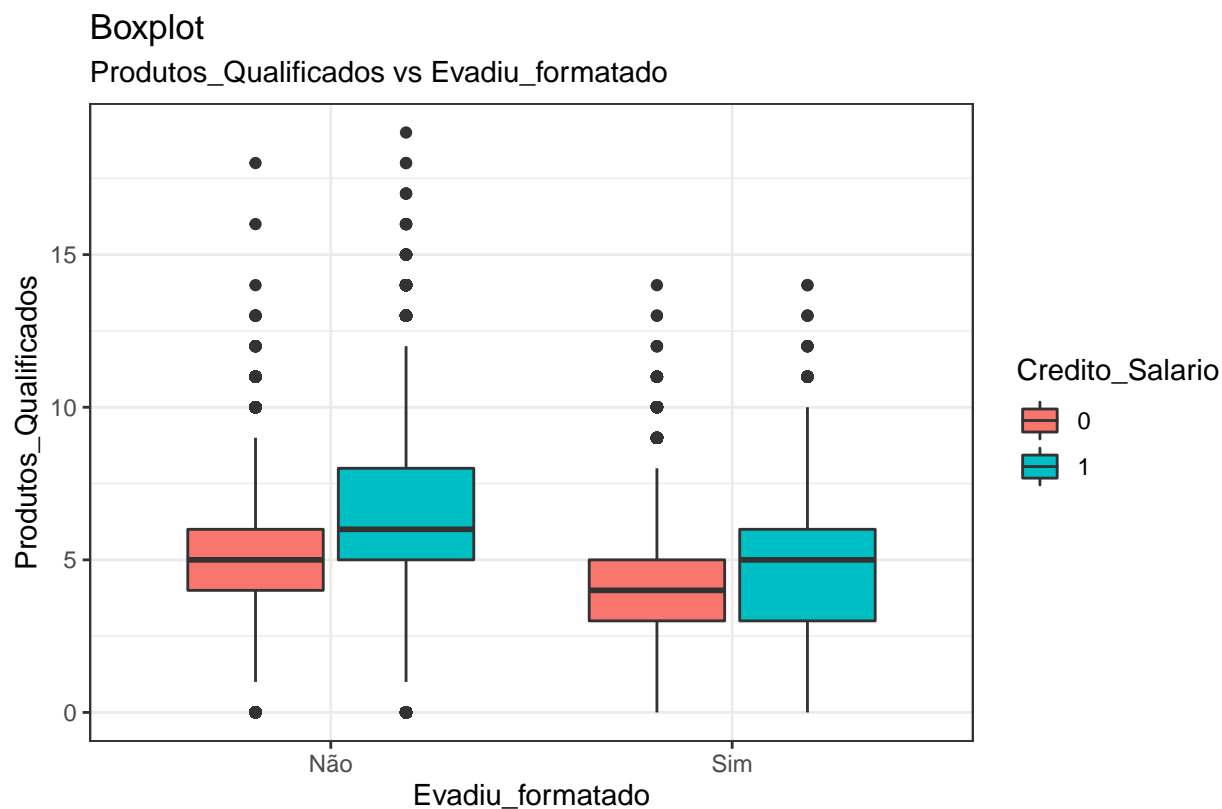
winsored_Movimentacoes vs Evadiu_formatado



```
box_plot(credit, c("Produtos_Qualificados"), variaveisSubgrupo)
```

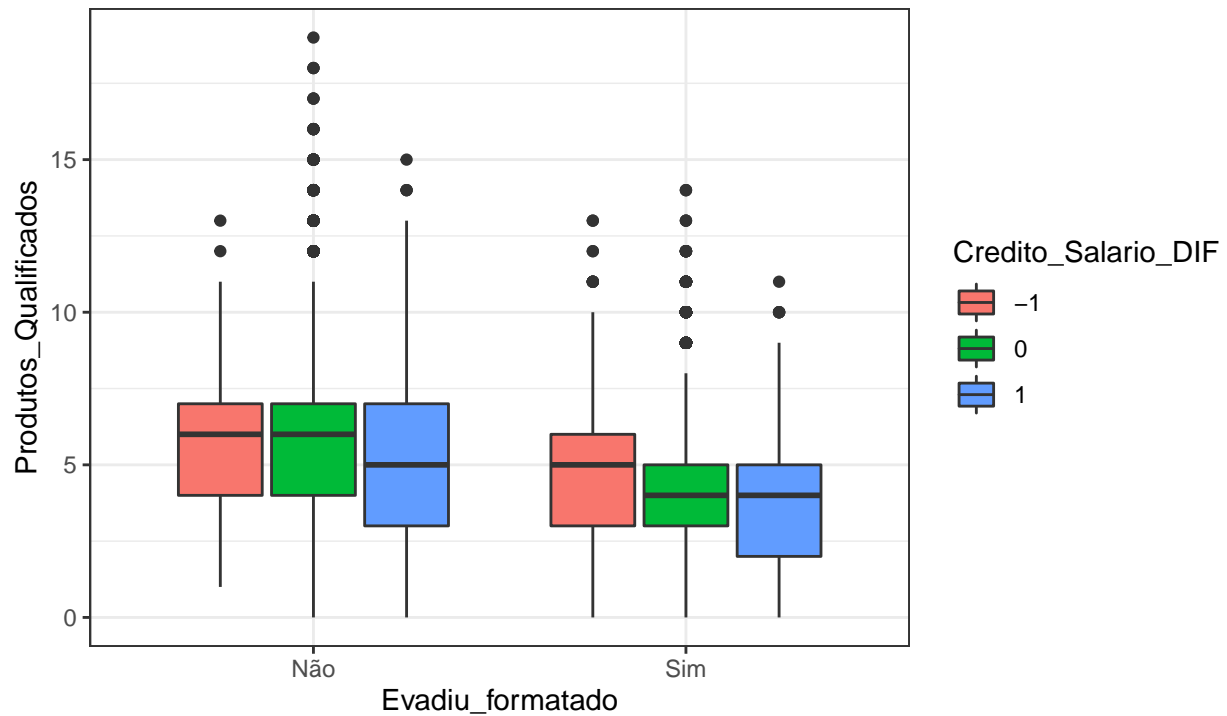






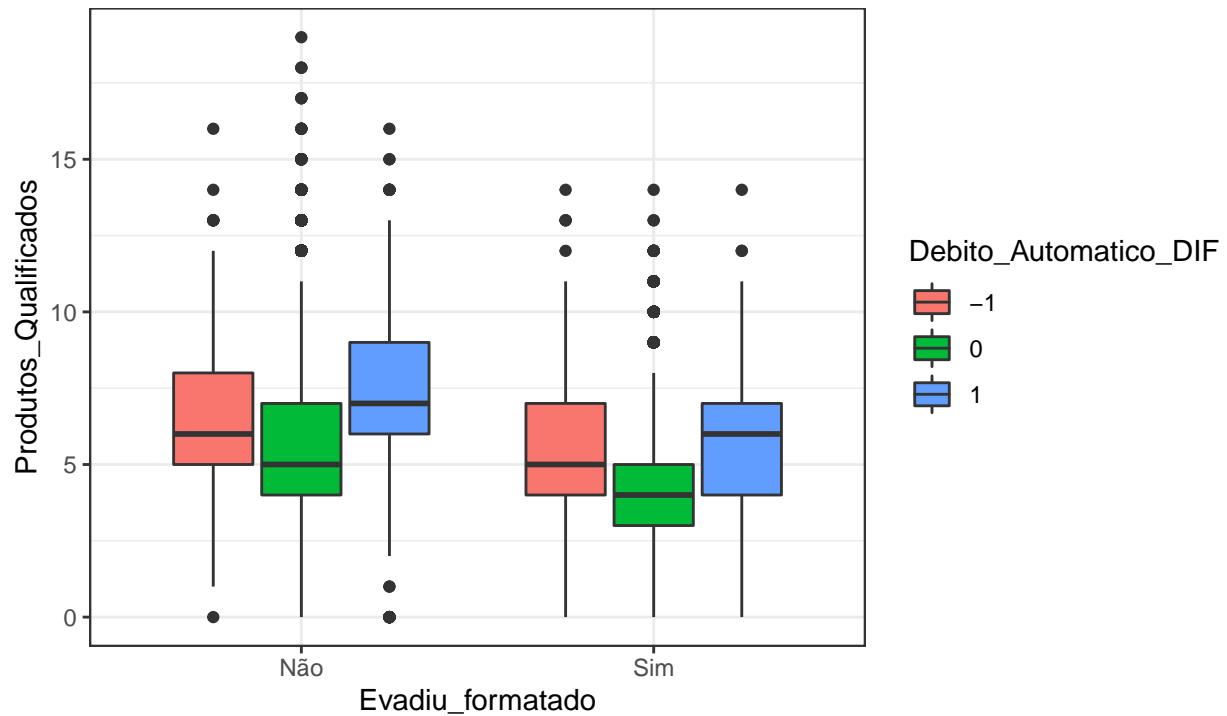
Boxplot

Produtos_Qualificados vs Evadiu_formatado



Boxplot

Produtos_Qualificados vs Evadiu_formatado



```
box_plot(credit, c("winsored_Rentabilidade"), variaveisSubgrupo)
```

