Aprendizagem Estatística em Altas Dimensões [MAE0501/MAE5904/IBI5904]

Ícaro Maia Santos de Castro¹ Rayssa de Carvalho Roberto² Rodrigo Aoyama Nakahara³ Rodrigo Araujo⁴ Vitor Hugo Vieira de Lima⁵

Novembro de 2020

Sumário

nportando os dados / Limpando / Inspecionando	2
Renomear coluna e ajustar níveis de categórica	2
Visualização dos Dados	3
nalise dos Missings	
Descritiva inicial	3
Possibilidades de imputação	
${f rain}$ / ${f test}$	16
Proporção de diabetes (Benign / Malignant)	
nálise Descritiva	16
Distribuição da variável Diabetes	16
Correlação entre cada variável	17
$oxed{ ext{fodelagem}}$	19
SVM	
RandomForest	22

 $^{^1\}mathrm{N\'umero}$ USP: 11866921

 $^{^2\}mathrm{N\'umero}$ USP: 10940828

 $^{^3\}mathrm{N\'umero}$ USP: 3510922

⁴Número USP: 3510922

⁵Número USP: 10263886

Importando os dados / Limpando / Inspecionando

```
diabetes <- read_csv("C:\\Users\\Rodrigo Araujo\\Documents\\IME-USP\\Aprendizagem Estatística em Altas
## Parsed with column specification:
## cols(
##
    Pregnancies = col_double(),
##
    Glucose = col_double(),
    BloodPressure = col_double(),
##
    SkinThickness = col_double(),
    Insulin = col_double(),
##
    BMI = col_double(),
##
    DiabetesPedigreeFunction = col_double(),
##
    Age = col_double(),
##
    Outcome = col_double()
## )
#diabetes <- read.csv("diabetes.csv")</pre>
```

Tabela 1: Dados.

Pregnancies	Glucose	${\bf BloodPressure}$	SkinThickness	Insulin	BMI	${\bf Diabetes Pedigree Function}$	Age	Outcome
6	148	72	35	0	33,6	0,63	50	1
1	85	66	29	0	26,6	0,35	31	0
8	183	64	0	0	23,3	0,67	32	1
1	89	66	23	94	28,1	0,17	21	0
0	137	40	35	168	43,1	2,29	33	1
5	116	74	0	0	25,6	0,20	30	0

```
dim(diabetes)
```

```
Renomear coluna e ajustar níveis de categórica
```

[1] 768 9

head(diabetes) %>% kable(caption="Dados.")

```
diabetes[, 2:6][diabetes[, 2:6] == 0] <- NA
colnames(diabetes)[9] <- "diabetes"
diabetes$diabetes <- as.factor(diabetes$diabetes)
levels(diabetes$diabetes) <- c("No","Yes")</pre>
```

Visualização dos Dados

Estrutura dos Dados

```
str(diabetes)
## tibble [768 x 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##
   $ Pregnancies
                              : num [1:768] 6 1 8 1 0 5 3 10 2 8 ...
   $ Glucose
                              : num [1:768] 148 85 183 89 137 116 78 115 197 125 ...
   $ BloodPressure
                              : num [1:768] 72 66 64 66 40 74 50 NA 70 96 ...
## $ SkinThickness
                              : num [1:768] 35 29 NA 23 35 NA 32 NA 45 NA ...
## $ Insulin
                              : num [1:768] NA NA NA 94 168 NA 88 NA 543 NA ...
                              : num [1:768] 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 NA ...
## $ DiabetesPedigreeFunction: num [1:768] 0.627 0.351 0.672 0.167 2.288 ...
                              : num [1:768] 50 31 32 21 33 30 26 29 53 54 ...
## $ Age
                              : Factor w/ 2 levels "No", "Yes": 2 1 2 1 2 1 2 1 2 2 ...
##
   $ diabetes
##
   - attr(*, "spec")=
     .. cols(
##
##
          Pregnancies = col_double(),
##
          Glucose = col double(),
##
          BloodPressure = col_double(),
     . .
##
         SkinThickness = col_double(),
     . .
##
         Insulin = col_double(),
         BMI = col_double(),
##
     . .
##
         DiabetesPedigreeFunction = col_double(),
##
         Age = col_double(),
     . .
          Outcome = col_double()
##
##
     ..)
```

Analise dos Missings

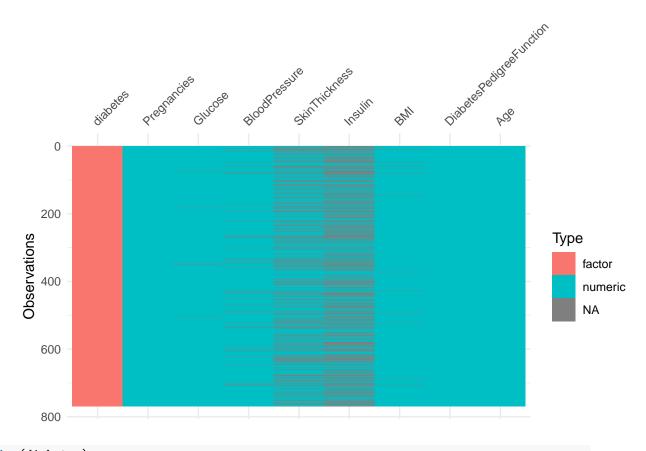
Descritiva inicial

```
summary(diabetes)
```

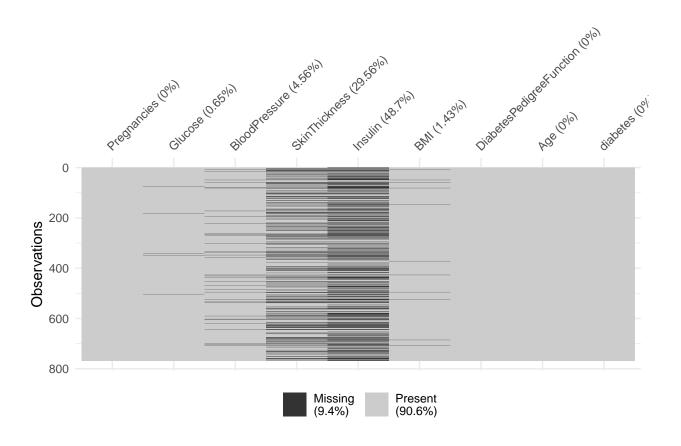
```
BloodPressure
##
    Pregnancies
                       Glucose
                                                     SkinThickness
##
   Min.
          : 0.000
                           : 44.0
                                    Min.
                                          : 24.00
                                                     Min.
                                                           : 7.00
                    Min.
   1st Qu.: 1.000
                    1st Qu.: 99.0
                                    1st Qu.: 64.00
                                                     1st Qu.:22.00
  Median : 3.000
                    Median :117.0
                                    Median : 72.00
                                                     Median :29.00
##
   Mean : 3.845
                    Mean
                          :121.7
                                    Mean
                                           : 72.41
                                                     Mean
                                                            :29.15
   3rd Qu.: 6.000
                                    3rd Qu.: 80.00
##
                    3rd Qu.:141.0
                                                     3rd Qu.:36.00
##
   Max.
          :17.000
                           :199.0
                                    Max.
                                           :122.00
                                                     Max.
                                                            :99.00
                    Max.
##
                    NA's
                           :5
                                    NA's
                                           :35
                                                     NA's
                                                            :227
##
      Insulin
                         BMI
                                    DiabetesPedigreeFunction
                                                                  Age
##
  Min. : 14.00
                    Min. :18.20
                                    Min.
                                           :0.0780
                                                             Min.
                                                                    :21.00
  1st Qu.: 76.25
                    1st Qu.:27.50
                                    1st Qu.:0.2437
                                                            1st Qu.:24.00
## Median :125.00
                    Median :32.30
                                    Median :0.3725
                                                            Median :29.00
```

```
##
    Mean
            :155.55
                      Mean
                              :32.46
                                        Mean
                                               :0.4719
                                                                   Mean
                                                                          :33.24
##
    3rd Qu.:190.00
                      3rd Qu.:36.60
                                        3rd Qu.:0.6262
                                                                   3rd Qu.:41.00
                                                                          :81.00
##
    Max.
            :846.00
                      Max.
                              :67.10
                                        Max.
                                               :2.4200
                                                                   Max.
##
    NA's
            :374
                      NA's
                              :11
    diabetes
##
    No :500
##
    Yes:268
##
##
##
##
##
##
```

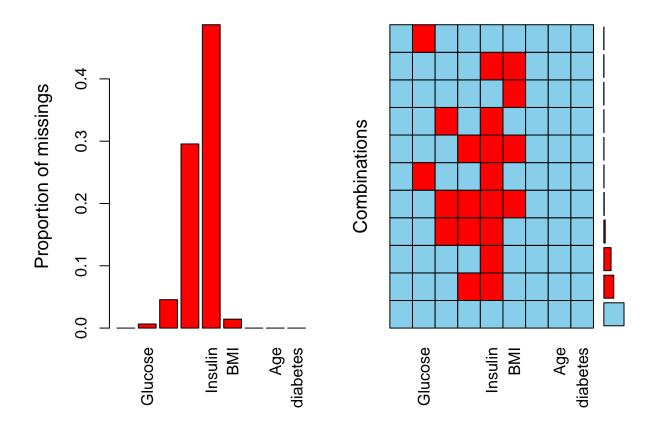
vis_dat(diabetes)



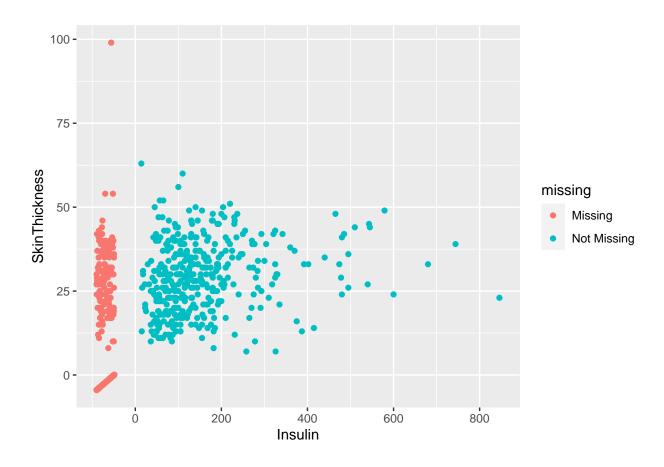
vis_miss(diabetes)



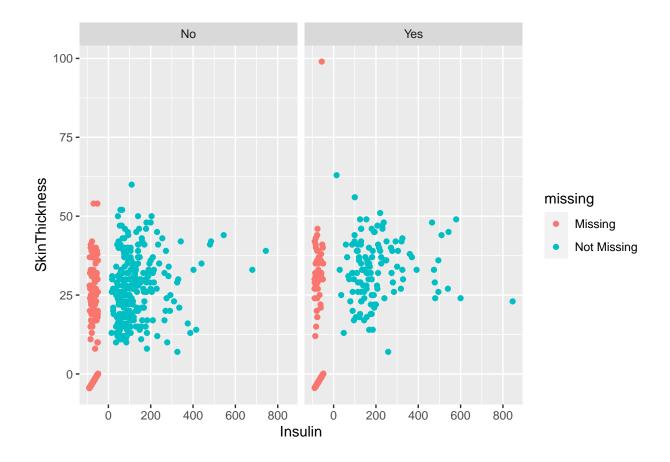
aggr(diabetes) # Missings têm padrões?



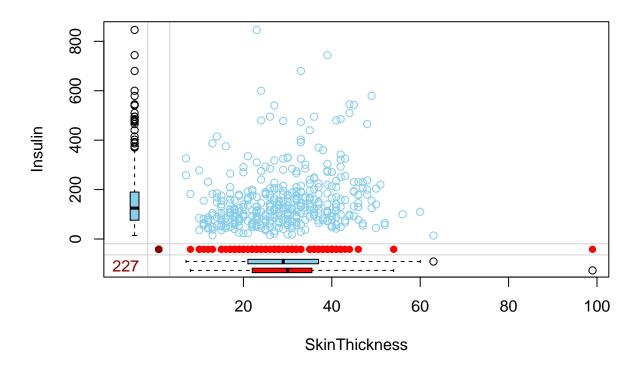
ggplot(diabetes, aes(x = Insulin, y = SkinThickness)) + # Padrão de missing entre 2 vars com mais missi
geom_miss_point()



ggplot(diabetes, aes(x = Insulin, y = SkinThickness)) + # Padrão de missing entre 2 vars com mais missi
geom_miss_point() + facet_wrap(~ diabetes)



marginplot(diabetes[c(4,5)])



Possibilidades de imputação

```
library(mice)
## Warning: package 'mice' was built under R version 4.0.3
## Attaching package: 'mice'
## The following objects are masked from 'package:base':
##
##
       cbind, rbind
imputacao1 <- mice(data = diabetes , m = 5, maxit = 50, meth = 'pmm', seed = 25)</pre>
##
##
    iter imp variable
##
         1 Glucose BloodPressure
                                   SkinThickness Insulin BMI
##
         2 Glucose BloodPressure
                                   SkinThickness Insulin BMI
         3 Glucose BloodPressure SkinThickness Insulin BMI
##
```

```
##
                                       SkinThickness
                                                        Insulin
     1
             Glucose
                       BloodPressure
          5
##
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
     1
##
     2
          1
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
     2
##
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     2
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     2
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     2
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     3
          1
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     3
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     3
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     3
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     3
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     4
                                                        Insulin
                                                                  BMI
          1
             Glucose
                       BloodPressure
                                       SkinThickness
##
     4
          2
             Glucose
                       BloodPressure
                                        SkinThickness
                                                        Insulin
                                                                  BMI
     4
##
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     4
          4
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
             Glucose
          5
##
     4
                                                        Insulin
                                                                  BMI
             Glucose
                       BloodPressure
                                       SkinThickness
##
     5
                       BloodPressure
          1
             Glucose
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     5
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     5
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     5
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     5
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
##
     6
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
          1
##
     6
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     6
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     6
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     6
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
     7
##
          1
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
     7
          2
##
             Glucose
                       BloodPressure
                                        SkinThickness
                                                        Insulin
                                                                  BMI
##
     7
          3
             Glucose
                                                        Insulin
                       BloodPressure
                                       SkinThickness
                                                                  BMI
##
     7
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     7
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     8
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     8
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     8
          3
                       BloodPressure
                                                        Insulin
                                                                  BMI
             Glucose
                                       SkinThickness
##
     8
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMI
##
     8
          5
             Glucose
                       BloodPressure
                                        SkinThickness
                                                        Insulin
##
     9
                                                        Insulin
          1
             Glucose
                       BloodPressure
                                       SkinThickness
                                                                  BMT
          2
##
     9
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     9
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
##
     9
             Glucose
                       BloodPressure
                                       SkinThickness
                                                        Insulin
                                                                  BMT
     9
                                                        Insulin
##
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                                  BMT
##
     10
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
                                                         Insulin
     10
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     10
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     10
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     10
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     11
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     11
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     11
           3
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
              Glucose
##
           4
                        BloodPressure
                                        SkinThickness
                                                         Insulin
     11
              Glucose
                                                                   BMI
##
     11
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     12
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
           1
                                                                   BMI
##
     12
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
```

```
##
     12
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
     12
##
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     12
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
     13
##
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMT
##
     13
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
                        BloodPressure
##
     13
           3
                                        SkinThickness
                                                         Insulin
              Glucose
                                                                  BMI
##
     13
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     13
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMT
##
     14
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     14
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     14
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     14
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     14
           5
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     15
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     15
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     15
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     15
           4
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     15
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
                        BloodPressure
                                        SkinThickness
                                                         Insulin
     16
           1
              Glucose
                                                                  BMT
##
     16
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     16
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     16
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
                        BloodPressure
                                        SkinThickness
                                                         Insulin
     16
           5
              Glucose
                                                                  BMI
##
     17
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMT
##
     17
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMT
##
     17
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMI
##
     17
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     17
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     18
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     18
           2
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     18
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     18
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     18
              Glucose
                        BloodPressure
                                                         Insulin
                                        SkinThickness
                                                                   BMI
##
     19
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
           1
##
     19
           2
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     19
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     19
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
                                                         Insulin
     19
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMT
##
     20
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     20
           2
              Glucose
                        BloodPressure
                                                         Insulin
                                        SkinThickness
                                                                  BMI
##
     20
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
     20
##
           4
              Glucose
                       BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                  BMT
##
     20
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
                                                         Insulin
     21
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     21
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     21
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     21
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     21
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
     22
##
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
           1
##
     22
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     22
           3
                        BloodPressure
                                        SkinThickness
                                                         Insulin
              Glucose
                                                                   BMI
##
     22
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     22
                       BloodPressure
                                        SkinThickness
                                                         Insulin
           5
              Glucose
                                                                  BMI
##
     23
                       BloodPressure
                                        SkinThickness
              Glucose
                                                         Insulin
```

```
##
     23
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                        BloodPressure
##
     23
           3
                                        SkinThickness
                                                         Insulin
                                                                   BMT
              Glucose
##
     23
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
     23
##
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     24
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
                        BloodPressure
##
           2
                                        SkinThickness
                                                         Insulin
     24
              Glucose
                                                                   BMI
##
     24
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     24
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     24
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     25
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     25
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     25
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     25
           4
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     25
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
     26
##
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     26
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
                                                         Insulin
     26
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     26
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
                        BloodPressure
                                        SkinThickness
##
     26
                                                         Insulin
           5
              Glucose
                                                                   BMT
##
     27
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     27
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     27
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     27
                        BloodPressure
                                        SkinThickness
                                                         Insulin
           4
              Glucose
                                                                   BMI
##
     27
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     28
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     28
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     28
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     28
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     28
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     29
                                                         Insulin
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     29
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     29
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     29
                        BloodPressure
                                                         Insulin
              Glucose
                                        SkinThickness
                                                                   BMI
##
     29
                                                         Insulin
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
     30
                                                         Insulin
##
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     30
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     30
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     30
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     30
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
                        BloodPressure
                                                         Insulin
     31
           1
              Glucose
                                        SkinThickness
                                                                   BMI
##
     31
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     31
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     31
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
           5
     31
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     32
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
           1
                                                                   BMI
           2
##
     32
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     32
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     32
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
     32
##
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     33
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
           1
              Glucose
##
     33
           2
                                        SkinThickness
                                                         Insulin
              Glucose
                        BloodPressure
                                                                   BMI
##
     33
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     33
                        BloodPressure
                                        SkinThickness
                                                         Insulin
           4
              Glucose
                                                                   BMI
##
     33
                        BloodPressure
                                        SkinThickness
              Glucose
                                                         Insulin
```

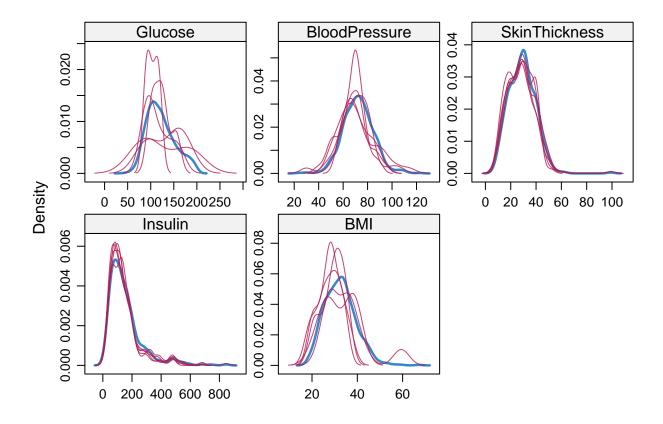
```
##
     34
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                        BloodPressure
##
     34
           2
              Glucose
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     34
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
     34
##
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     34
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
                        BloodPressure
##
     35
                                        SkinThickness
                                                         Insulin
           1
              Glucose
                                                                   BMI
##
     35
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     35
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     35
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     35
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     36
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
           1
##
     36
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     36
           3
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     36
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     36
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     37
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
           1
                                                                   BMI
##
     37
           2
                                                         Insulin
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     37
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
                        BloodPressure
                                        SkinThickness
##
     37
                                                         Insulin
           4
              Glucose
                                                                   BMT
##
     37
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     38
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     38
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     38
           3
                        BloodPressure
                                        SkinThickness
                                                         Insulin
              Glucose
                                                                   BMI
##
     38
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     38
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     39
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     39
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     39
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     39
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     39
                                                         Insulin
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     40
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     40
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     40
           3
                        BloodPressure
                                                         Insulin
              Glucose
                                        SkinThickness
                                                                   BMI
##
     40
                                                         Insulin
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
                                                         Insulin
##
     40
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     41
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     41
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
##
     41
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     41
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
           5
                        BloodPressure
                                                         Insulin
     41
              Glucose
                                        SkinThickness
                                                                   BMI
##
     42
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     42
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     42
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
                                                         Insulin
     42
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                                   BMI
##
     42
           5
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     43
           1
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     43
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     43
           3
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     43
           4
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMT
##
     43
           5
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
              Glucose
##
     44
                                        SkinThickness
                                                         Insulin
           1
              Glucose
                        BloodPressure
                                                                   BMI
##
     44
           2
              Glucose
                        BloodPressure
                                        SkinThickness
                                                         Insulin
                                                                   BMI
##
     44
           3
                        BloodPressure
                                        SkinThickness
                                                         Insulin
              Glucose
                                                                   BMI
##
     44
                        BloodPressure
                                        SkinThickness
              Glucose
                                                         Insulin
```

```
##
     44
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
     45
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMT
          1
             Glucose
                                       SkinThickness
##
     45
             Glucose
                       BloodPressure
                                                       Insulin
                                                                BMI
##
     45
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMI
##
     45
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
     45
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
                       BloodPressure
                                       SkinThickness
                                                      Insulin
##
     46
             Glucose
##
     46
          2
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
                                                                BMT
##
     46
          3
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
##
     46
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMI
##
     46
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMI
##
     47
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMI
##
     47
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
                                                                BMI
                                       SkinThickness
##
     47
             Glucose
                       BloodPressure
                                                       Insulin
                                                                BMI
##
     47
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
          4
                                                                BMT
##
     47
          5
             Glucose
                       BloodPressure
                                       SkinThickness
                                                       Insulin
                                                                BMI
##
     48
                                                      Insulin
          1
             Glucose
                       BloodPressure
                                       SkinThickness
##
     48
             Glucose
                       BloodPressure
                                       SkinThickness
                                                       Insulin
##
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
     48
##
     48
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
     48
          5
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
     49
             Glucose
                       BloodPressure
                                       SkinThickness
                                                       Insulin
##
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
     49
          2
                                                                BMT
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
##
     49
     49
##
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
     49
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
##
     50
             Glucose
                       BloodPressure
                                      SkinThickness
                                                      Insulin
                                                                BMI
                                                      Insulin
##
     50
             Glucose
                       BloodPressure
                                      SkinThickness
##
     50
          3
             Glucose
                       BloodPressure
                                      SkinThickness
                                                       Insulin
                                                                BMI
##
     50
          4
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
                                                                BMI
##
     50
             Glucose
                       BloodPressure
                                       SkinThickness
                                                      Insulin
                                                                BMI
```

summary(imputacao1)

```
## Class: mids
   Number of multiple imputations: 5
   Imputation methods:
##
                 Pregnancies
                                                 Glucose
                                                                      BloodPressure
##
                                                    "pmm"
                                                                                "pmm"
##
               SkinThickness
                                                  Insulin
                                                                                 BMI
                        "pmm"
                                                    "pmm"
                                                                               "pmm"
## DiabetesPedigreeFunction
                                                                            diabetes
                                                      Age
##
##
  PredictorMatrix:
##
                  Pregnancies Glucose BloodPressure SkinThickness Insulin BMI
  Pregnancies
##
                             0
                                                                              1
                                                                                   1
                                      1
                                                      1
                                                                     1
   Glucose
                             1
                                      0
                                                      1
                                                                     1
                                                                              1
                                                                                   1
   BloodPressure
                             1
                                                      0
                                                                              1
                                                                                   1
                                      1
                                                                     1
  SkinThickness
                             1
                                                                     0
                                                                              1
                                      1
                                                      1
                                                                                   1
## Insulin
                             1
                                      1
                                                      1
                                                                     1
                                                                              0
                                                                                   1
##
  BMI
                              1
                                      1
                                                      1
##
                  DiabetesPedigreeFunction Age diabetes
## Pregnancies
                                            1
                                                1
```

```
## Glucose
                                             1
## BloodPressure
                                                  1
                                                            1
## SkinThickness
                                                  1
## Insulin
                                                            1
                                                  1
## BMI
imputado1 <- complete(imputacao1, 1)</pre>
imputado2 <- complete(imputacao1, 2)</pre>
imputado3 <- complete(imputacao1, 3)</pre>
imputado4 <- complete(imputacao1, 4)</pre>
imputado5 <- complete(imputacao1, 5)</pre>
library(lattice)
densityplot(imputacao1)
```



diabetes\$Glucose <- apply(cbind(imputado1\$Glucose, imputado2\$Glucose, imputado3\$Glucose, imputado4\$Gluc
diabetes\$BloodPressure <- apply(cbind(imputado1\$BloodPressure, imputado2\$BloodPressure, imputado3\$Blood
diabetes\$SkinThickness <- apply(cbind(imputado1\$SkinThickness, imputado2\$SkinThickness, imputado3\$SkinThickness, imputado3\$SkinThickness, imputado3\$SkinThickness, imputado3\$Insulin, imputado3\$Insulin, imputado4\$Insu
diabetes\$BMI <- apply(cbind(imputado1\$BMI, imputado2\$BMI, imputado3\$BMI, imputado4\$BMI, imputado5\$BMI),

train / test

```
# para reprodução
set.seed(23)
nrows <- nrow(diabetes)
index <- sample(1:nrows, 0.7 * nrows) # shuffle and divide
# train <- diab # 768 test data (100%)
train <- diabetes[index,] # 537 test data (70%)
test <- diabetes[-index,] # 231 test data (30%)</pre>
```

Proporção de diabetes (Benign / Malignant)

 train

```
prop.table(table(train$diabetes))

##
## No Yes
## 0.6405959 0.3594041

test

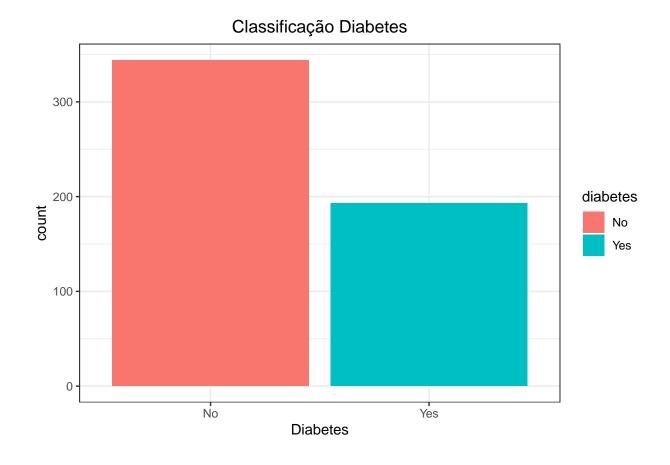
prop.table(table(test$diabetes))

##
## No Yes
## 0.6753247 0.3246753
```

Análise Descritiva

Distribuição da variável Diabetes

```
ggplot(train, aes(diabetes, fill = diabetes)) +
  geom_bar() +
  theme_bw() +
  labs(title = "Classificação Diabetes", x = "Diabetes") +
  theme(plot.title = element_text(hjust = 0.5))
```



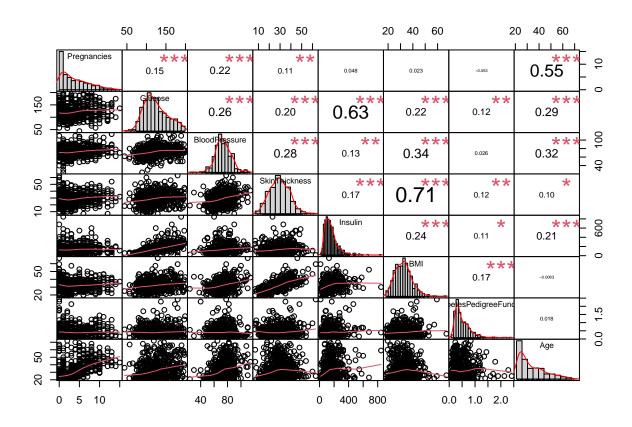
Correlação entre cada variável

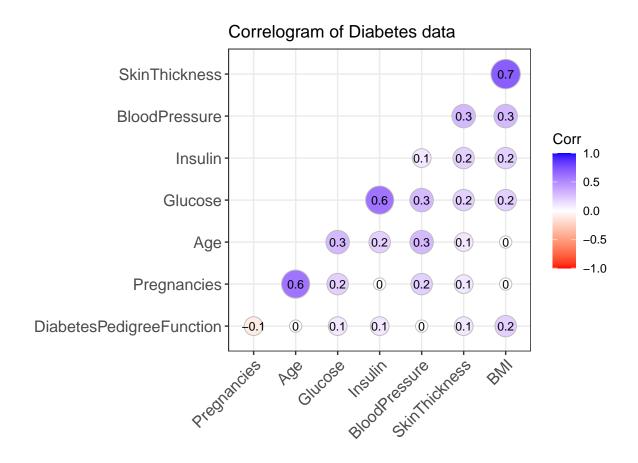
```
library(PerformanceAnalytics)
## Loading required package: xts
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
## as.Date, as.Date.numeric
##
## Attaching package: 'xts'
## The following objects are masked from 'package:dplyr':
##
## first, last
```

```
##
## Attaching package: 'PerformanceAnalytics'
## The following object is masked from 'package:graphics':
##
## legend
```

chart.Correlation(train[,-9], histogram=TRUE, col="grey10", pch=1, main="Correlação entre ás variáveis





${\bf Modelagem}$

SVM

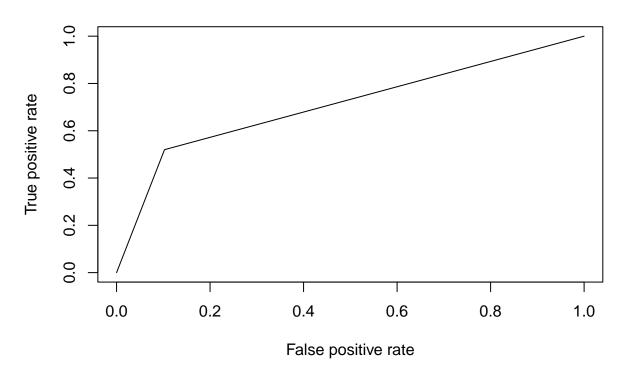
```
##
## Attaching package: 'caret'
## The following object is masked from 'package:vegan':
##
## tolerance
## The following objects are masked from 'package:MLmetrics':
##
## MAE, RMSE
## The following object is masked from 'package:survival':
##
## cluster
```

```
## The following object is masked from 'package:purrr':
##
##
       lift
library(e1071)
##
## Attaching package: 'e1071'
## The following objects are masked from 'package:PerformanceAnalytics':
##
##
       kurtosis, skewness
## The following object is masked from 'package:Hmisc':
##
       impute
set.seed(123)
linear.tune <- e1071::tune.svm(diabetes ~.,</pre>
                                 data = train,
                                 kernel = 'linear',
                                 cost = c(0.001, 0.01, 0.1, 1, 5, 10))
summary(linear.tune)
##
## Parameter tuning of 'svm':
## - sampling method: 10-fold cross validation
##
## - best parameters:
## cost
## 0.01
## - best performance: 0.2309574
##
## - Detailed performance results:
              error dispersion
      cost
## 1 1e-03 0.3591894 0.07461689
## 2 1e-02 0.2309574 0.05370355
## 3 1e-01 0.2310273 0.05114627
## 4 1e+00 0.2365828 0.05083653
## 5 5e+00 0.2347310 0.05025263
## 6 1e+01 0.2347310 0.05025263
        Matriz de Confusão
svm.real <- test$diabetes</pre>
best.linear <- linear.tune$best.model</pre>
```

```
## Confusion Matrix and Statistics
##
##
             Reference
## Prediction No Yes
          No 140 36
##
          Yes 16 39
##
##
##
                  Accuracy : 0.7749
                    95% CI: (0.7155, 0.8271)
##
       No Information Rate: 0.6753
##
       P-Value [Acc > NIR] : 0.000570
##
##
##
                     Kappa: 0.4485
##
##
   Mcnemar's Test P-Value: 0.008418
##
##
               Sensitivity: 0.5200
##
               Specificity: 0.8974
##
            Pos Pred Value: 0.7091
            Neg Pred Value: 0.7955
##
##
                Prevalence: 0.3247
##
            Detection Rate: 0.1688
##
      Detection Prevalence: 0.2381
##
         Balanced Accuracy: 0.7087
##
##
          'Positive' Class : Yes
##
```

Curva ROC

ROC curve



RandomForest

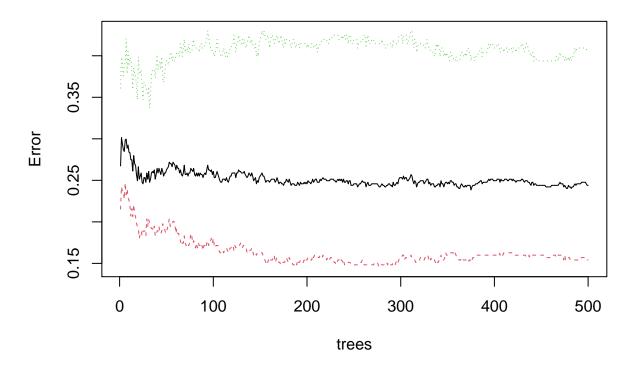
Area under the curve: 0.7087

```
library(randomForest)
## randomForest 4.6-14
```

Type rfNews() to see new features/changes/bug fixes.

```
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:psych':
##
##
       outlier
## The following object is masked from 'package:dplyr':
##
##
       combine
## The following object is masked from 'package:ggplot2':
##
##
       margin
learn_rf <- randomForest(diabetes~., data=train, ntree=500, proximity=T, importance=T)</pre>
pre_rf
         <- predict(learn_rf, test[,-9])
         <- confusionMatrix(pre_rf, test$diabetes)</pre>
cm_rf
cm_rf
## Confusion Matrix and Statistics
##
##
             Reference
## Prediction No Yes
          No 127 30
##
          Yes 29 45
##
##
                  Accuracy: 0.7446
##
                    95% CI: (0.6833, 0.7995)
       No Information Rate: 0.6753
##
##
       P-Value [Acc > NIR] : 0.01342
##
##
                     Kappa: 0.4155
##
   Mcnemar's Test P-Value : 1.00000
##
##
##
               Sensitivity: 0.8141
               Specificity: 0.6000
##
            Pos Pred Value: 0.8089
##
            Neg Pred Value: 0.6081
##
##
                Prevalence: 0.6753
##
            Detection Rate: 0.5498
      Detection Prevalence : 0.6797
##
##
         Balanced Accuracy: 0.7071
##
##
          'Positive' Class : No
##
```

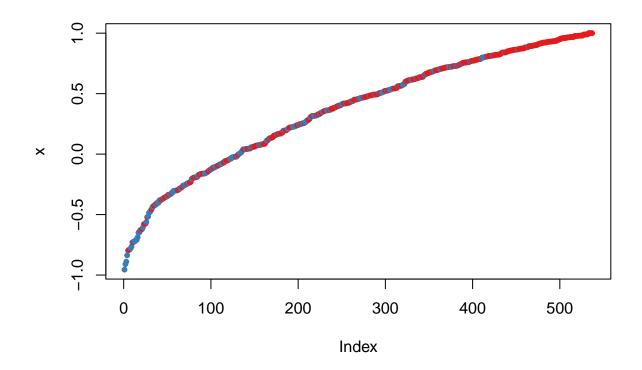
Random Forest (Error Rate vs. Number of Trees)



Prediction Plot

```
plot(margin(learn_rf,test$diabetes))
```

Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested pa



Variance Importance Plot

varImpPlot(learn_rf)

learn_rf

