Trabalho Redes Neurais

Leonardo da Silva Correa

25/05/2021

Pre-Processamento

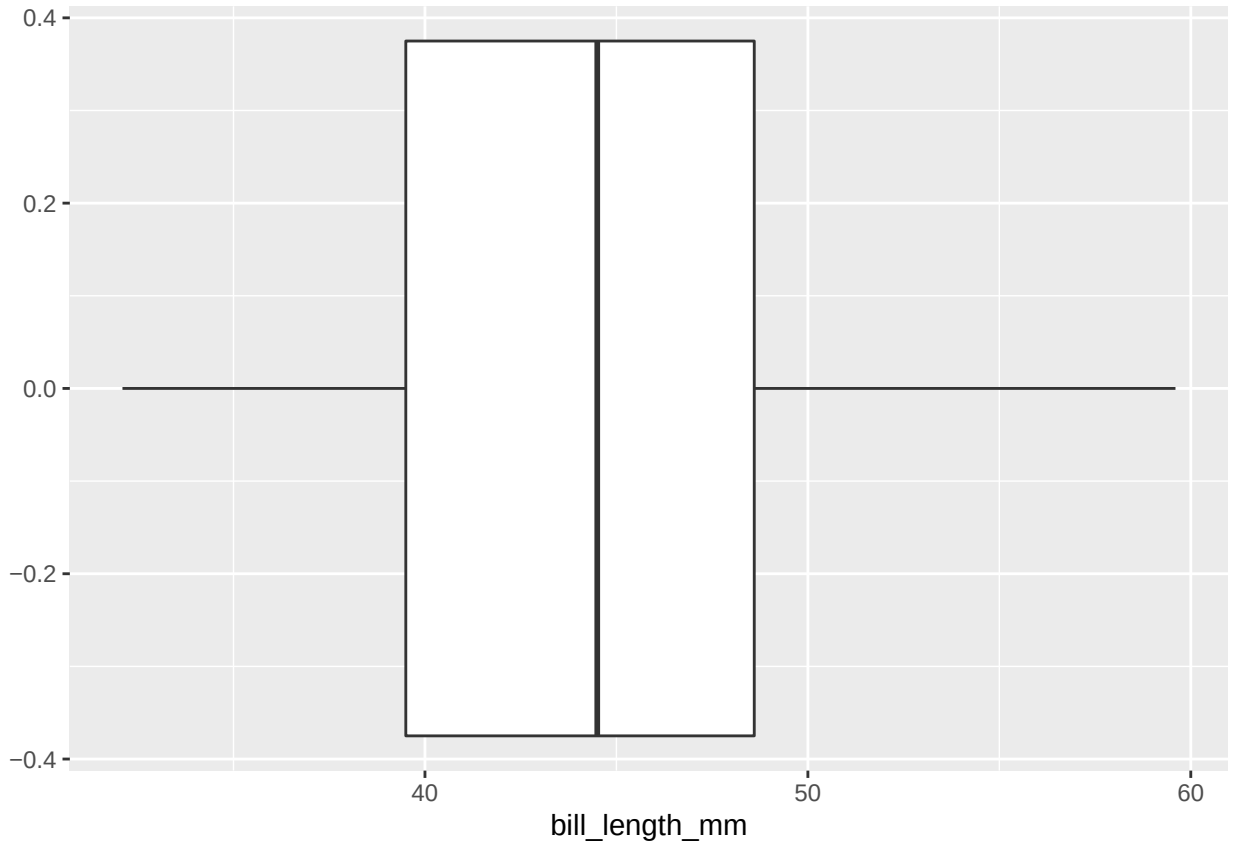
```
summary(df)
```

```
##       species          island    bill_length_mm  bill_depth_mm
##  Adelie   :152   Biscoe   :168   Min.   :32.10   Min.   :13.10
##  Chinstrap: 68   Dream    :124   1st Qu.:39.23   1st Qu.:15.60
##  Gentoo   :124   Torgersen: 52   Median :44.45   Median :17.30
##                                  Mean   :43.92   Mean   :17.15
##                                  3rd Qu.:48.50   3rd Qu.:18.70
##                                  Max.   :59.60   Max.   :21.50
##                                  NA's   :2       NA's   :2
##  flipper_length_mm  body_mass_g       sex           year
##  Min.   :172.0     Min.   :2700   female:165   Min.   :2007
##  1st Qu.:190.0     1st Qu.:3550   male  :168   1st Qu.:2007
##  Median :197.0     Median :4050   NA's  : 11   Median :2008
##  Mean   :200.9     Mean   :4202                Mean   :2008
##  3rd Qu.:213.0     3rd Qu.:4750                3rd Qu.:2009
##  Max.   :231.0     Max.   :6300                Max.   :2009
##  NA's   :2         NA's   :2
```

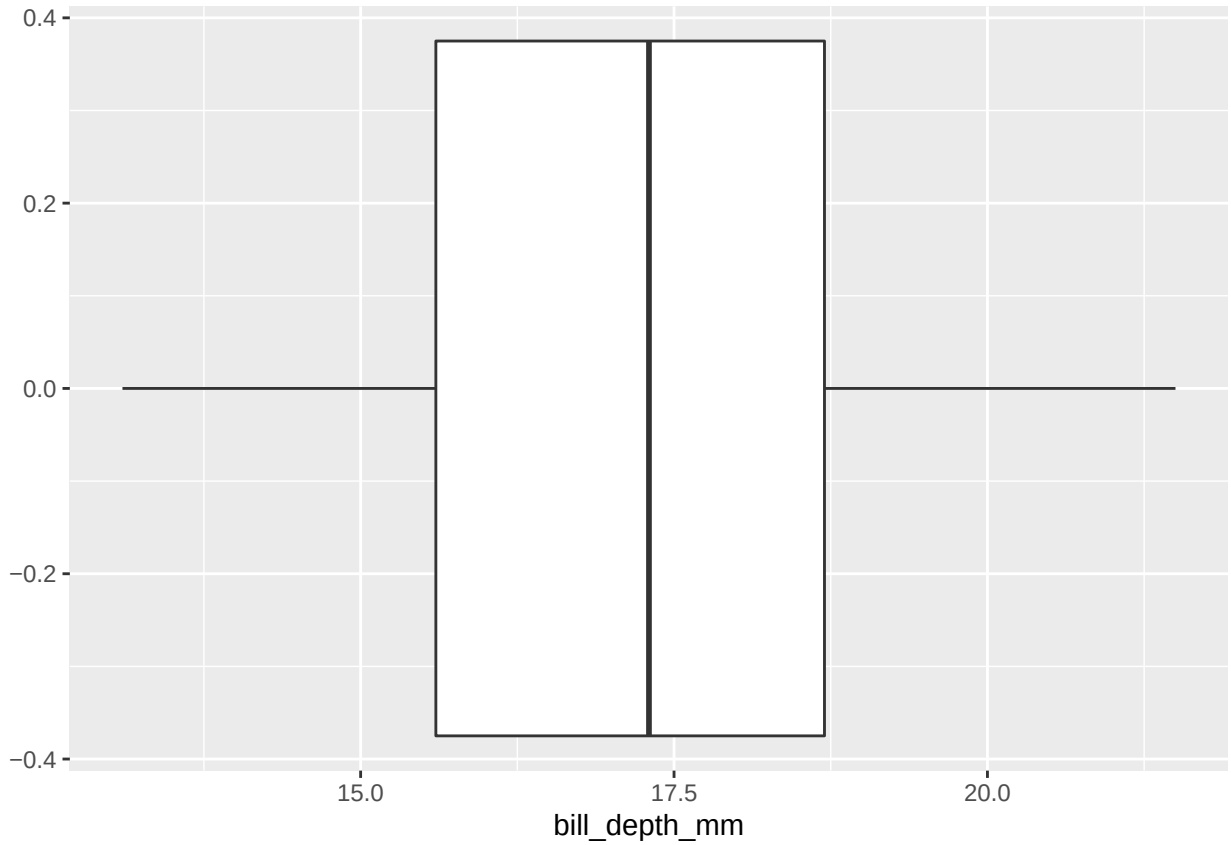
```
#df <- filter(df, !is.na(bill_length_mm))
df <- filter(df, !is.na(sex))
summary(df)
```

```
##       species          island    bill_length_mm  bill_depth_mm
##  Adelie   :146   Biscoe   :163   Min.   :32.10   Min.   :13.10
##  Chinstrap: 68   Dream    :123   1st Qu.:39.50   1st Qu.:15.60
##  Gentoo   :119   Torgersen: 47   Median :44.50   Median :17.30
##                                  Mean   :43.99   Mean   :17.16
##                                  3rd Qu.:48.60   3rd Qu.:18.70
##                                  Max.   :59.60   Max.   :21.50
##  flipper_length_mm  body_mass_g       sex           year
##  Min.   :172       Min.   :2700   female:165   Min.   :2007
##  1st Qu.:190       1st Qu.:3550   male  :168   1st Qu.:2007
##  Median :197       Median :4050                Median :2008
##  Mean   :201       Mean   :4207                Mean   :2008
##  3rd Qu.:213       3rd Qu.:4775                3rd Qu.:2009
##  Max.   :231       Max.   :6300                Max.   :2009
```

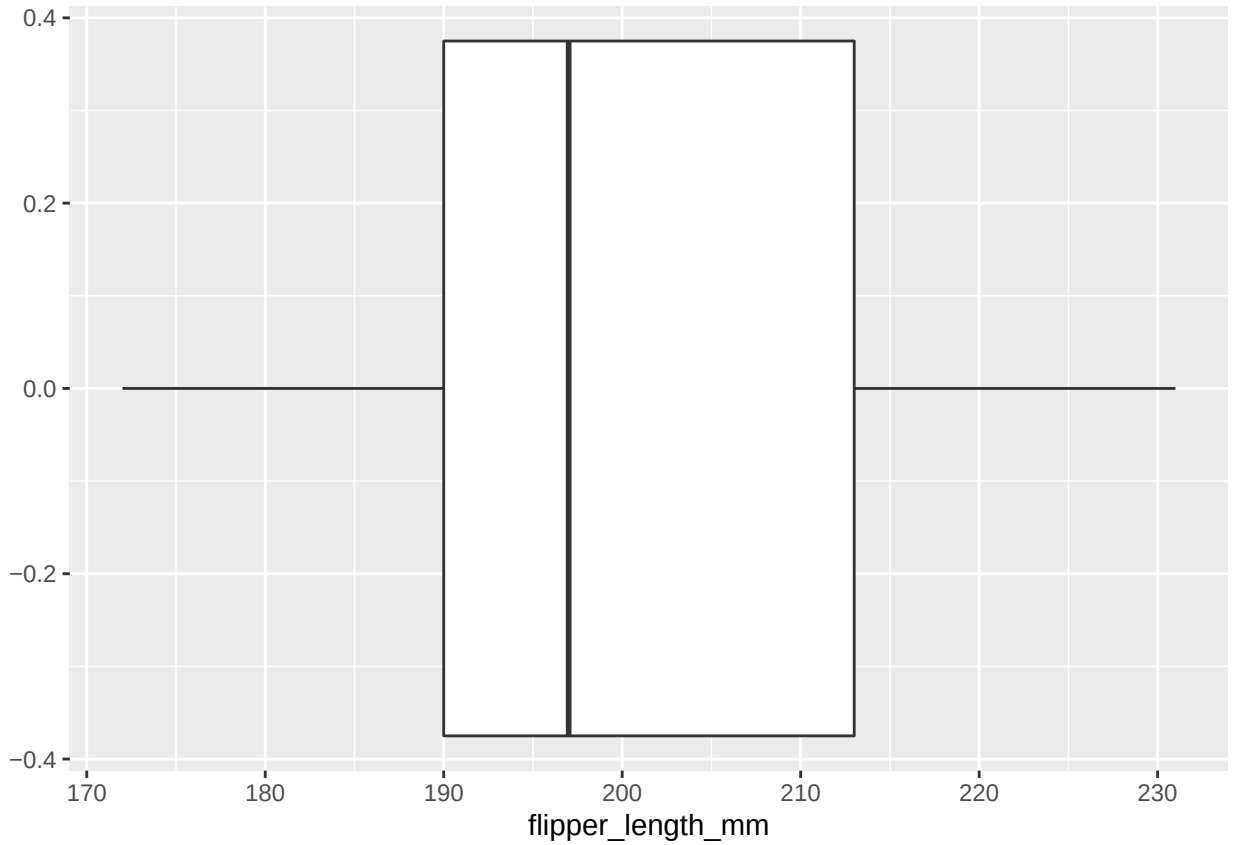
```
df <- subset( df, select = -sex )
#verificando Outliers
ggplot(select(df, bill_length_mm), aes(bill_length_mm)) + geom_boxplot()
```



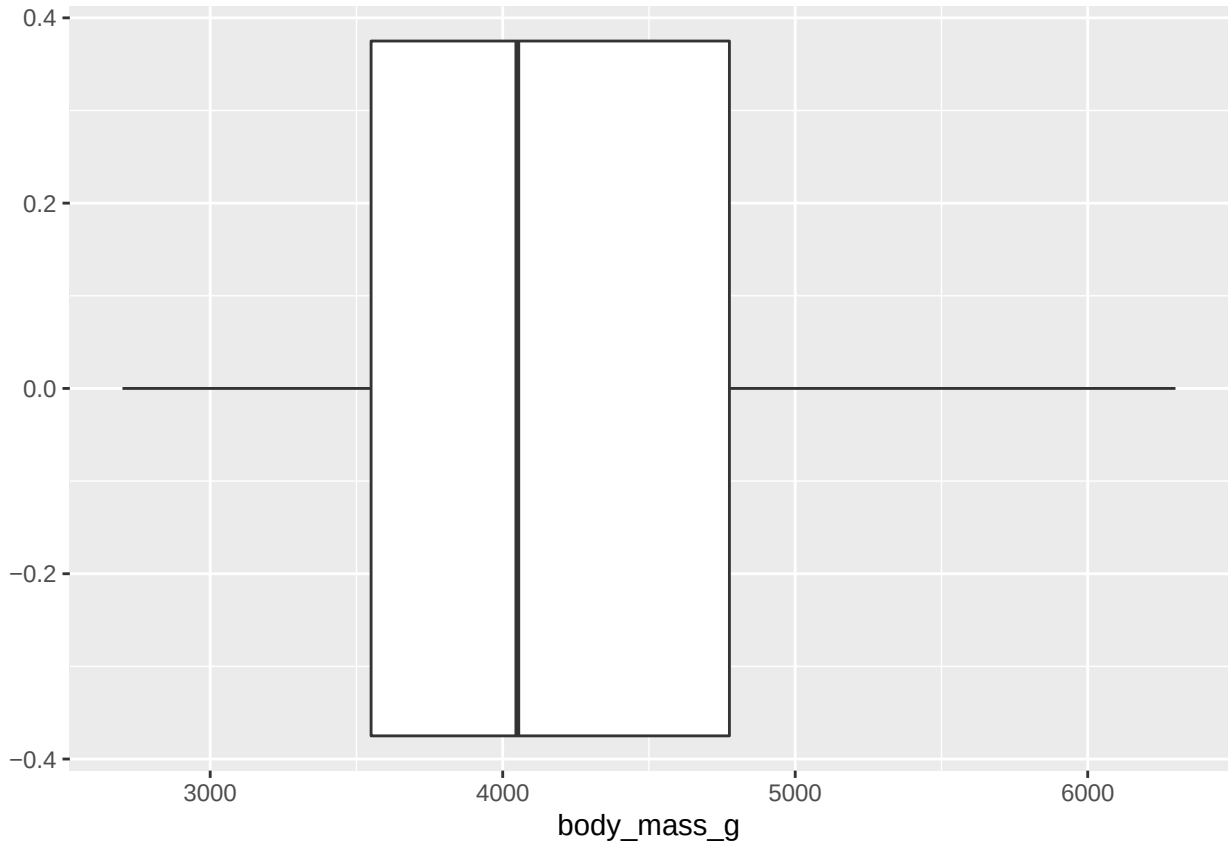

```
ggplot(select(df, bill_depth_mm), aes(bill_depth_mm)) + geom_boxplot()
```



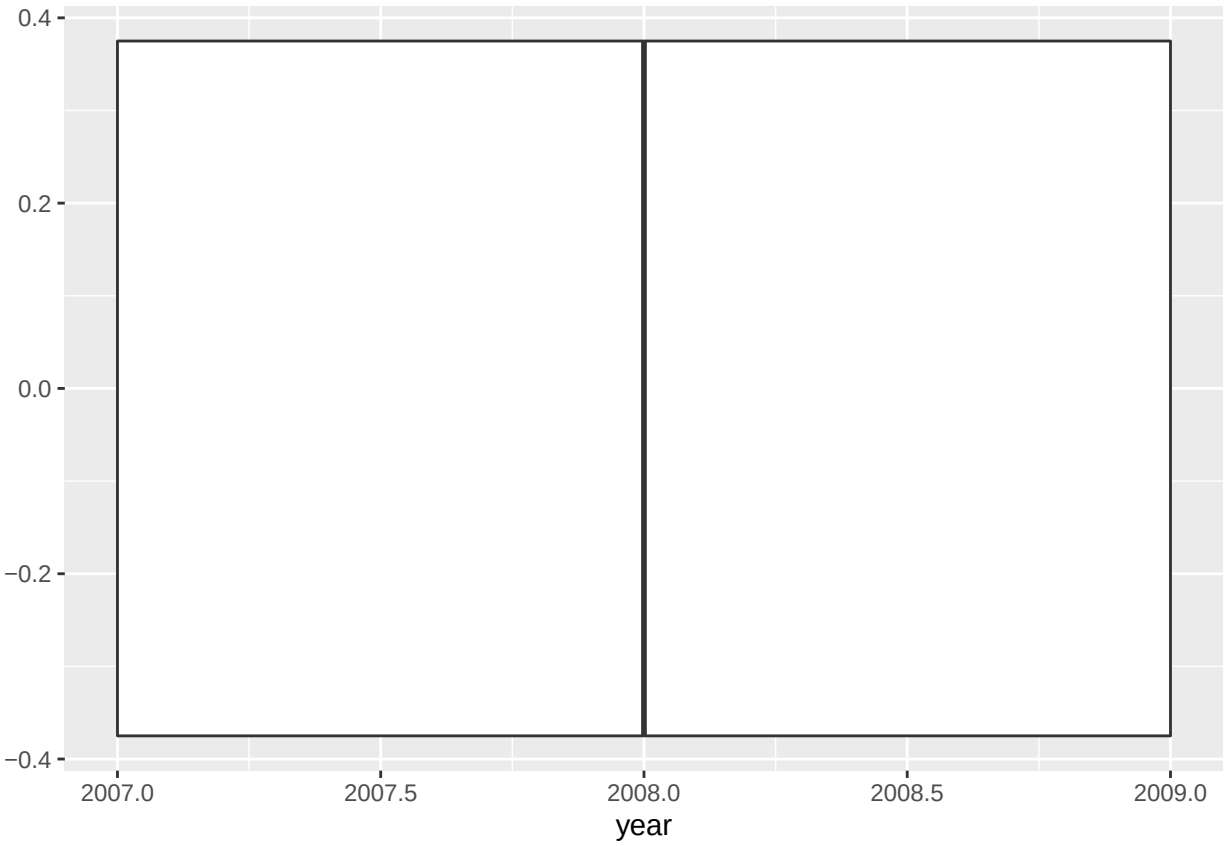

```
ggplot(select(df, flipper_length_mm), aes(flipper_length_mm)) + geom_boxplot()
```




```
ggplot(select(df, body_mass_g), aes(body_mass_g)) + geom_boxplot()
```




```
ggplot(select(df, year), aes(year)) + geom_boxplot()
```



Criando dataframes de treino e teste

```
train_idx <- sample(nrow(df), 2/3 * nrow(df))
df_train <- df[train_idx, ]
df_test <- df[-train_idx, ]
```

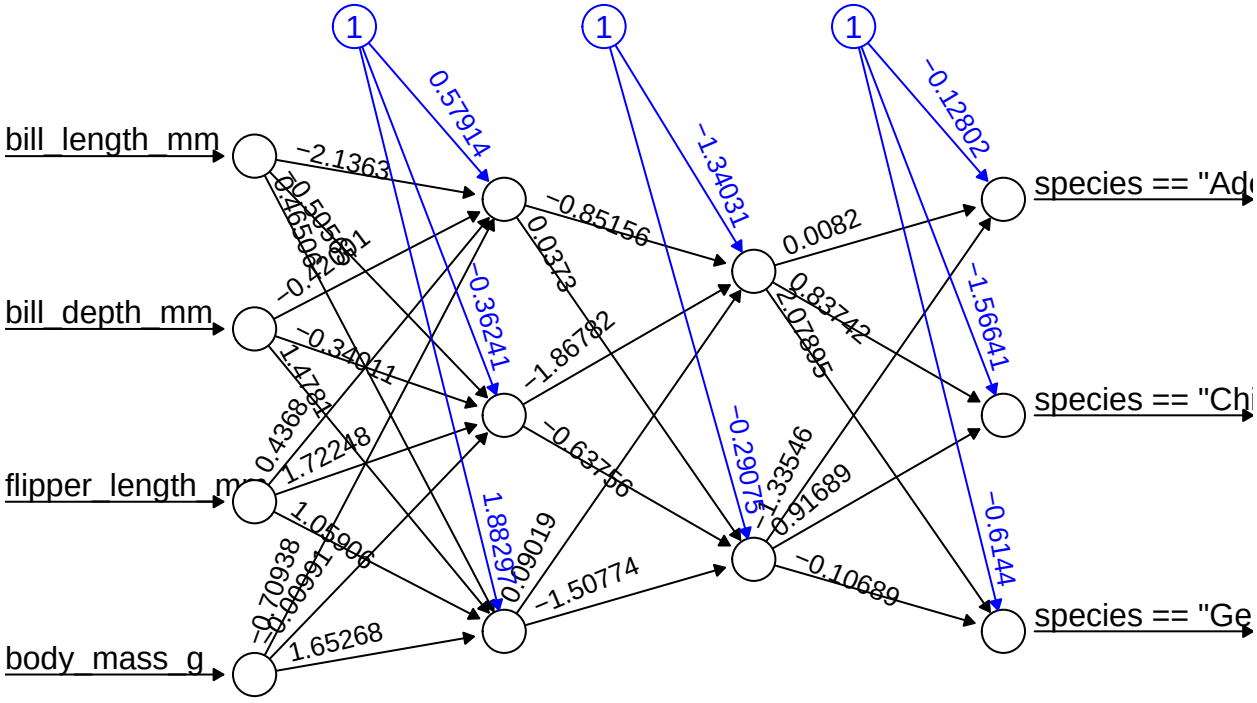
Criando Redes Neurais

```
nn <- neuralnet((species == "Adelie") +
                  (species == "Chinstrap") +
                  (species == "Gentoo") ~
                   bill_length_mm +
                   bill_depth_mm+flipper_length_mm+body_mass_g,
                 df_train,
                 hidden = c(3,2),
                 threshold = 0.01,
                 stepmax = 1e+05,
                 learningrate=0.1,
                 algorithm = "backprop",
                 linear.output = FALSE)
pred <- predict(nn, df_test)
a<-apply(pred, 1, which.max)
a[a==1]<-"Adelie"
a[a==2]<-"Chinstrap"
```

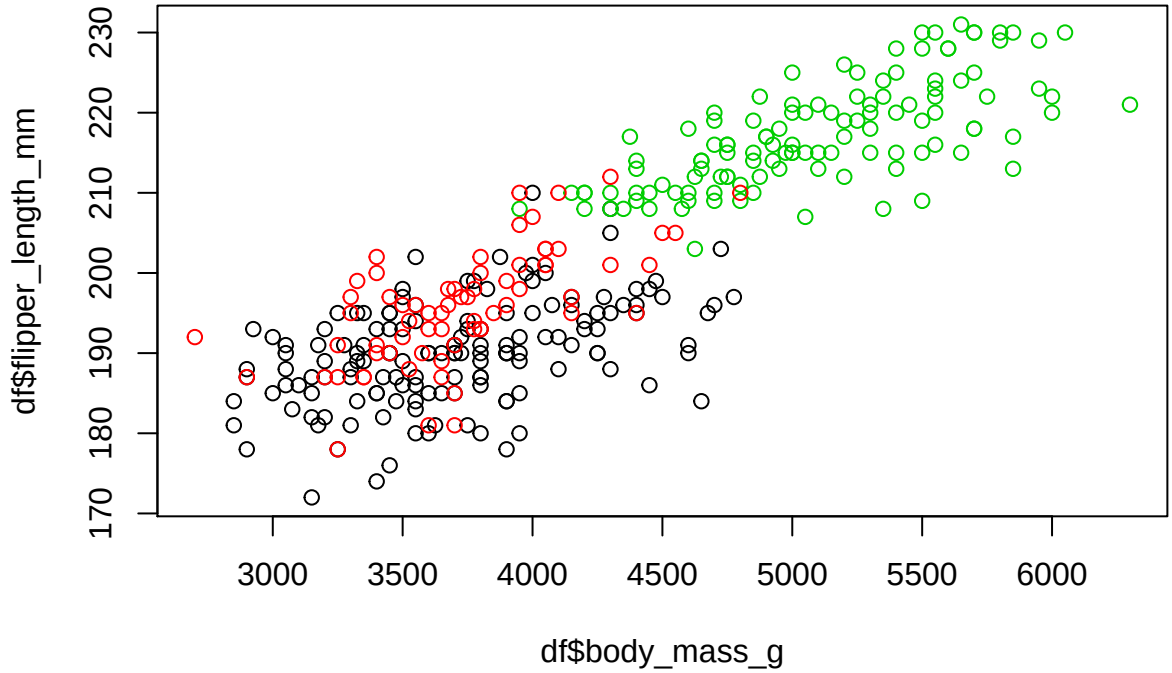
```r
a[a==3]<-"Gentoo"
a<-factor(a,levels = c("Adelie","Chinstrap","Gentoo"))
result<-table(df_test$species,a)
confusionMatrix(result)
```

```
## Confusion Matrix and Statistics
##
##            a
##             Adelie Chinstrap Gentoo
##   Adelie        48         0      0
##   Chinstrap     26         0      0
##   Gentoo        37         0      0
##
## Overall Statistics
##
##                Accuracy : 0.4324
##                  95% CI : (0.3387, 0.5298)
##     No Information Rate : 1
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Adelie Class: Chinstrap Class: Gentoo
## Sensitivity                 0.4324              NA            NA
## Specificity                     NA          0.7658        0.6667
## Pos Pred Value                  NA              NA            NA
## Neg Pred Value                  NA              NA            NA
## Prevalence                  1.0000          0.0000        0.0000
## Detection Rate              0.4324          0.0000        0.0000
## Detection Prevalence        0.4324          0.2342        0.3333
## Balanced Accuracy               NA              NA            NA
```

```r
plot(nn,rep = "best")
```




```
plot(df$body_mass_g, df$flipper_length_mm, col=df$species)
```



Quando observamos o gráfico, nota-se uma prevalência de tamanho de massa e de barbatanas na espécie Gentoo.