ELISAvsMNA

# Подгрузим необходимые библиотеки

library(dplyr)

##   
## Присоединяю пакет: 'dplyr'

## Следующие объекты скрыты от 'package:stats':  
##   
## filter, lag

## Следующие объекты скрыты от 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble 3.1.5 v purrr 0.3.4  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## Warning: пакет 'tidyr' был собран под R версии 4.1.2

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(mlbench)

## Warning: пакет 'mlbench' был собран под R версии 4.1.2

library(Hmisc)

## Warning: пакет 'Hmisc' был собран под R версии 4.1.2

## Загрузка требуемого пакета: lattice

## Загрузка требуемого пакета: survival

## Загрузка требуемого пакета: Formula

##   
## Присоединяю пакет: 'Hmisc'

## Следующие объекты скрыты от 'package:dplyr':  
##   
## src, summarize

## Следующие объекты скрыты от 'package:base':  
##   
## format.pval, units

library(GGally)

## Warning: пакет 'GGally' был собран под R версии 4.1.2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(corrplot)

## corrplot 0.91 loaded

library(rpart.plot)

## Warning: пакет 'rpart.plot' был собран под R версии 4.1.2

## Загрузка требуемого пакета: rpart

library(treeheatr)

## Warning: пакет 'treeheatr' был собран под R версии 4.1.2

library(caret)

##   
## Присоединяю пакет: 'caret'

## Следующий объект скрыт от 'package:survival':  
##   
## cluster

## Следующий объект скрыт от 'package:purrr':  
##   
## lift

library(Metrics)

## Warning: пакет 'Metrics' был собран под R версии 4.1.2

##   
## Присоединяю пакет: 'Metrics'

## Следующие объекты скрыты от 'package:caret':  
##   
## precision, recall

library(pROC)

## Type 'citation("pROC")' for a citation.

##   
## Присоединяю пакет: 'pROC'

## Следующий объект скрыт от 'package:Metrics':  
##   
## auc

## Следующие объекты скрыты от 'package:stats':  
##   
## cov, smooth, var

library(rocc)

## Warning: пакет 'rocc' был собран под R версии 4.1.2

## Загрузка требуемого пакета: ROCR

## Warning: пакет 'ROCR' был собран под R версии 4.1.2

library(caTools)

## Warning: пакет 'caTools' был собран под R версии 4.1.2

library(gridExtra)

##   
## Присоединяю пакет: 'gridExtra'

## Следующий объект скрыт от 'package:dplyr':  
##   
## combine

library(spearmanCI)

## Warning: пакет 'spearmanCI' был собран под R версии 4.1.2

library(survminer)

## Загрузка требуемого пакета: ggpubr

library(randomForest)

## Warning: пакет 'randomForest' был собран под R версии 4.1.2

## randomForest 4.6-14

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

##   
## Присоединяю пакет: 'randomForest'

## Следующий объект скрыт от 'package:gridExtra':  
##   
## combine

## Следующий объект скрыт от 'package:ggplot2':  
##   
## margin

## Следующий объект скрыт от 'package:dplyr':  
##   
## combine

#Загрузим данные

load("data/other\_test\_results\_matched\_to\_phone\_survey\_ids.rdata")  
load("data/phone\_survey\_data.rdata")  
load("data/test\_B\_results\_matched\_to\_phone\_survey\_ids.rdata")  
load("data/test\_C\_results\_matched\_to\_phone\_survey\_ids.rdata")  
load("data/test\_D\_results\_matched\_to\_phone\_survey\_ids.rdata")  
load("data/test\_nab.rda")

#Сделаем базу данных, с которой мы будем работать  
  
  
```r  
data\_for\_ab\_IgG1 <-merge(test\_B\_results\_matched\_to\_phone\_survey\_ids, test\_C\_results\_matched\_to\_phone\_survey\_ids, by = "ID", all.x = T)  
  
data\_for\_ab\_IgG2 <-merge(data\_for\_ab\_IgG1, test\_D\_results\_matched\_to\_phone\_survey\_ids, by = "ID", all.x = T)  
  
data\_all<-merge(phone\_survey\_data, other\_test\_results\_matched\_to\_phone\_survey\_ids, by = "ID", all.y = T)  
  
data\_for\_ab<- merge(data\_for\_ab\_IgG2, data\_all, by = "ID", all.y = T)

## Warning in merge.data.table(data\_for\_ab\_IgG2, data\_all, by = "ID", all.y = T):  
## column names 'draw\_sample\_date.x', 'visited\_clinic.x', 'draw\_sample\_date.y',  
## 'visited\_clinic.y' are duplicated in the result

data\_for\_ab<- data\_for\_ab %>%   
 dplyr::select(c(ID, IgG\_testB\_quantitative, IgA\_or\_G\_or\_M\_testC\_quantitative, IgG\_or\_M\_testD\_quantitative, male, age, neutralizing\_antibodies\_quantitative))  
  
data\_ab\_only<- data\_for\_ab %>%   
 dplyr::select(c(ID, IgG\_testB\_quantitative,IgA\_or\_G\_or\_M\_testC\_quantitative, IgG\_or\_M\_testD\_quantitative, neutralizing\_antibodies\_quantitative))  
  
data\_ab\_only = na.omit(data\_ab\_only)

#Выставим пороги чувствительности по методам ИФА - 1, а по MNA - 80. Оставим в базах данных только ответчиков.

data\_Abbott<-data\_for\_ab %>%   
 dplyr::select(c(ID, IgG\_testB\_quantitative, neutralizing\_antibodies\_quantitative)) %>% filter(IgG\_testB\_quantitative>=1&neutralizing\_antibodies\_quantitative>=80)  
  
data\_Coronapass<-data\_for\_ab %>%   
 dplyr::select(c(ID, IgA\_or\_G\_or\_M\_testC\_quantitative, neutralizing\_antibodies\_quantitative)) %>% filter(IgA\_or\_G\_or\_M\_testC\_quantitative>=1&neutralizing\_antibodies\_quantitative>=80)  
  
data\_VectorBest<-data\_for\_ab %>%   
 dplyr::select(c(ID, IgG\_or\_M\_testD\_quantitative, neutralizing\_antibodies\_quantitative)) %>% filter(IgG\_or\_M\_testD\_quantitative>=1&neutralizing\_antibodies\_quantitative>=80)

# Корреляция MNA и различных наборов для ИФА

g1<- ggplot(data\_ab\_only, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgG\_testB\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 0.75, color = "red")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 0.75, color = "red")+  
 ylab("ELISA Abbott")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 10)  
  
cor\_abbot<-cor.test( data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgG\_testB\_quantitative, method = c("spearman"), conf.level = 0.95)

## Warning in cor.test.default(data\_ab\_only$neutralizing\_antibodies\_quantitative, :  
## Cannot compute exact p-value with ties

abbotCI<-spearmanCI(data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgG\_testB\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## 0.5967001 0.7199568  
## sample estimate   
## 0.6583371

g1\_1<- ggplot(data\_Abbott, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgG\_testB\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 1, color = "grey50")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 1, color = "grey50")+  
 ylab("ELISA Abbott")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 13)  
  
cor\_abbot\_pos<-cor.test( data\_Abbott$neutralizing\_antibodies\_quantitative, data\_Abbott$IgG\_testB\_quantitative, method = c("spearman"), conf.level = 0.95)

## Warning in cor.test.default(data\_Abbott$neutralizing\_antibodies\_quantitative, :  
## Cannot compute exact p-value with ties

abbotCI\_pos<-spearmanCI(data\_Abbott$neutralizing\_antibodies\_quantitative, data\_Abbott$IgG\_testB\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## -0.02605964 0.5055637  
## sample estimate   
## 0.2398885

g2<- ggplot(data\_ab\_only, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgA\_or\_G\_or\_M\_testC\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 0.75, color = "red")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 0.75, color = "red")+  
 ylab("ELISA Genetico")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 10)  
  
cor\_gen<-cor.test( data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgA\_or\_G\_or\_M\_testC\_quantitative, method = c("spearman"), conf.level = 0.95)

## Warning in cor.test.default(data\_ab\_only$neutralizing\_antibodies\_quantitative, :  
## Cannot compute exact p-value with ties

genCI<-spearmanCI(data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgA\_or\_G\_or\_M\_testC\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## 0.5505387 0.6808581  
## sample estimate   
## 0.6157073

g2\_2<-ggplot(data\_Coronapass, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgA\_or\_G\_or\_M\_testC\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 1, color = "grey50")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 1, color = "grey50")+  
 ylab("ELISA Genetico")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 13)  
  
cor\_gen\_pos<-cor.test( data\_Coronapass$neutralizing\_antibodies\_quantitative, data\_Coronapass$IgA\_or\_G\_or\_M\_testC\_quantitative, method = c("spearman"), conf.level = 0.95)

## Warning in  
## cor.test.default(data\_Coronapass$neutralizing\_antibodies\_quantitative, : Cannot  
## compute exact p-value with ties

genCI<-spearmanCI(data\_Coronapass$neutralizing\_antibodies\_quantitative, data\_Coronapass$IgA\_or\_G\_or\_M\_testC\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## -0.1141756 0.3726211  
## sample estimate   
## 0.1293057

g3<- ggplot(data\_ab\_only, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgG\_or\_M\_testD\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 0.75, color = "red")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 0.75, color = "red")+  
 ylab("ELISA Vector")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 10)  
  
cor\_vec<-cor.test( data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgG\_or\_M\_testD\_quantitative, method = c("spearman"), conf.level = 0.95)

## Warning in cor.test.default(data\_ab\_only$neutralizing\_antibodies\_quantitative, :  
## Cannot compute exact p-value with ties

vecCI<-spearmanCI(data\_ab\_only$neutralizing\_antibodies\_quantitative, data\_ab\_only$IgG\_or\_M\_testD\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## 0.7107691 0.8092698  
## sample estimate   
## 0.7600288

g3\_3<-ggplot(data\_VectorBest, aes(x = neutralizing\_antibodies\_quantitative,  
 y=IgG\_or\_M\_testD\_quantitative)) +  
 geom\_jitter(shape = 21,  
 size=1,  
 color = "black",  
 alpha = 0.8)+  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")+  
 geom\_hline(yintercept = 1, linetype = "dashed", size = 1, color = "grey50")+  
 geom\_vline(xintercept = 80, linetype = "dashed", size = 1, color = "grey50")+  
 ylab("ELISA Vector")+  
 xlab("MNA")+  
 geom\_smooth(method = lm, se = F) +  
 theme\_classic2(base\_size = 13)  
  
cor\_vec\_pos<-cor.test( data\_VectorBest$neutralizing\_antibodies\_quantitative, data\_VectorBest$IgG\_or\_M\_testD\_quantitative, method = c("spearman"), conf.level = 0.95)

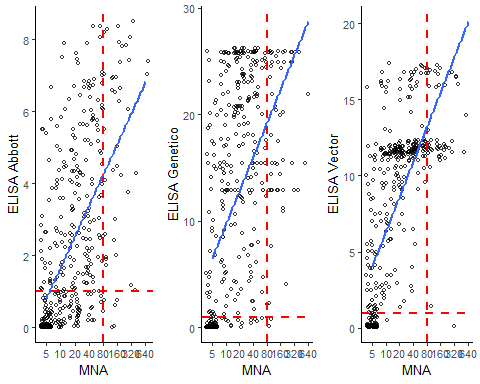
## Warning in  
## cor.test.default(data\_VectorBest$neutralizing\_antibodies\_quantitative, : Cannot  
## compute exact p-value with ties

vecCI\_pos<-spearmanCI(data\_VectorBest$neutralizing\_antibodies\_quantitative, data\_VectorBest$IgG\_or\_M\_testD\_quantitative, method = "Euclidean", level = 0.95, plot = F)

## confidence interval   
## 2.5 % 97.5 %  
## -0.137554 0.3418079  
## sample estimate   
## 0.1021959

ggarrange(g1, g2, g3, ncol = 3, nrow = 1)

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'



#Применим метод машинного обучения “Дерево решений” в задаче классификации “ответчик не ответчик к SARS-CoV”, взяв в качестве разделителя титр антител 1:80 в MNA

#Создадим базы данных, с которыми нам будет удобно работать

data\_for\_ab<- data\_for\_ab %>%   
rename\_all(function(x)x %>% stringi::stri\_replace\_all\_regex(c("IgG\_testB\_quantitative", "IgA\_or\_G\_or\_M\_testC\_quantitative", "IgG\_or\_M\_testD\_quantitative", "male", "age", "neutralizing\_antibodies\_quantitative"), c("Abbott", "Genetico", "Vector", "gender", "age", "Nab"), vectorize\_all = FALSE))  
  
data\_for\_ab<- data\_for\_ab %>%   
mutate('MNT 1:80' = ifelse (Nab < 80, "No SARS-Cov-2", "SARS-Cov-2"))  
  
dataset <- data\_for\_ab %>%   
 dplyr::select(!c( Nab))  
  
dataset$`MNT 1:80` <- as.factor(dataset$`MNT 1:80`)  
dataset$gender <- as.factor(dataset$gender)

dataset\_80 <- data\_for\_ab%>%   
 dplyr::select(c(Abbott, Genetico, Vector, 'MNT 1:80'))  
  
dataset\_av <- data\_for\_ab%>%   
 dplyr::select(c(Abbott, Vector, 'MNT 1:80'))  
  
dataset\_80\_all <- data\_for\_ab%>%   
 dplyr::select(c(Abbott, Genetico, Vector, gender, age ,'MNT 1:80'))  
  
dataset\_male <- dataset\_80\_all%>%   
 filter(gender == 1)  
  
dataset\_female <- dataset\_80\_all%>%   
 filter(gender == 0)  
  
dataset\_abbott<-data\_for\_ab %>%   
 dplyr::select(c(Abbott, gender, age , 'MNT 1:80'))  
  
dataset\_genetico<-data\_for\_ab %>%   
 dplyr::select(c(Genetico, gender, age , 'MNT 1:80'))  
  
dataset\_vector <- data\_for\_ab%>%   
 dplyr::select(c( Vector, gender, age , 'MNT 1:80'))

#Построим РОК-кривые для трех тестов против MNA.

roc.list <- roc(`MNT 1:80` ~ Abbott + Genetico +Vector, data = dataset\_80)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

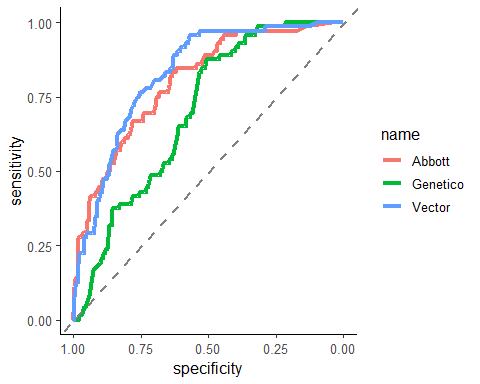
## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

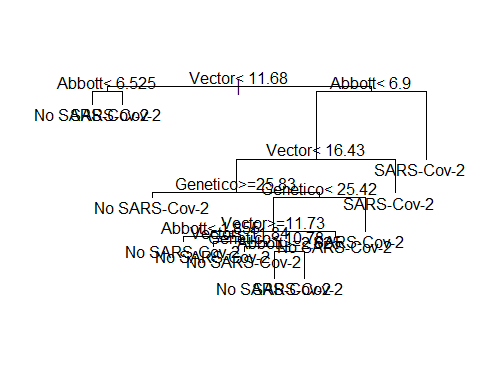
ggroc(roc.list, cex= 1.5)+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()



# Сделаем машинное обучение методом “Дерево решений”, посмотрев, можно ли предсказать, будет ли у человека в MNA титр >=1:80, зная результаты трех тестов ELISA.

#Рекурсивное разбиение и регрессионное дерево

model <- rpart(`MNT 1:80` ~., data = dataset\_80)  
par(xpd = NA)   
plot(model)  
text(model, digits = 3)



# Посмотрим, сколько у нас ответчиков в MNA по порогу 1:80.

response.column = 'MNT 1:80'  
response = dataset[[response.column]]  
table(response)

## response  
## No SARS-Cov-2 SARS-Cov-2   
## 354 72

# Сделаем тестовую и тренировочную выборки

set.seed(1265)   
index <- sample(2, nrow(dataset\_80), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_80 <- dataset\_80[index==1, ]   
test\_base\_80 <- dataset\_80[index == 2, ]   
  
table(train\_base\_80$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 285 58

table(test\_base\_80$`MNT 1:80`)

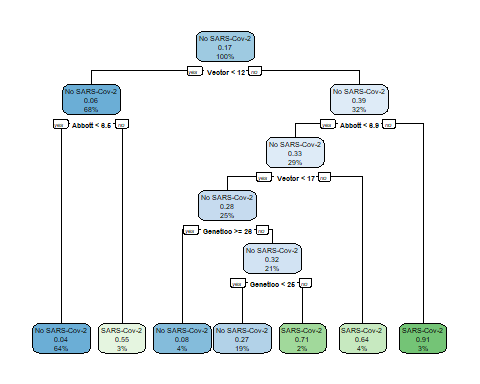
##   
## No SARS-Cov-2 SARS-Cov-2   
## 69 14

# Сделаем модель Дерево решений

Nab\_model\_80 <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_80,   
 method = "class")  
summary(Nab\_model\_80)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_80, method = "class")  
## n=343 (850 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.07758621 0 1.0000000 1.0000000 0.1196910  
## 2 0.06896552 2 0.8448276 1.0172414 0.1205064  
## 3 0.02586207 3 0.7758621 0.9310345 0.1162976  
## 4 0.01724138 5 0.7241379 0.8448276 0.1117369  
## 5 0.01000000 6 0.7068966 1.0000000 0.1196910  
##   
## Variable importance  
## Vector Abbott Genetico   
## 51 36 13   
##   
## Node number 1: 343 observations, complexity param=0.07758621  
## predicted class=No SARS-Cov-2 expected loss=0.1690962 P(node) =1  
## class counts: 285 58  
## probabilities: 0.831 0.169   
## left son=2 (232 obs) right son=3 (111 obs)  
## Primary splits:  
## Vector < 11.685 to the left, improve=15.786470, (3 missing)  
## Abbott < 6.92 to the left, improve=15.482630, (0 missing)  
## Genetico < 10.8755 to the left, improve= 8.064598, (3 missing)  
## Surrogate splits:  
## Abbott < 4.825 to the left, agree=0.706, adj=0.091, (3 split)  
## Genetico < 23.4655 to the left, agree=0.700, adj=0.073, (0 split)  
##   
## Node number 2: 232 observations, complexity param=0.01724138  
## predicted class=No SARS-Cov-2 expected loss=0.06465517 P(node) =0.6763848  
## class counts: 217 15  
## probabilities: 0.935 0.065   
## left son=4 (221 obs) right son=5 (11 obs)  
## Primary splits:  
## Abbott < 6.525 to the left, improve=5.338831, (0 missing)  
## Vector < 10.45 to the left, improve=2.452415, (2 missing)  
## Genetico < 15.337 to the left, improve=1.131184, (3 missing)  
##   
## Node number 3: 111 observations, complexity param=0.07758621  
## predicted class=No SARS-Cov-2 expected loss=0.3873874 P(node) =0.3236152  
## class counts: 68 43  
## probabilities: 0.613 0.387   
## left son=6 (100 obs) right son=7 (11 obs)  
## Primary splits:  
## Abbott < 6.92 to the left, improve=6.646503, (0 missing)  
## Vector < 16.84 to the left, improve=4.861466, (1 missing)  
## Genetico < 25.869 to the right, improve=1.495142, (0 missing)  
##   
## Node number 4: 221 observations  
## predicted class=No SARS-Cov-2 expected loss=0.04072398 P(node) =0.6443149  
## class counts: 212 9  
## probabilities: 0.959 0.041   
##   
## Node number 5: 11 observations  
## predicted class=SARS-Cov-2 expected loss=0.4545455 P(node) =0.03206997  
## class counts: 5 6  
## probabilities: 0.455 0.545   
##   
## Node number 6: 100 observations, complexity param=0.06896552  
## predicted class=No SARS-Cov-2 expected loss=0.33 P(node) =0.2915452  
## class counts: 67 33  
## probabilities: 0.670 0.330   
## left son=12 (86 obs) right son=13 (14 obs)  
## Primary splits:  
## Vector < 16.515 to the left, improve=3.124370, (1 missing)  
## Abbott < 4.04 to the left, improve=1.740879, (0 missing)  
## Genetico < 25.827 to the right, improve=1.140266, (0 missing)  
##   
## Node number 7: 11 observations  
## predicted class=SARS-Cov-2 expected loss=0.09090909 P(node) =0.03206997  
## class counts: 1 10  
## probabilities: 0.091 0.909   
##   
## Node number 12: 86 observations, complexity param=0.02586207  
## predicted class=No SARS-Cov-2 expected loss=0.2790698 P(node) =0.2507289  
## class counts: 62 24  
## probabilities: 0.721 0.279   
## left son=24 (13 obs) right son=25 (73 obs)  
## Primary splits:  
## Genetico < 25.827 to the right, improve=1.251648, (0 missing)  
## Vector < 16.035 to the right, improve=1.216290, (1 missing)  
## Abbott < 6.425 to the right, improve=1.186930, (0 missing)  
##   
## Node number 13: 14 observations  
## predicted class=SARS-Cov-2 expected loss=0.3571429 P(node) =0.04081633  
## class counts: 5 9  
## probabilities: 0.357 0.643   
##   
## Node number 24: 13 observations  
## predicted class=No SARS-Cov-2 expected loss=0.07692308 P(node) =0.03790087  
## class counts: 12 1  
## probabilities: 0.923 0.077   
##   
## Node number 25: 73 observations, complexity param=0.02586207  
## predicted class=No SARS-Cov-2 expected loss=0.3150685 P(node) =0.212828  
## class counts: 50 23  
## probabilities: 0.685 0.315   
## left son=50 (66 obs) right son=51 (7 obs)  
## Primary splits:  
## Genetico < 25.358 to the left, improve=2.467888, (0 missing)  
## Vector < 13.015 to the right, improve=1.269033, (0 missing)  
## Abbott < 4.04 to the left, improve=1.225209, (0 missing)  
##   
## Node number 50: 66 observations  
## predicted class=No SARS-Cov-2 expected loss=0.2727273 P(node) =0.1924198  
## class counts: 48 18  
## probabilities: 0.727 0.273   
##   
## Node number 51: 7 observations  
## predicted class=SARS-Cov-2 expected loss=0.2857143 P(node) =0.02040816  
## class counts: 2 5  
## probabilities: 0.286 0.714

rpart.plot(x = Nab\_model\_80, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)

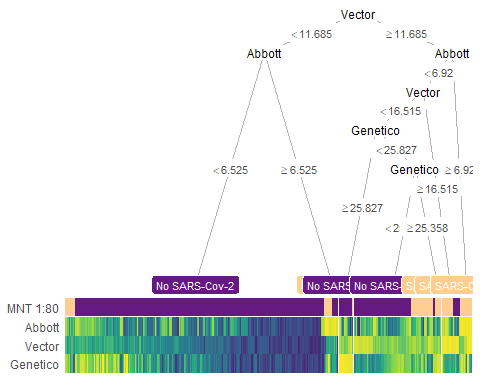


# Нарисуем тепловое дерево для сделанной модели

x <- partykit::as.party(Nab\_model\_80)  
heat\_tree(x = x)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



# Посмотрим метрики

class\_predicted\_80 <- predict(object = Nab\_model\_80,   
 newdata = test\_base\_80,   
 type = "class")  
confusionMatrix(data = class\_predicted\_80,   
 reference = as.factor(test\_base\_80$`MNT 1:80`))

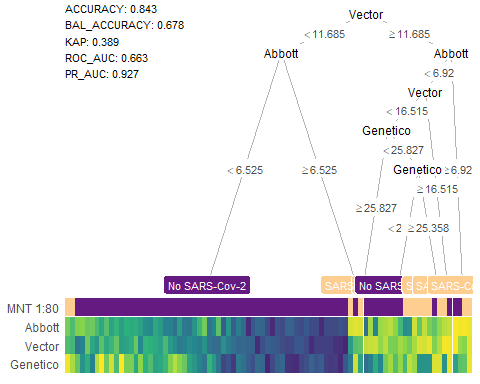
## Registered S3 methods overwritten by 'proxy':  
## method from   
## print.registry\_field registry  
## print.registry\_entry registry

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 64 8  
## SARS-Cov-2 5 6  
##   
## Accuracy : 0.8434   
## 95% CI : (0.7471, 0.9139)  
## No Information Rate : 0.8313   
## P-Value [Acc > NIR] : 0.4545   
##   
## Kappa : 0.3894   
##   
## Mcnemar's Test P-Value : 0.5791   
##   
## Sensitivity : 0.9275   
## Specificity : 0.4286   
## Pos Pred Value : 0.8889   
## Neg Pred Value : 0.5455   
## Prevalence : 0.8313   
## Detection Rate : 0.7711   
## Detection Prevalence : 0.8675   
## Balanced Accuracy : 0.6781   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = x,  
 data\_test = test\_base\_80)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



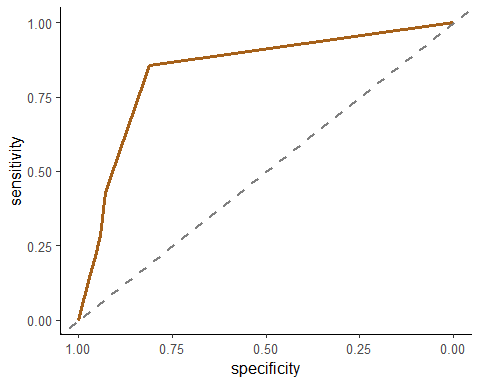
# Построим РОК-кривую для нашей модели

pred\_prob\_80 <-as.data.frame( predict(Nab\_model\_80, test\_base\_80, type="prob"))  
  
roc\_80 <- roc(test\_base\_80$`MNT 1:80`, pred\_prob\_80$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

ggroc(roc\_80, cex= 1.2, color = '#A6611A')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()



## Дальше сделаем машиной обучение индивидуально для каждого теста. Но будем учитывать пол и возраст. Вдруг, это поможет увеличить специфичность нашей модели.

#Для системы Abbott

index <- sample(2, nrow(dataset\_abbott), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_abbott <- dataset\_abbott[index==1, ]   
test\_base\_abbott <- dataset\_abbott[index == 2, ]   
  
table(train\_base\_abbott$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 287 54

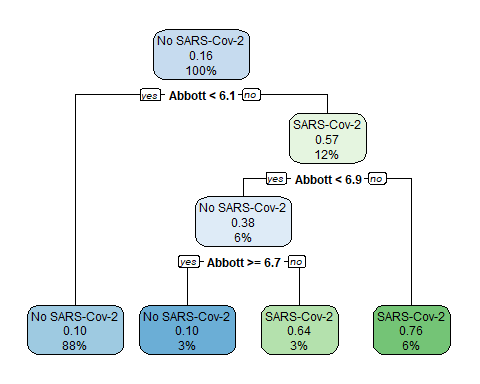
table(test\_base\_abbott$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 67 18

Nab\_model\_abbott <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_abbott,   
 method = "class")  
summary(Nab\_model\_abbott)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_abbott, method = "class")  
## n=341 (871 пропущенное наблюдение удалено)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.11111111 0 1.0000000 1.0000000 0.1248438  
## 2 0.09259259 1 0.8888889 0.9814815 0.1238977  
## 3 0.05555556 2 0.7962963 0.9259259 0.1209652  
## 4 0.01000000 3 0.7407407 0.8703704 0.1178831  
##   
## Variable importance  
## Abbott age gender   
## 92 5 2   
##   
## Node number 1: 341 observations, complexity param=0.1111111  
## predicted class=No SARS-Cov-2 expected loss=0.1583578 P(node) =1  
## class counts: 287 54  
## probabilities: 0.842 0.158   
## left son=2 (299 obs) right son=3 (42 obs)  
## Primary splits:  
## Abbott < 6.055 to the left, improve=16.3460000, (0 missing)  
## age < 65.5 to the left, improve= 1.5758220, (0 missing)  
## gender < 0.5 to the left, improve= 0.1617748, (0 missing)  
##   
## Node number 2: 299 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1003344 P(node) =0.8768328  
## class counts: 269 30  
## probabilities: 0.900 0.100   
##   
## Node number 3: 42 observations, complexity param=0.09259259  
## predicted class=SARS-Cov-2 expected loss=0.4285714 P(node) =0.1231672  
## class counts: 18 24  
## probabilities: 0.429 0.571   
## left son=6 (21 obs) right son=7 (21 obs)  
## Primary splits:  
## Abbott < 6.9 to the left, improve=3.04761900, (0 missing)  
## age < 48.5 to the right, improve=0.76190480, (0 missing)  
## gender < 0.5 to the left, improve=0.09920635, (0 missing)  
## Surrogate splits:  
## age < 33.5 to the left, agree=0.619, adj=0.238, (0 split)  
## gender < 0.5 to the right, agree=0.548, adj=0.095, (0 split)  
##   
## Node number 6: 21 observations, complexity param=0.05555556  
## predicted class=No SARS-Cov-2 expected loss=0.3809524 P(node) =0.06158358  
## class counts: 13 8  
## probabilities: 0.619 0.381   
## left son=12 (10 obs) right son=13 (11 obs)  
## Primary splits:  
## Abbott < 6.67 to the right, improve=3.01385300, (0 missing)  
## age < 53 to the right, improve=1.69322300, (0 missing)  
## gender < 0.5 to the left, improve=0.01385281, (0 missing)  
## Surrogate splits:  
## age < 58 to the left, agree=0.619, adj=0.2, (0 split)  
## gender < 0.5 to the left, agree=0.571, adj=0.1, (0 split)  
##   
## Node number 7: 21 observations  
## predicted class=SARS-Cov-2 expected loss=0.2380952 P(node) =0.06158358  
## class counts: 5 16  
## probabilities: 0.238 0.762   
##   
## Node number 12: 10 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1 P(node) =0.02932551  
## class counts: 9 1  
## probabilities: 0.900 0.100   
##   
## Node number 13: 11 observations  
## predicted class=SARS-Cov-2 expected loss=0.3636364 P(node) =0.03225806  
## class counts: 4 7  
## probabilities: 0.364 0.636

rpart.plot(x = Nab\_model\_abbott, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)



abbott <- partykit::as.party(Nab\_model\_abbott)  
heat\_tree(x = abbott)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



class\_predicted\_abbott <- predict(object = Nab\_model\_abbott,   
 newdata = test\_base\_abbott,   
 type = "class")  
confusionMatrix(data = class\_predicted\_abbott,   
 reference = as.factor(test\_base\_abbott$`MNT 1:80`))

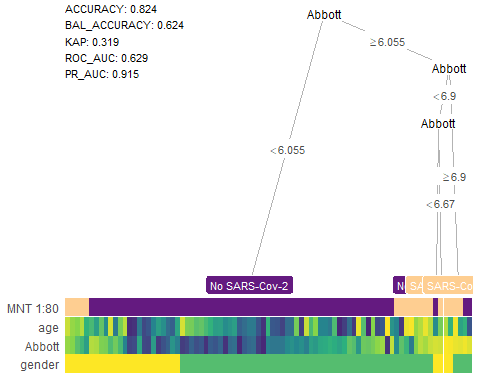
## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 65 13  
## SARS-Cov-2 2 5  
##   
## Accuracy : 0.8235   
## 95% CI : (0.7257, 0.8977)  
## No Information Rate : 0.7882   
## P-Value [Acc > NIR] : 0.258213   
##   
## Kappa : 0.3193   
##   
## Mcnemar's Test P-Value : 0.009823   
##   
## Sensitivity : 0.9701   
## Specificity : 0.2778   
## Pos Pred Value : 0.8333   
## Neg Pred Value : 0.7143   
## Prevalence : 0.7882   
## Detection Rate : 0.7647   
## Detection Prevalence : 0.9176   
## Balanced Accuracy : 0.6240   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = abbott,  
 data\_test = test\_base\_abbott)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 2 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



pred\_prob\_abbott <-as.data.frame( predict(Nab\_model\_abbott, test\_base\_abbott, type="prob"))  
  
roc\_abbott <- roc(test\_base\_abbott$`MNT 1:80`, pred\_prob\_abbott$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

r\_ab<- ggroc(roc\_abbott, cex= 1.2, color = 'green')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()+  
 ggtitle("ELISA Abbott") +  
 labs(x = "1 - Specificity",  
 y = "Sensitivity")

#Для системы Genetico

index <- sample(2, nrow(dataset\_genetico), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_genetico <- dataset\_genetico[index==1, ]   
test\_base\_genetico <- dataset\_genetico[index == 2, ]   
  
table(train\_base\_genetico$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 294 55

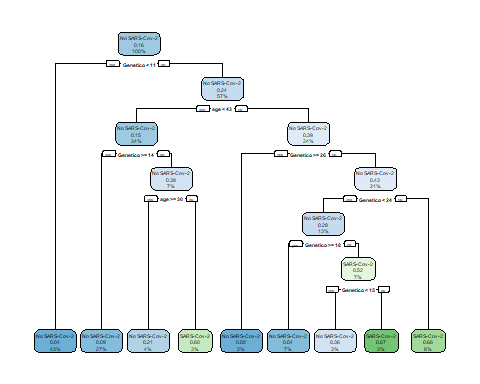
table(test\_base\_genetico$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 60 17

Nab\_model\_genetico <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_genetico,   
 method = "class")  
summary(Nab\_model\_genetico)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_genetico, method = "class")  
## n=349 (858 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.04090909 0 1.0000000 1.000000 0.1237598  
## 2 0.03636364 4 0.8363636 1.054545 0.1264402  
## 3 0.01818182 6 0.7636364 1.145455 0.1306405  
## 4 0.01000000 8 0.7272727 1.181818 0.1322332  
##   
## Variable importance  
## Genetico age gender   
## 72 27 1   
##   
## Node number 1: 349 observations, complexity param=0.04090909  
## predicted class=No SARS-Cov-2 expected loss=0.1575931 P(node) =1  
## class counts: 294 55  
## probabilities: 0.842 0.158   
## left son=2 (149 obs) right son=3 (200 obs)  
## Primary splits:  
## Genetico < 10.8755 to the left, improve=7.3733070, (3 missing)  
## age < 53.5 to the left, improve=2.8992090, (0 missing)  
## gender < 0.5 to the left, improve=0.8401735, (0 missing)  
## Surrogate splits:  
## age < 70.5 to the right, agree=0.578, adj=0.02, (3 split)  
##   
## Node number 2: 149 observations  
## predicted class=No SARS-Cov-2 expected loss=0.04026846 P(node) =0.4269341  
## class counts: 143 6  
## probabilities: 0.960 0.040   
##   
## Node number 3: 200 observations, complexity param=0.04090909  
## predicted class=No SARS-Cov-2 expected loss=0.245 P(node) =0.5730659  
## class counts: 151 49  
## probabilities: 0.755 0.245   
## left son=6 (117 obs) right son=7 (83 obs)  
## Primary splits:  
## age < 42.5 to the left, improve=5.6048700, (0 missing)  
## Genetico < 13.634 to the right, improve=2.1112630, (3 missing)  
## gender < 0.5 to the left, improve=0.1132346, (0 missing)  
## Surrogate splits:  
## Genetico < 23.7025 to the left, agree=0.62, adj=0.084, (0 split)  
##   
## Node number 6: 117 observations, complexity param=0.01818182  
## predicted class=No SARS-Cov-2 expected loss=0.1452991 P(node) =0.3352436  
## class counts: 100 17  
## probabilities: 0.855 0.145   
## left son=12 (93 obs) right son=13 (24 obs)  
## Primary splits:  
## Genetico < 13.634 to the right, improve=3.13050600, (2 missing)  
## age < 22.5 to the right, improve=0.68945870, (0 missing)  
## gender < 0.5 to the right, improve=0.02094017, (0 missing)  
##   
## Node number 7: 83 observations, complexity param=0.04090909  
## predicted class=No SARS-Cov-2 expected loss=0.3855422 P(node) =0.2378223  
## class counts: 51 32  
## probabilities: 0.614 0.386   
## left son=14 (8 obs) right son=15 (75 obs)  
## Primary splits:  
## Genetico < 25.847 to the right, improve=2.7000660, (1 missing)  
## age < 63.5 to the left, improve=2.2173200, (0 missing)  
## gender < 0.5 to the left, improve=0.3565137, (0 missing)  
##   
## Node number 12: 93 observations  
## predicted class=No SARS-Cov-2 expected loss=0.08602151 P(node) =0.2664756  
## class counts: 85 8  
## probabilities: 0.914 0.086   
##   
## Node number 13: 24 observations, complexity param=0.01818182  
## predicted class=No SARS-Cov-2 expected loss=0.375 P(node) =0.06876791  
## class counts: 15 9  
## probabilities: 0.625 0.375   
## left son=26 (14 obs) right son=27 (10 obs)  
## Primary splits:  
## age < 29.5 to the right, improve=1.735714000, (0 missing)  
## Genetico < 12.761 to the left, improve=0.672222200, (0 missing)  
## gender < 0.5 to the right, improve=0.005244755, (0 missing)  
## Surrogate splits:  
## Genetico < 11.809 to the right, agree=0.625, adj=0.1, (0 split)  
##   
## Node number 14: 8 observations  
## predicted class=No SARS-Cov-2 expected loss=0 P(node) =0.02292264  
## class counts: 8 0  
## probabilities: 1.000 0.000   
##   
## Node number 15: 75 observations, complexity param=0.04090909  
## predicted class=No SARS-Cov-2 expected loss=0.4266667 P(node) =0.2148997  
## class counts: 43 32  
## probabilities: 0.573 0.427   
## left son=30 (46 obs) right son=31 (29 obs)  
## Primary splits:  
## Genetico < 24.2275 to the left, improve=4.7319870, (1 missing)  
## age < 63.5 to the left, improve=1.6457140, (0 missing)  
## gender < 0.5 to the left, improve=0.1976938, (0 missing)  
## Surrogate splits:  
## age < 63.5 to the left, agree=0.662, adj=0.138, (1 split)  
##   
## Node number 26: 14 observations  
## predicted class=No SARS-Cov-2 expected loss=0.2142857 P(node) =0.04011461  
## class counts: 11 3  
## probabilities: 0.786 0.214   
##   
## Node number 27: 10 observations  
## predicted class=SARS-Cov-2 expected loss=0.4 P(node) =0.0286533  
## class counts: 4 6  
## probabilities: 0.400 0.600   
##   
## Node number 30: 46 observations, complexity param=0.03636364  
## predicted class=No SARS-Cov-2 expected loss=0.2826087 P(node) =0.1318052  
## class counts: 33 13  
## probabilities: 0.717 0.283   
## left son=60 (23 obs) right son=61 (23 obs)  
## Primary splits:  
## Genetico < 18.0915 to the right, improve=5.1015370, (1 missing)  
## age < 50.5 to the left, improve=1.0069500, (0 missing)  
## gender < 0.5 to the right, improve=0.0244936, (0 missing)  
## Surrogate splits:  
## age < 50.5 to the left, agree=0.600, adj=0.182, (1 split)  
## gender < 0.5 to the right, agree=0.533, adj=0.045, (0 split)  
##   
## Node number 31: 29 observations  
## predicted class=SARS-Cov-2 expected loss=0.3448276 P(node) =0.08309456  
## class counts: 10 19  
## probabilities: 0.345 0.655   
##   
## Node number 60: 23 observations  
## predicted class=No SARS-Cov-2 expected loss=0.04347826 P(node) =0.06590258  
## class counts: 22 1  
## probabilities: 0.957 0.043   
##   
## Node number 61: 23 observations, complexity param=0.03636364  
## predicted class=SARS-Cov-2 expected loss=0.4782609 P(node) =0.06590258  
## class counts: 11 12  
## probabilities: 0.478 0.522   
## left son=122 (11 obs) right son=123 (12 obs)  
## Primary splits:  
## Genetico < 15.324 to the left, improve=1.0540180, (0 missing)  
## age < 51.5 to the left, improve=0.5244147, (0 missing)  
## gender < 0.5 to the right, improve=0.1766736, (0 missing)  
## Surrogate splits:  
## age < 51.5 to the left, agree=0.696, adj=0.364, (0 split)  
##   
## Node number 122: 11 observations  
## predicted class=No SARS-Cov-2 expected loss=0.3636364 P(node) =0.03151862  
## class counts: 7 4  
## probabilities: 0.636 0.364   
##   
## Node number 123: 12 observations  
## predicted class=SARS-Cov-2 expected loss=0.3333333 P(node) =0.03438395  
## class counts: 4 8  
## probabilities: 0.333 0.667

rpart.plot(x = Nab\_model\_genetico, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)

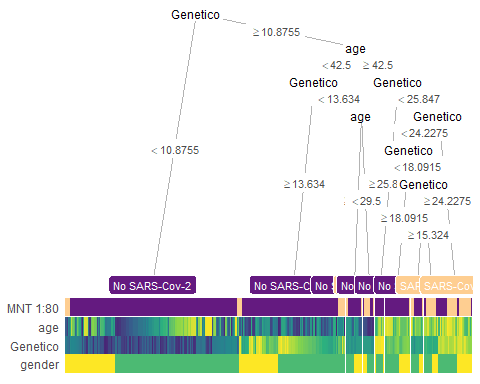


genetico <- partykit::as.party(Nab\_model\_genetico)  
heat\_tree(x = genetico)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



class\_predicted\_genetico <- predict(object = Nab\_model\_genetico,   
 newdata = test\_base\_genetico,   
 type = "class")  
confusionMatrix(data = class\_predicted\_genetico,   
 reference = as.factor(test\_base\_genetico$`MNT 1:80`))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 57 13  
## SARS-Cov-2 3 4  
##   
## Accuracy : 0.7922   
## 95% CI : (0.6846, 0.8763)  
## No Information Rate : 0.7792   
## P-Value [Acc > NIR] : 0.45548   
##   
## Kappa : 0.2348   
##   
## Mcnemar's Test P-Value : 0.02445   
##   
## Sensitivity : 0.9500   
## Specificity : 0.2353   
## Pos Pred Value : 0.8143   
## Neg Pred Value : 0.5714   
## Prevalence : 0.7792   
## Detection Rate : 0.7403   
## Detection Prevalence : 0.9091   
## Balanced Accuracy : 0.5926   
##   
## 'Positive' Class : No SARS-Cov-2   
##

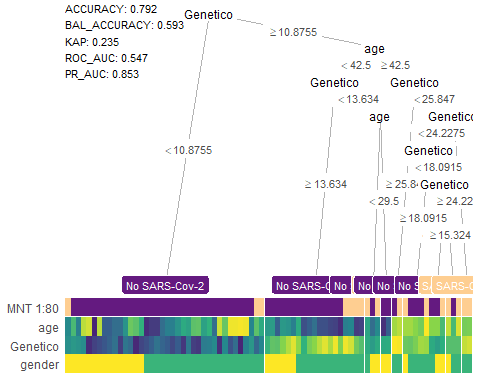
heat\_tree(  
 x = genetico,  
 data\_test = test\_base\_genetico)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 2 treated as  
## interval scaled

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 2 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



pred\_prob\_genetico <-as.data.frame( predict(Nab\_model\_genetico, test\_base\_genetico, type="prob"))  
  
roc\_genetico <- roc(test\_base\_genetico$`MNT 1:80`, pred\_prob\_genetico$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

r\_gen<-ggroc(roc\_genetico, cex= 1.2, color = 'red')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()+  
 ggtitle("ELISA Genetico") +  
 labs(x = "1 - Specificity",  
 y = "Sensitivity")

# Для системы Vector

index <- sample(2, nrow(dataset\_vector), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_vector <- dataset\_vector[index==1, ]   
test\_base\_vector <- dataset\_vector[index == 2, ]   
  
table(train\_base\_vector$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 271 59

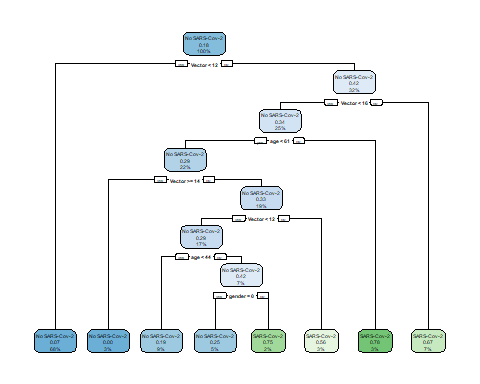
table(test\_base\_vector$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 83 13

Nab\_model\_vector <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_vector,   
 method = "class")  
summary(Nab\_model\_vector)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_vector, method = "class")  
## n=330 (874 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.06779661 0 1.0000000 1.0000000 0.1179782  
## 2 0.02118644 3 0.7796610 0.8644068 0.1112957  
## 3 0.01000000 7 0.6949153 0.8644068 0.1112957  
##   
## Variable importance  
## Vector age gender   
## 74 18 8   
##   
## Node number 1: 330 observations, complexity param=0.06779661  
## predicted class=No SARS-Cov-2 expected loss=0.1787879 P(node) =1  
## class counts: 271 59  
## probabilities: 0.821 0.179   
## left son=2 (224 obs) right son=3 (106 obs)  
## Primary splits:  
## Vector < 11.685 to the left, improve=17.3297200, (2 missing)  
## age < 42.5 to the left, improve= 3.1717990, (0 missing)  
## gender < 0.5 to the left, improve= 0.3292208, (0 missing)  
## Surrogate splits:  
## age < 63.5 to the left, agree=0.683, adj=0.019, (2 split)  
##   
## Node number 2: 224 observations  
## predicted class=No SARS-Cov-2 expected loss=0.06696429 P(node) =0.6787879  
## class counts: 209 15  
## probabilities: 0.933 0.067   
##   
## Node number 3: 106 observations, complexity param=0.06779661  
## predicted class=No SARS-Cov-2 expected loss=0.4150943 P(node) =0.3212121  
## class counts: 62 44  
## probabilities: 0.585 0.415   
## left son=6 (82 obs) right son=7 (24 obs)  
## Primary splits:  
## Vector < 15.875 to the left, improve=3.9269830, (0 missing)  
## age < 61 to the left, improve=3.0355280, (0 missing)  
## gender < 0.5 to the left, improve=0.5211951, (0 missing)  
##   
## Node number 6: 82 observations, complexity param=0.06779661  
## predicted class=No SARS-Cov-2 expected loss=0.3414634 P(node) =0.2484848  
## class counts: 54 28  
## probabilities: 0.659 0.341   
## left son=12 (73 obs) right son=13 (9 obs)  
## Primary splits:  
## age < 61 to the left, improve=3.84912900, (0 missing)  
## Vector < 14.06 to the right, improve=2.65582700, (0 missing)  
## gender < 0.5 to the left, improve=0.07632464, (0 missing)  
##   
## Node number 7: 24 observations  
## predicted class=SARS-Cov-2 expected loss=0.3333333 P(node) =0.07272727  
## class counts: 8 16  
## probabilities: 0.333 0.667   
##   
## Node number 12: 73 observations, complexity param=0.02118644  
## predicted class=No SARS-Cov-2 expected loss=0.2876712 P(node) =0.2212121  
## class counts: 52 21  
## probabilities: 0.712 0.288   
## left son=24 (9 obs) right son=25 (64 obs)  
## Primary splits:  
## Vector < 14.06 to the right, improve=1.6990580, (0 missing)  
## age < 52.5 to the left, improve=0.9200012, (0 missing)  
## gender < 0.5 to the left, improve=0.2430256, (0 missing)  
##   
## Node number 13: 9 observations  
## predicted class=SARS-Cov-2 expected loss=0.2222222 P(node) =0.02727273  
## class counts: 2 7  
## probabilities: 0.222 0.778   
##   
## Node number 24: 9 observations  
## predicted class=No SARS-Cov-2 expected loss=0 P(node) =0.02727273  
## class counts: 9 0  
## probabilities: 1.000 0.000   
##   
## Node number 25: 64 observations, complexity param=0.02118644  
## predicted class=No SARS-Cov-2 expected loss=0.328125 P(node) =0.1939394  
## class counts: 43 21  
## probabilities: 0.672 0.328   
## left son=50 (55 obs) right son=51 (9 obs)  
## Primary splits:  
## Vector < 12.265 to the left, improve=1.0833960, (0 missing)  
## age < 52.5 to the left, improve=0.5104167, (0 missing)  
## gender < 0.5 to the left, improve=0.4667032, (0 missing)  
##   
## Node number 50: 55 observations, complexity param=0.02118644  
## predicted class=No SARS-Cov-2 expected loss=0.2909091 P(node) =0.1666667  
## class counts: 39 16  
## probabilities: 0.709 0.291   
## left son=100 (31 obs) right son=101 (24 obs)  
## Primary splits:  
## age < 44 to the left, improve=1.3468230, (0 missing)  
## Vector < 12.125 to the right, improve=0.5153772, (0 missing)  
## gender < 0.5 to the left, improve=0.4909091, (0 missing)  
## Surrogate splits:  
## Vector < 11.925 to the left, agree=0.582, adj=0.042, (0 split)  
## gender < 0.5 to the left, agree=0.582, adj=0.042, (0 split)  
##   
## Node number 51: 9 observations  
## predicted class=SARS-Cov-2 expected loss=0.4444444 P(node) =0.02727273  
## class counts: 4 5  
## probabilities: 0.444 0.556   
##   
## Node number 100: 31 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1935484 P(node) =0.09393939  
## class counts: 25 6  
## probabilities: 0.806 0.194   
##   
## Node number 101: 24 observations, complexity param=0.02118644  
## predicted class=No SARS-Cov-2 expected loss=0.4166667 P(node) =0.07272727  
## class counts: 14 10  
## probabilities: 0.583 0.417   
## left son=202 (16 obs) right son=203 (8 obs)  
## Primary splits:  
## gender < 0.5 to the left, improve=2.6666670, (0 missing)  
## age < 55.5 to the right, improve=2.0416670, (0 missing)  
## Vector < 11.835 to the left, improve=0.6666667, (0 missing)  
## Surrogate splits:  
## age < 48.5 to the right, agree=0.708, adj=0.125, (0 split)  
##   
## Node number 202: 16 observations  
## predicted class=No SARS-Cov-2 expected loss=0.25 P(node) =0.04848485  
## class counts: 12 4  
## probabilities: 0.750 0.250   
##   
## Node number 203: 8 observations  
## predicted class=SARS-Cov-2 expected loss=0.25 P(node) =0.02424242  
## class counts: 2 6  
## probabilities: 0.250 0.750

rpart.plot(x = Nab\_model\_vector, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)

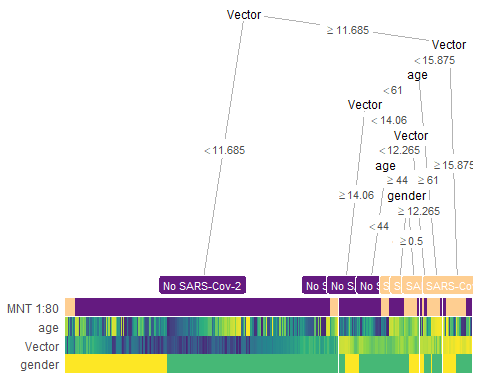


vector <- partykit::as.party(Nab\_model\_vector)  
heat\_tree(x = vector)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 3 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



class\_predicted\_vector <- predict(object = Nab\_model\_vector,   
 newdata = test\_base\_vector,   
 type = "class")  
confusionMatrix(data = class\_predicted\_vector,   
 reference = as.factor(test\_base\_vector$`MNT 1:80`))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 74 7  
## SARS-Cov-2 9 6  
##   
## Accuracy : 0.8333   
## 95% CI : (0.7435, 0.9016)  
## No Information Rate : 0.8646   
## P-Value [Acc > NIR] : 0.8515   
##   
## Kappa : 0.3316   
##   
## Mcnemar's Test P-Value : 0.8026   
##   
## Sensitivity : 0.8916   
## Specificity : 0.4615   
## Pos Pred Value : 0.9136   
## Neg Pred Value : 0.4000   
## Prevalence : 0.8646   
## Detection Rate : 0.7708   
## Detection Prevalence : 0.8438   
## Balanced Accuracy : 0.6766   
##   
## 'Positive' Class : No SARS-Cov-2   
##

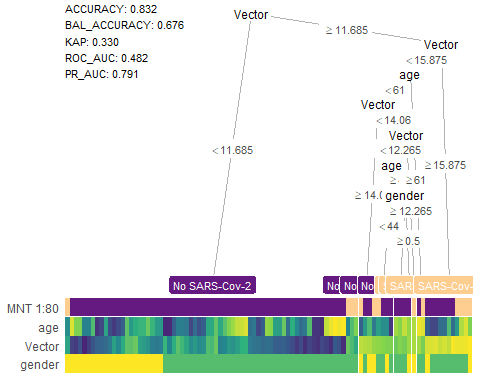
heat\_tree(  
 x = vector,  
 data\_test = test\_base\_vector)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 2 treated as  
## interval scaled

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 2 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



pred\_prob\_vector <-as.data.frame( predict(Nab\_model\_vector, test\_base\_vector, type="prob"))  
  
roc\_vector <- roc(test\_base\_vector$`MNT 1:80`, pred\_prob\_vector$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

r\_vec<- ggroc(roc\_vector, cex= 1.2, color = 'blue')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2() +  
 ggtitle("ELISA Vector") +  
 labs(x = "1 - Specificity",  
 y = "Sensitivity")

roc.list<-ggarrange(r\_ab, r\_gen, r\_vec, ncol = 3, nrow = 1)

# Возьмем в обучение все (ELISA, пол, возраст)

index <- sample(2, nrow(dataset\_80\_all), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_80\_all <- dataset\_80\_all[index==1, ]   
test\_base\_80\_all <- dataset\_80\_all[index == 2, ]   
  
table(train\_base\_80\_all$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 276 59

table(test\_base\_80\_all$`MNT 1:80`)

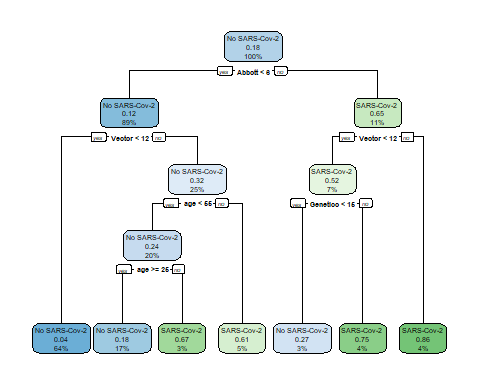
##   
## No SARS-Cov-2 SARS-Cov-2   
## 78 13

# Сделаем модель Дерево решений

Nab\_model\_80\_all <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_80\_all,   
 method = "class")  
summary(Nab\_model\_80\_all)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_80\_all, method = "class")  
## n=335 (864 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.18644068 0 1.0000000 1.0000000 0.1181697  
## 2 0.04237288 1 0.8135593 0.9830508 0.1173760  
## 3 0.03389831 3 0.7288136 1.0847458 0.1219552  
## 4 0.01000000 6 0.6101695 0.8305085 0.1096243  
##   
## Variable importance  
## Abbott Vector age Genetico gender   
## 41 31 19 8 1   
##   
## Node number 1: 335 observations, complexity param=0.1864407  
## predicted class=No SARS-Cov-2 expected loss=0.1761194 P(node) =1  
## class counts: 276 59  
## probabilities: 0.824 0.176   
## left son=2 (298 obs) right son=3 (37 obs)  
## Primary splits:  
## Abbott < 6.025 to the left, improve=18.5745200, (0 missing)  
## Vector < 11.64 to the left, improve=16.8909000, (3 missing)  
## Genetico < 10.7805 to the left, improve= 9.1421160, (2 missing)  
## age < 54.5 to the left, improve= 4.8583480, (0 missing)  
## gender < 0.5 to the left, improve= 0.5690343, (0 missing)  
## Surrogate splits:  
## Vector < 16.38 to the left, agree=0.901, adj=0.108, (0 split)  
##   
## Node number 2: 298 observations, complexity param=0.03389831  
## predicted class=No SARS-Cov-2 expected loss=0.1174497 P(node) =0.8895522  
## class counts: 263 35  
## probabilities: 0.883 0.117   
## left son=4 (214 obs) right son=5 (84 obs)  
## Primary splits:  
## Vector < 11.64 to the left, improve=9.68388900, (2 missing)  
## Genetico < 10.7805 to the left, improve=5.14843500, (2 missing)  
## Abbott < 1.83 to the left, improve=4.82501400, (0 missing)  
## age < 65.5 to the left, improve=3.56245300, (0 missing)  
## gender < 0.5 to the left, improve=0.02417945, (0 missing)  
## Surrogate splits:  
## Genetico < 23.006 to the left, agree=0.733, adj=0.060, (1 split)  
## Abbott < 4.47 to the left, agree=0.726, adj=0.036, (1 split)  
## age < 65.5 to the left, agree=0.720, adj=0.012, (0 split)  
##   
## Node number 3: 37 observations, complexity param=0.04237288  
## predicted class=SARS-Cov-2 expected loss=0.3513514 P(node) =0.1104478  
## class counts: 13 24  
## probabilities: 0.351 0.649   
## left son=6 (23 obs) right son=7 (14 obs)  
## Primary splits:  
## Vector < 12.265 to the left, improve=2.6755850, (1 missing)  
## Abbott < 6.67 to the right, improve=1.3728010, (0 missing)  
## age < 51.5 to the left, improve=1.1557740, (0 missing)  
## Genetico < 15.324 to the left, improve=0.8942766, (0 missing)  
## gender < 0.5 to the left, improve=0.3794847, (0 missing)  
## Surrogate splits:  
## Genetico < 25.79 to the left, agree=0.750, adj=0.308, (1 split)  
## Abbott < 6.08 to the right, agree=0.694, adj=0.154, (0 split)  
## age < 67 to the left, agree=0.694, adj=0.154, (0 split)  
##   
## Node number 4: 214 observations  
## predicted class=No SARS-Cov-2 expected loss=0.03738318 P(node) =0.638806  
## class counts: 206 8  
## probabilities: 0.963 0.037   
##   
## Node number 5: 84 observations, complexity param=0.03389831  
## predicted class=No SARS-Cov-2 expected loss=0.3214286 P(node) =0.2507463  
## class counts: 57 27  
## probabilities: 0.679 0.321   
## left son=10 (66 obs) right son=11 (18 obs)  
## Primary splits:  
## age < 54.5 to the left, improve=3.8448770000, (0 missing)  
## Genetico < 10.7805 to the left, improve=1.3452230000, (0 missing)  
## Abbott < 0.605 to the left, improve=1.0714290000, (0 missing)  
## Vector < 11.735 to the right, improve=0.8928571000, (0 missing)  
## gender < 0.5 to the right, improve=0.0001452785, (0 missing)  
##   
## Node number 6: 23 observations, complexity param=0.04237288  
## predicted class=SARS-Cov-2 expected loss=0.4782609 P(node) =0.06865672  
## class counts: 11 12  
## probabilities: 0.478 0.522   
## left son=12 (11 obs) right son=13 (12 obs)  
## Primary splits:  
## Genetico < 15.337 to the left, improve=2.6146250, (0 missing)  
## Abbott < 6.9 to the left, improve=1.0540180, (0 missing)  
## Vector < 11.34 to the right, improve=0.7461180, (0 missing)  
## age < 51.5 to the left, improve=0.6211180, (0 missing)  
## gender < 0.5 to the left, improve=0.2167224, (0 missing)  
## Surrogate splits:  
## age < 36 to the left, agree=0.696, adj=0.364, (0 split)  
## Abbott < 6.9 to the left, agree=0.652, adj=0.273, (0 split)  
## Vector < 10.9 to the left, agree=0.652, adj=0.273, (0 split)  
## gender < 0.5 to the right, agree=0.609, adj=0.182, (0 split)  
##   
## Node number 7: 14 observations  
## predicted class=SARS-Cov-2 expected loss=0.1428571 P(node) =0.04179104  
## class counts: 2 12  
## probabilities: 0.143 0.857   
##   
## Node number 10: 66 observations, complexity param=0.03389831  
## predicted class=No SARS-Cov-2 expected loss=0.2424242 P(node) =0.1970149  
## class counts: 50 16  
## probabilities: 0.758 0.242   
## left son=20 (57 obs) right son=21 (9 obs)  
## Primary splits:  
## age < 24.5 to the right, improve=3.7511960, (0 missing)  
## Vector < 15.885 to the left, improve=1.2314350, (0 missing)  
## Genetico < 20.5755 to the right, improve=0.7696132, (0 missing)  
## gender < 0.5 to the right, improve=0.3812820, (0 missing)  
## Abbott < 0.605 to the left, improve=0.2540353, (0 missing)  
##   
## Node number 11: 18 observations  
## predicted class=SARS-Cov-2 expected loss=0.3888889 P(node) =0.05373134  
## class counts: 7 11  
## probabilities: 0.389 0.611   
##   
## Node number 12: 11 observations  
## predicted class=No SARS-Cov-2 expected loss=0.2727273 P(node) =0.03283582  
## class counts: 8 3  
## probabilities: 0.727 0.273   
##   
## Node number 13: 12 observations  
## predicted class=SARS-Cov-2 expected loss=0.25 P(node) =0.0358209  
## class counts: 3 9  
## probabilities: 0.250 0.750   
##   
## Node number 20: 57 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1754386 P(node) =0.1701493  
## class counts: 47 10  
## probabilities: 0.825 0.175   
##   
## Node number 21: 9 observations  
## predicted class=SARS-Cov-2 expected loss=0.3333333 P(node) =0.02686567  
## class counts: 3 6  
## probabilities: 0.333 0.667

rpart.plot(x = Nab\_model\_80\_all, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)



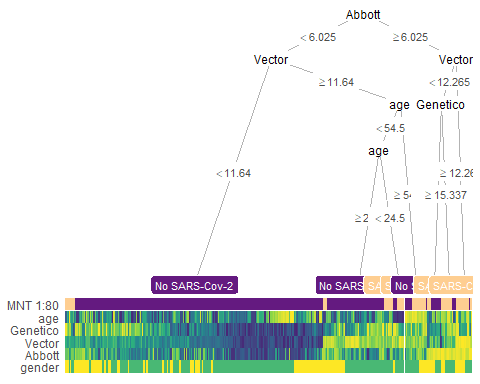
# Нарисуем тепловое дерево для сделанной модели

all <- partykit::as.party(Nab\_model\_80\_all)  
heat\_tree(x = all)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 5 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



# Посмотрим метрики

class\_predicted\_80\_all <- predict(object = Nab\_model\_80\_all,   
 newdata = test\_base\_80\_all,   
 type = "class")  
confusionMatrix(data = class\_predicted\_80\_all,   
 reference = as.factor(test\_base\_80\_all$`MNT 1:80`))

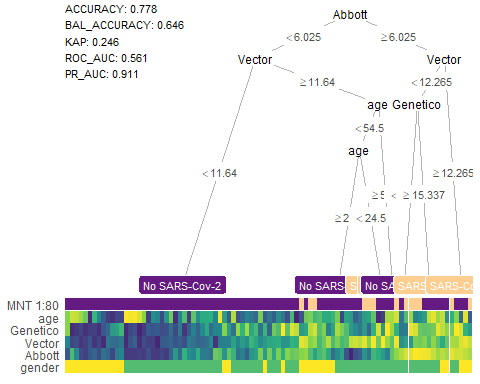
## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 65 7  
## SARS-Cov-2 13 6  
##   
## Accuracy : 0.7802   
## 95% CI : (0.6812, 0.8603)  
## No Information Rate : 0.8571   
## P-Value [Acc > NIR] : 0.9836   
##   
## Kappa : 0.2473   
##   
## Mcnemar's Test P-Value : 0.2636   
##   
## Sensitivity : 0.8333   
## Specificity : 0.4615   
## Pos Pred Value : 0.9028   
## Neg Pred Value : 0.3158   
## Prevalence : 0.8571   
## Detection Rate : 0.7143   
## Detection Prevalence : 0.7912   
## Balanced Accuracy : 0.6474   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = all,  
 data\_test = test\_base\_80\_all)

## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 4 treated as  
## interval scaled  
  
## Warning in cluster::daisy(., metric = "gower"): binary variable(s) 4 treated as  
## interval scaled

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



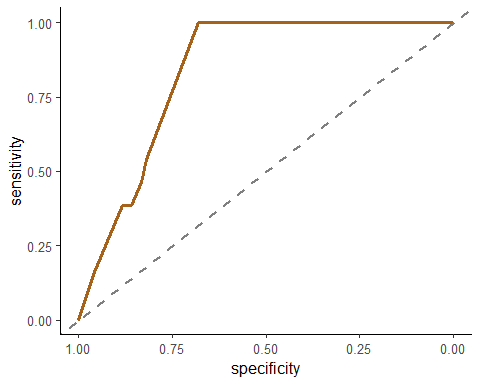
# Построим РОК-кривую для нашей модели

pred\_prob\_80\_all <-as.data.frame( predict(Nab\_model\_80\_all, test\_base\_80\_all, type="prob"))  
  
roc\_80\_all <- roc(test\_base\_80\_all$`MNT 1:80`, pred\_prob\_80\_all$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

ggroc(roc\_80\_all, color = '#A6611A', cex= 1.2 )+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()



##Попробуем убрать тест Генетико. Кажется, что он все портит. # Сделаем тестовую и тренировочную выборки

index <- sample(2, nrow(dataset\_av), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_av <- dataset\_av[index==1, ]   
test\_base\_av <- dataset\_av[index == 2, ]  
  
table(train\_base\_av$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 284 57

table(test\_base\_av$`MNT 1:80`)

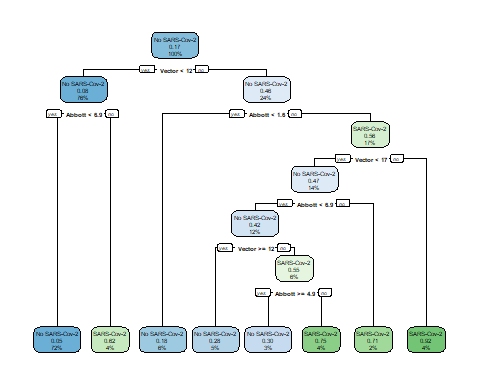
##   
## No SARS-Cov-2 SARS-Cov-2   
## 70 15

# Сделаем модель Дерево решений

Nab\_model\_av <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_av,   
 method = "class")  
summary(Nab\_model\_av)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_av, method = "class")  
## n=341 (879 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.06140351 0 1.0000000 1.0000000 0.1208772  
## 2 0.05263158 2 0.8771930 1.0175439 0.1217181  
## 3 0.01000000 7 0.6140351 0.9824561 0.1200230  
##   
## Variable importance  
## Vector Abbott   
## 59 41   
##   
## Node number 1: 341 observations, complexity param=0.06140351  
## predicted class=No SARS-Cov-2 expected loss=0.1671554 P(node) =1  
## class counts: 284 57  
## probabilities: 0.833 0.167   
## left son=2 (260 obs) right son=3 (81 obs)  
## Primary splits:  
## Vector < 11.835 to the left, improve=18.10092, (3 missing)  
## Abbott < 6.04 to the left, improve=16.85084, (0 missing)  
## Surrogate splits:  
## Abbott < 6.49 to the left, agree=0.769, adj=0.025, (3 split)  
##   
## Node number 2: 260 observations, complexity param=0.05263158  
## predicted class=No SARS-Cov-2 expected loss=0.07692308 P(node) =0.7624633  
## class counts: 240 20  
## probabilities: 0.923 0.077   
## left son=4 (247 obs) right son=5 (13 obs)  
## Primary splits:  
## Abbott < 6.92 to the left, improve=7.935223, (0 missing)  
## Vector < 10.45 to the left, improve=3.290267, (2 missing)  
##   
## Node number 3: 81 observations, complexity param=0.06140351  
## predicted class=No SARS-Cov-2 expected loss=0.4567901 P(node) =0.2375367  
## class counts: 44 37  
## probabilities: 0.543 0.457   
## left son=6 (22 obs) right son=7 (59 obs)  
## Primary splits:  
## Abbott < 1.64 to the left, improve=4.567331, (0 missing)  
## Vector < 16.635 to the left, improve=3.306250, (1 missing)  
##   
## Node number 4: 247 observations  
## predicted class=No SARS-Cov-2 expected loss=0.048583 P(node) =0.7243402  
## class counts: 235 12  
## probabilities: 0.951 0.049   
##   
## Node number 5: 13 observations  
## predicted class=SARS-Cov-2 expected loss=0.3846154 P(node) =0.03812317  
## class counts: 5 8  
## probabilities: 0.385 0.615   
##   
## Node number 6: 22 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1818182 P(node) =0.06451613  
## class counts: 18 4  
## probabilities: 0.818 0.182   
##   
## Node number 7: 59 observations, complexity param=0.05263158  
## predicted class=SARS-Cov-2 expected loss=0.440678 P(node) =0.1730205  
## class counts: 26 33  
## probabilities: 0.441 0.559   
## left son=14 (47 obs) right son=15 (12 obs)  
## Primary splits:  
## Vector < 16.635 to the left, improve=3.658421, (1 missing)  
## Abbott < 6.9 to the left, improve=2.794950, (0 missing)  
## Surrogate splits:  
## Abbott < 7.72 to the left, agree=0.81, adj=0.083, (1 split)  
##   
## Node number 14: 47 observations, complexity param=0.05263158  
## predicted class=No SARS-Cov-2 expected loss=0.4680851 P(node) =0.1378299  
## class counts: 25 22  
## probabilities: 0.532 0.468   
## left son=28 (40 obs) right son=29 (7 obs)  
## Primary splits:  
## Abbott < 6.86 to the left, improve=0.9971125, (0 missing)  
## Vector < 14.06 to the right, improve=0.9350164, (1 missing)  
##   
## Node number 15: 12 observations  
## predicted class=SARS-Cov-2 expected loss=0.08333333 P(node) =0.03519062  
## class counts: 1 11  
## probabilities: 0.083 0.917   
##   
## Node number 28: 40 observations, complexity param=0.05263158  
## predicted class=No SARS-Cov-2 expected loss=0.425 P(node) =0.1173021  
## class counts: 23 17  
## probabilities: 0.575 0.425   
## left son=56 (18 obs) right son=57 (22 obs)  
## Primary splits:  
## Vector < 12.315 to the right, improve=1.418687, (0 missing)  
## Abbott < 4.88 to the right, improve=0.550000, (0 missing)  
## Surrogate splits:  
## Abbott < 4.305 to the left, agree=0.6, adj=0.111, (0 split)  
##   
## Node number 29: 7 observations  
## predicted class=SARS-Cov-2 expected loss=0.2857143 P(node) =0.02052786  
## class counts: 2 5  
## probabilities: 0.286 0.714   
##   
## Node number 56: 18 observations  
## predicted class=No SARS-Cov-2 expected loss=0.2777778 P(node) =0.05278592  
## class counts: 13 5  
## probabilities: 0.722 0.278   
##   
## Node number 57: 22 observations, complexity param=0.05263158  
## predicted class=SARS-Cov-2 expected loss=0.4545455 P(node) =0.06451613  
## class counts: 10 12  
## probabilities: 0.455 0.545   
## left son=114 (10 obs) right son=115 (12 obs)  
## Primary splits:  
## Abbott < 4.88 to the right, improve=2.209091, (0 missing)  
## Vector < 12.05 to the left, improve=1.051948, (0 missing)  
## Surrogate splits:  
## Vector < 11.965 to the right, agree=0.636, adj=0.2, (0 split)  
##   
## Node number 114: 10 observations  
## predicted class=No SARS-Cov-2 expected loss=0.3 P(node) =0.02932551  
## class counts: 7 3  
## probabilities: 0.700 0.300   
##   
## Node number 115: 12 observations  
## predicted class=SARS-Cov-2 expected loss=0.25 P(node) =0.03519062  
## class counts: 3 9  
## probabilities: 0.250 0.750

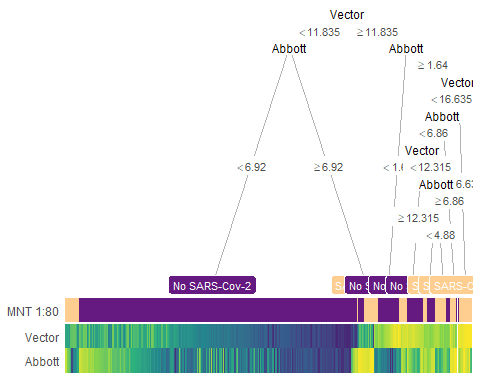
rpart.plot(x = Nab\_model\_av, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)



x\_av <- partykit::as.party(Nab\_model\_av)  
heat\_tree(x = x\_av)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.

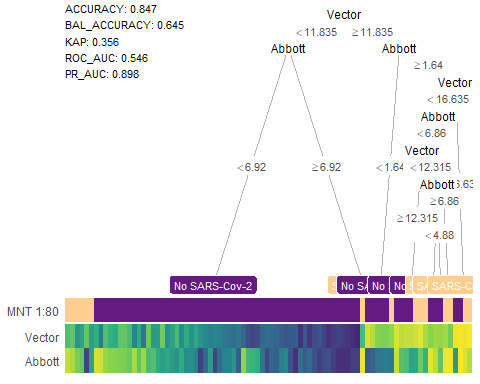


class\_predicted\_av <- predict(object = Nab\_model\_av,   
 newdata = test\_base\_av,   
 type = "class")  
confusionMatrix(data = class\_predicted\_av,   
 reference = as.factor(test\_base\_av$`MNT 1:80`))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 67 10  
## SARS-Cov-2 3 5  
##   
## Accuracy : 0.8471   
## 95% CI : (0.7527, 0.916)  
## No Information Rate : 0.8235   
## P-Value [Acc > NIR] : 0.34429   
##   
## Kappa : 0.3557   
##   
## Mcnemar's Test P-Value : 0.09609   
##   
## Sensitivity : 0.9571   
## Specificity : 0.3333   
## Pos Pred Value : 0.8701   
## Neg Pred Value : 0.6250   
## Prevalence : 0.8235   
## Detection Rate : 0.7882   
## Detection Prevalence : 0.9059   
## Balanced Accuracy : 0.6452   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = x\_av,  
 data\_test = test\_base\_av)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



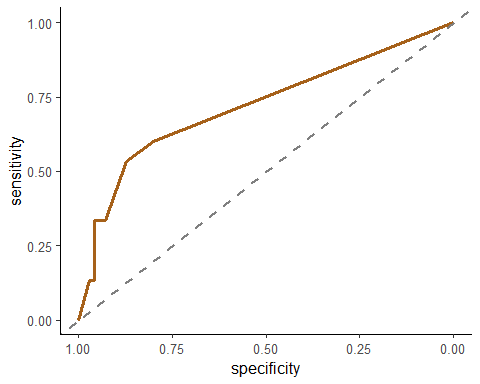
# Построим РОК-кривую

pred\_prob\_av <-as.data.frame( predict(Nab\_model\_av, test\_base\_av, type="prob"))  
  
roc\_av <- roc(test\_base\_av$`MNT 1:80`, pred\_prob\_av$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

ggroc(roc\_av, cex= 1.2, color = '#A6611A')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()



#А что если наша тест-системы работают лцучше на мужчинах? проверим эту гипотезу.

##Разделим на мальчиков и девочек

#Посмотрим для мальчиков

index <- sample(2, nrow(dataset\_male), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_male <- dataset\_male[index==1, ] # Train data  
test\_base\_male <- dataset\_male[index == 2, ] # Test data  
#check the ratio of pos/neg cases in the train/test data:  
table(train\_base\_male$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 95 25

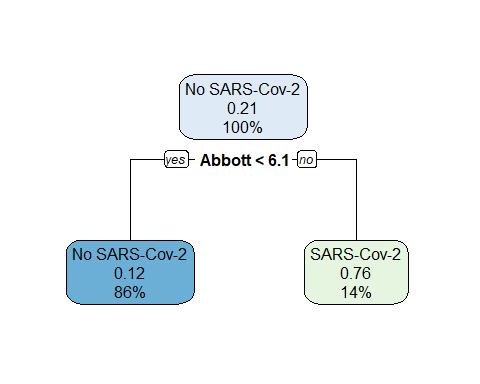
table(test\_base\_male$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 26 5

Nab\_model\_male <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_male,   
 method = "class")  
summary(Nab\_model\_male)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_male, method = "class")  
## n=120 (291 пропущенное наблюдение удалено)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.36 0 1.00 1.00 0.1779513  
## 2 0.01 1 0.64 0.92 0.1724722  
##   
## Variable importance  
## Abbott Vector age   
## 65 27 8   
##   
## Node number 1: 120 observations, complexity param=0.36  
## predicted class=No SARS-Cov-2 expected loss=0.2083333 P(node) =1  
## class counts: 95 25  
## probabilities: 0.792 0.208   
## left son=2 (103 obs) right son=3 (17 obs)  
## Primary splits:  
## Abbott < 6.11 to the left, improve=12.261800, (0 missing)  
## Vector < 11.94 to the left, improve=10.930040, (1 missing)  
## Genetico < 24.2385 to the left, improve= 3.611445, (1 missing)  
## age < 53 to the left, improve= 2.469336, (0 missing)  
## Surrogate splits:  
## Vector < 16.4 to the left, agree=0.917, adj=0.412, (0 split)  
## age < 66.5 to the left, agree=0.875, adj=0.118, (0 split)  
##   
## Node number 2: 103 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1165049 P(node) =0.8583333  
## class counts: 91 12  
## probabilities: 0.883 0.117   
##   
## Node number 3: 17 observations  
## predicted class=SARS-Cov-2 expected loss=0.2352941 P(node) =0.1416667  
## class counts: 4 13  
## probabilities: 0.235 0.765

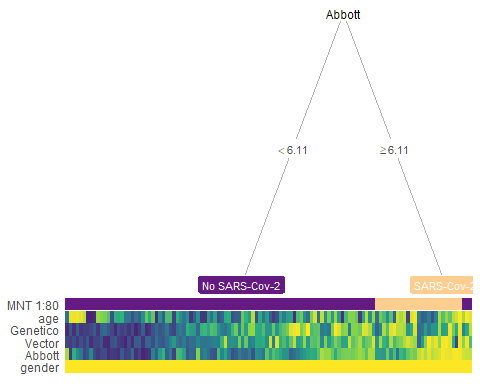
rpart.plot(x = Nab\_model\_male, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)



x\_male <- partykit::as.party(Nab\_model\_male)  
heat\_tree(x = x\_male)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.

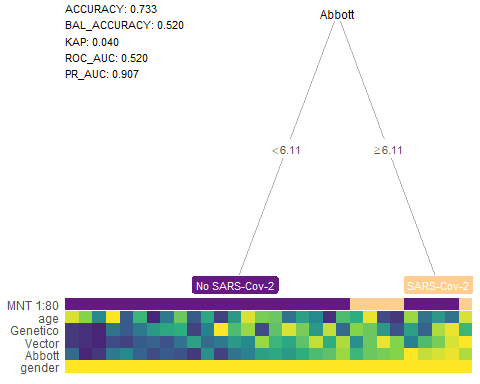


class\_predicted\_male <- predict(object = Nab\_model\_male,   
 newdata = test\_base\_male,   
 type = "class")  
confusionMatrix(data = class\_predicted\_male,   
 reference = as.factor(test\_base\_male$`MNT 1:80`))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 22 4  
## SARS-Cov-2 4 1  
##   
## Accuracy : 0.7419   
## 95% CI : (0.5539, 0.8814)  
## No Information Rate : 0.8387   
## P-Value [Acc > NIR] : 0.9488   
##   
## Kappa : 0.0462   
##   
## Mcnemar's Test P-Value : 1.0000   
##   
## Sensitivity : 0.8462   
## Specificity : 0.2000   
## Pos Pred Value : 0.8462   
## Neg Pred Value : 0.2000   
## Prevalence : 0.8387   
## Detection Rate : 0.7097   
## Detection Prevalence : 0.8387   
## Balanced Accuracy : 0.5231   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = x\_male,  
 data\_test = test\_base\_male)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.



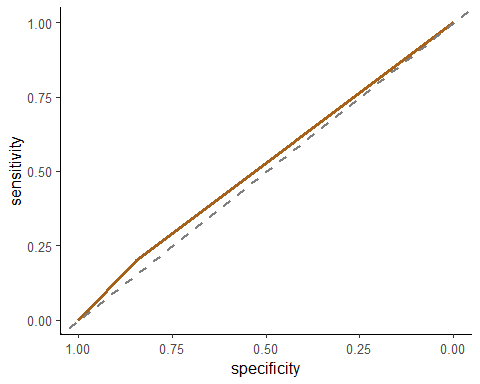
# Построим РОК-кривую

pred\_prob\_male <-as.data.frame( predict(Nab\_model\_male, test\_base\_male, type="prob"))  
# ROC value  
roc\_male <- roc(test\_base\_male$`MNT 1:80`, pred\_prob\_male$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

ggroc(roc\_male, cex= 1.2, color = '#A6611A')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()



#Посмотрим для девочек

index <- sample(2, nrow(dataset\_female), prob = c(0.8, 0.2), replace = TRUE)  
  
train\_base\_female <- dataset\_female[index==1, ]   
test\_base\_female <- dataset\_female[index == 2, ]  
  
table(train\_base\_female$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 184 35

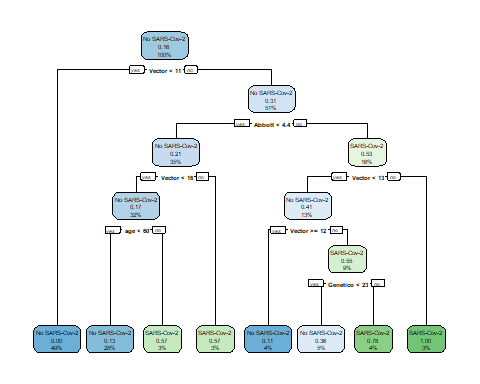
table(test\_base\_female$`MNT 1:80`)

##   
## No SARS-Cov-2 SARS-Cov-2   
## 49 7

Nab\_model\_female <- rpart(formula = `MNT 1:80` ~.,   
 data = train\_base\_female,   
 method = "class")  
summary(Nab\_model\_female)

## Call:  
## rpart(formula = `MNT 1:80` ~ ., data = train\_base\_female, method = "class")  
## n=219 (576 пропущенных наблюдений удалены)  
##   
## CP nsplit rel error xerror xstd  
## 1 0.06666667 0 1.0000000 1.000000 0.1549362  
## 2 0.02857143 5 0.6571429 1.028571 0.1567064  
## 3 0.01000000 7 0.6000000 1.114286 0.1617627  
##   
## Variable importance  
## Vector Abbott Genetico age   
## 43 23 22 11   
##   
## Node number 1: 219 observations, complexity param=0.06666667  
## predicted class=No SARS-Cov-2 expected loss=0.1598174 P(node) =1  
## class counts: 184 35  
## probabilities: 0.840 0.160   
## left son=2 (107 obs) right son=3 (112 obs)  
## Primary splits:  
## Vector < 10.82 to the left, improve=10.636470, (1 missing)  
## Abbott < 5.35 to the left, improve= 9.315488, (0 missing)  
## Genetico < 12.761 to the left, improve= 5.397592, (2 missing)  
## age < 45.5 to the left, improve= 2.542254, (0 missing)  
## Surrogate splits:  
## Genetico < 12.1685 to the left, agree=0.849, adj=0.689, (0 split)  
## Abbott < 0.445 to the left, agree=0.752, adj=0.491, (1 split)  
## age < 34.5 to the left, agree=0.569, adj=0.113, (0 split)  
##   
## Node number 2: 107 observations  
## predicted class=No SARS-Cov-2 expected loss=0 P(node) =0.4885845  
## class counts: 107 0  
## probabilities: 1.000 0.000   
##   
## Node number 3: 112 observations, complexity param=0.06666667  
## predicted class=No SARS-Cov-2 expected loss=0.3125 P(node) =0.5114155  
## class counts: 77 35  
## probabilities: 0.688 0.312   
## left son=6 (76 obs) right son=7 (36 obs)  
## Primary splits:  
## Abbott < 4.43 to the left, improve=4.917398, (0 missing)  
## Vector < 16.405 to the left, improve=4.197007, (0 missing)  
## age < 61 to the left, improve=3.371667, (0 missing)  
## Genetico < 16.9165 to the right, improve=1.144048, (0 missing)  
## Surrogate splits:  
## age < 54.5 to the left, agree=0.714, adj=0.111, (0 split)  
## Genetico < 25.869 to the left, agree=0.696, adj=0.056, (0 split)  
##   
## Node number 6: 76 observations, complexity param=0.02857143  
## predicted class=No SARS-Cov-2 expected loss=0.2105263 P(node) =0.347032  
## class counts: 60 16  
## probabilities: 0.789 0.211   
## left son=12 (69 obs) right son=13 (7 obs)  
## Primary splits:  
## Vector < 16.405 to the left, improve=2.008500, (0 missing)  
## age < 59.5 to the left, improve=2.008500, (0 missing)  
## Genetico < 16.9165 to the right, improve=1.090539, (0 missing)  
## Abbott < 2.64 to the right, improve=0.340599, (0 missing)  
##   
## Node number 7: 36 observations, complexity param=0.06666667  
## predicted class=SARS-Cov-2 expected loss=0.4722222 P(node) =0.1643836  
## class counts: 17 19  
## probabilities: 0.472 0.528   
## left son=14 (29 obs) right son=15 (7 obs)  
## Primary splits:  
## Vector < 12.78 to the left, improve=3.8754790, (0 missing)  
## age < 40 to the left, improve=1.5873020, (0 missing)  
## Abbott < 6.92 to the left, improve=0.9920635, (0 missing)  
## Genetico < 23.737 to the left, improve=0.6067821, (0 missing)  
##   
## Node number 12: 69 observations, complexity param=0.02857143  
## predicted class=No SARS-Cov-2 expected loss=0.173913 P(node) =0.3150685  
## class counts: 57 12  
## probabilities: 0.826 0.174   
## left son=24 (62 obs) right son=25 (7 obs)  
## Primary splits:  
## age < 59.5 to the left, improve=2.4620320, (0 missing)  
## Abbott < 2.64 to the right, improve=0.8648625, (0 missing)  
## Genetico < 25.79 to the left, improve=0.7318247, (0 missing)  
## Vector < 11.67 to the left, improve=0.6038647, (0 missing)  
##   
## Node number 13: 7 observations  
## predicted class=SARS-Cov-2 expected loss=0.4285714 P(node) =0.03196347  
## class counts: 3 4  
## probabilities: 0.429 0.571   
##   
## Node number 14: 29 observations, complexity param=0.06666667  
## predicted class=No SARS-Cov-2 expected loss=0.4137931 P(node) =0.1324201  
## class counts: 17 12  
## probabilities: 0.586 0.414   
## left son=28 (9 obs) right son=29 (20 obs)  
## Primary splits:  
## Vector < 11.88 to the right, improve=2.391188, (0 missing)  
## Abbott < 6.92 to the left, improve=1.915119, (0 missing)  
## Genetico < 23.737 to the left, improve=1.915119, (0 missing)  
## age < 42.5 to the left, improve=1.354680, (0 missing)  
## Surrogate splits:  
## Genetico < 26.089 to the right, agree=0.759, adj=0.222, (0 split)  
## age < 67 to the right, agree=0.759, adj=0.222, (0 split)  
##   
## Node number 15: 7 observations  
## predicted class=SARS-Cov-2 expected loss=0 P(node) =0.03196347  
## class counts: 0 7  
## probabilities: 0.000 1.000   
##   
## Node number 24: 62 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1290323 P(node) =0.283105  
## class counts: 54 8  
## probabilities: 0.871 0.129   
##   
## Node number 25: 7 observations  
## predicted class=SARS-Cov-2 expected loss=0.4285714 P(node) =0.03196347  
## class counts: 3 4  
## probabilities: 0.429 0.571   
##   
## Node number 28: 9 observations  
## predicted class=No SARS-Cov-2 expected loss=0.1111111 P(node) =0.04109589  
## class counts: 8 1  
## probabilities: 0.889 0.111   
##   
## Node number 29: 20 observations, complexity param=0.06666667  
## predicted class=SARS-Cov-2 expected loss=0.45 P(node) =0.0913242  
## class counts: 9 11  
## probabilities: 0.450 0.550   
## left son=58 (11 obs) right son=59 (9 obs)  
## Primary splits:  
## Genetico < 23.1065 to the left, improve=1.6979800, (0 missing)  
## Vector < 11.72 to the left, improve=0.5813187, (0 missing)  
## Abbott < 6.92 to the left, improve=0.3646465, (0 missing)  
## age < 51.5 to the left, improve=0.1000000, (0 missing)  
## Surrogate splits:  
## Vector < 11.61 to the left, agree=0.70, adj=0.333, (0 split)  
## Abbott < 7.985 to the left, agree=0.65, adj=0.222, (0 split)  
## age < 34.5 to the right, agree=0.65, adj=0.222, (0 split)  
##   
## Node number 58: 11 observations  
## predicted class=No SARS-Cov-2 expected loss=0.3636364 P(node) =0.05022831  
## class counts: 7 4  
## probabilities: 0.636 0.364   
##   
## Node number 59: 9 observations  
## predicted class=SARS-Cov-2 expected loss=0.2222222 P(node) =0.04109589  
## class counts: 2 7  
## probabilities: 0.222 0.778

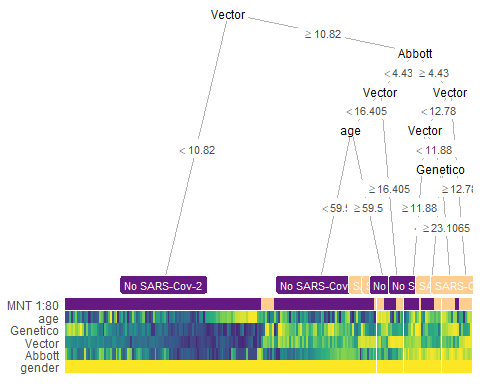
rpart.plot(x = Nab\_model\_female, yesno = 2,   
 fallen.leaves = TRUE, clip.right.labs = T)



x\_female <- partykit::as.party(Nab\_model\_female)  
heat\_tree(x = x\_female)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.

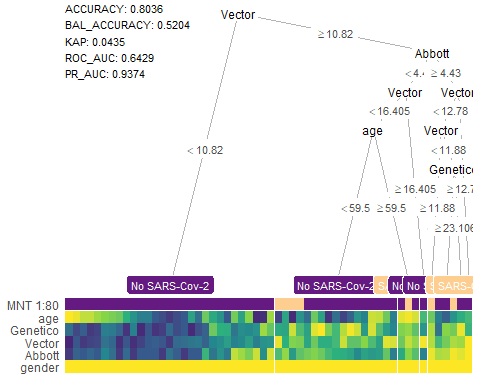


class\_predicted\_female <- predict(object = Nab\_model\_female,   
 newdata = test\_base\_female,   
 type = "class")  
confusionMatrix(data = class\_predicted\_female,   
 reference = as.factor(test\_base\_female$`MNT 1:80`))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 44 6  
## SARS-Cov-2 5 1  
##   
## Accuracy : 0.8036   
## 95% CI : (0.6757, 0.8977)  
## No Information Rate : 0.875   
## P-Value [Acc > NIR] : 0.9583   
##   
## Kappa : 0.0435   
##   
## Mcnemar's Test P-Value : 1.0000   
##   
## Sensitivity : 0.8980   
## Specificity : 0.1429   
## Pos Pred Value : 0.8800   
## Neg Pred Value : 0.1667   
## Prevalence : 0.8750   
## Detection Rate : 0.7857   
## Detection Prevalence : 0.8929   
## Balanced Accuracy : 0.5204   
##   
## 'Positive' Class : No SARS-Cov-2   
##

heat\_tree(  
 x = x\_female,  
 data\_test = test\_base\_female)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.  
  
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.

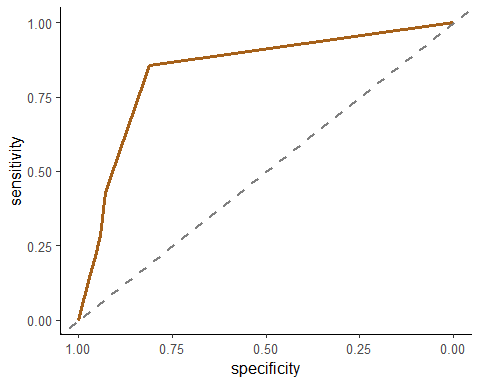


pred\_prob\_80 <-as.data.frame( predict(Nab\_model\_80, test\_base\_80, type="prob"))  
  
roc\_80 <- roc(test\_base\_80$`MNT 1:80`, pred\_prob\_80$`SARS-Cov-2`)

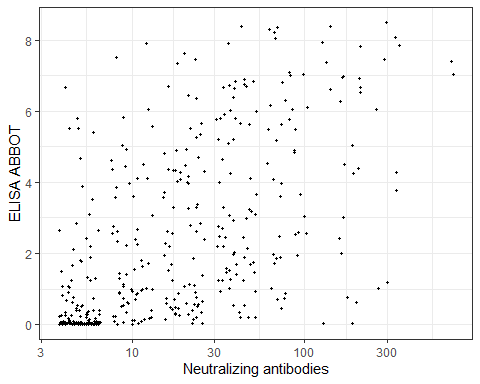
## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

ggroc(roc\_80, cex= 1.2, color = '#A6611A')+  
 geom\_abline(intercept = 1, slope = 1, color = "grey50", size = 0.8, linetype = 2)+  
 theme\_classic2()

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ #Визуально оценим корреляцию признаков

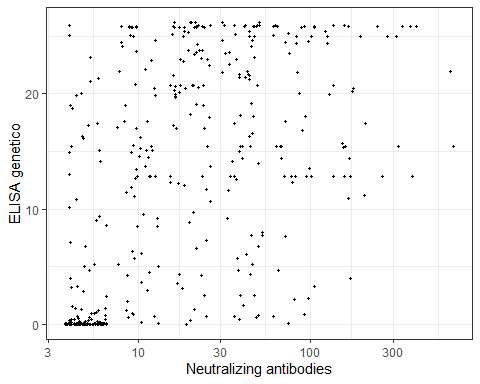
datagg = data\_ab\_only  
  
ggplot(datagg, aes(y = IgG\_testB\_quantitative, x = neutralizing\_antibodies\_quantitative)) +  
 geom\_jitter(size = 0.8) +  
 ylab("ELISA ABBOT") +   
 xlab("Neutralizing antibodies")+  
 scale\_x\_log10()+  
 theme\_bw()



scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")

## <ScaleContinuousPosition>  
## Range:   
## Limits: 0 -- 1

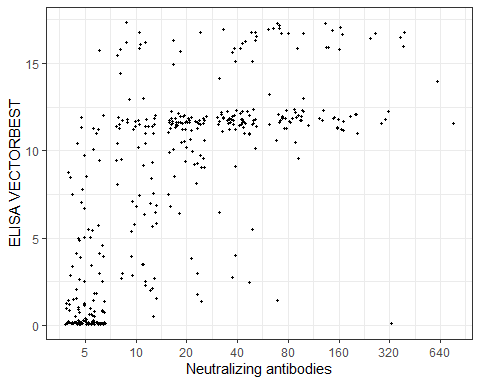
ggplot(datagg, aes(y = IgA\_or\_G\_or\_M\_testC\_quantitative, x = neutralizing\_antibodies\_quantitative)) +  
 geom\_jitter(size = 0.8) +  
 ylab("ELISA genetico") +   
 xlab("Neutralizing antibodies")+  
 scale\_x\_log10()+  
 theme\_bw()



scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")

## <ScaleContinuousPosition>  
## Range:   
## Limits: 0 -- 1

ggplot(datagg, aes(y = IgG\_or\_M\_testD\_quantitative, x = neutralizing\_antibodies\_quantitative)) +  
 geom\_jitter(size = 0.8) +  
 ylab("ELISA VECTORBEST") +   
 xlab("Neutralizing antibodies")+  
 theme\_bw() +  
 scale\_x\_continuous(breaks = c(5, 10, 20, 40, 80, 160, 320, 640), trans = "log2")



#Проанализируем данные с помощью линейной регрессии

lm\_vector = lm(Nab~Vector,data\_for\_ab)  
  
summary(lm\_vector)

##   
## Call:  
## lm(formula = Nab ~ Vector, data = data\_for\_ab)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -73.55 -34.13 -12.29 5.20 584.65   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.7109 5.5643 -0.128 0.898   
## Vector 4.8535 0.5672 8.557 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 64.72 on 421 degrees of freedom  
## (1087 пропущенных наблюдений удалены)  
## Multiple R-squared: 0.1482, Adjusted R-squared: 0.1461   
## F-statistic: 73.23 on 1 and 421 DF, p-value: < 2.2e-16

lm\_abbott =lm(Nab~Abbott, data\_for\_ab)  
  
summary(lm\_abbott)

##   
## Call:  
## lm(formula = Nab ~ Abbott, data = data\_for\_ab)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -96.31 -23.55 -5.36 1.68 544.45   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.651 4.219 2.288 0.0226 \*   
## Abbott 12.219 1.239 9.859 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 63.08 on 424 degrees of freedom  
## (1084 пропущенных наблюдений удалены)  
## Multiple R-squared: 0.1865, Adjusted R-squared: 0.1846   
## F-statistic: 97.2 on 1 and 424 DF, p-value: < 2.2e-16

lm\_genetico = lm(Nab~Genetico, data\_for\_ab)  
  
summary(lm\_genetico)

##   
## Call:  
## lm(formula = Nab ~ Genetico, data = data\_for\_ab)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -62.15 -30.08 -9.01 -3.71 594.64   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.4086 5.1304 2.614 0.00928 \*\*   
## Genetico 2.0744 0.3271 6.342 5.84e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 67 on 421 degrees of freedom  
## (1087 пропущенных наблюдений удалены)  
## Multiple R-squared: 0.08721, Adjusted R-squared: 0.08505   
## F-statistic: 40.23 on 1 and 421 DF, p-value: 5.843e-10

#Создадим модель с помощью randomForest на основе 3 тест систем

datasetrf<-test\_nab  
  
set.seed(100)  
  
ind <- sample(2, nrow(datasetrf), replace = TRUE, prob = c(0.8, 0.2))  
  
train <- datasetrf[ind==1,]  
test <- datasetrf[ind==2,]  
  
rf <- randomForest(`MNT 1:80` ~ vector+abbott+genetico, data=train, proximity=TRUE)  
  
print(rf)

##   
## Call:  
## randomForest(formula = `MNT 1:80` ~ vector + abbott + genetico, data = train, proximity = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 1  
##   
## OOB estimate of error rate: 14.97%  
## Confusion matrix:  
## No SARS-Cov-2 SARS-Cov-2 class.error  
## No SARS-Cov-2 276 17 0.05802048  
## SARS-Cov-2 36 25 0.59016393

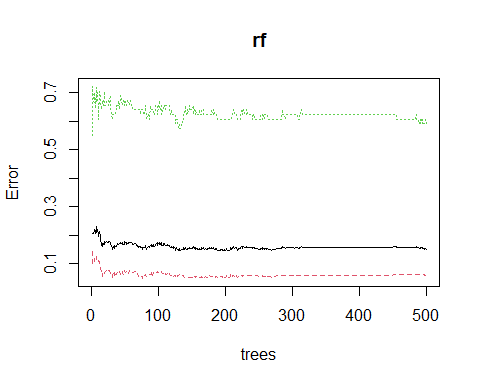
p1 <- predict(rf, train)  
  
confusionMatrix(p1, train$ `MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 293 0  
## SARS-Cov-2 0 61  
##   
## Accuracy : 1   
## 95% CI : (0.9896, 1)   
## No Information Rate : 0.8277   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 1   
##   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.8277   
## Detection Rate : 0.8277   
## Detection Prevalence : 0.8277   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : No SARS-Cov-2  
##

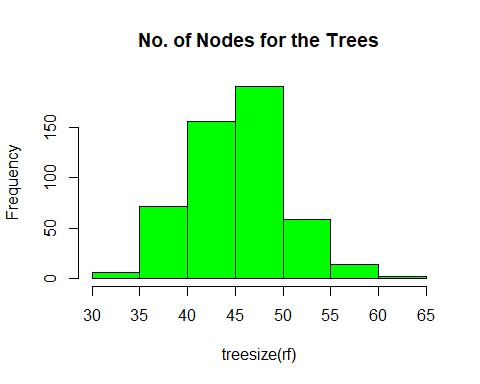
p2 <- predict(rf, test)  
  
confusionMatrix(p2, test$`MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 71 5  
## SARS-Cov-2 4 5  
##   
## Accuracy : 0.8941   
## 95% CI : (0.8085, 0.9504)  
## No Information Rate : 0.8824   
## P-Value [Acc > NIR] : 0.45   
##   
## Kappa : 0.4669   
##   
## Mcnemar's Test P-Value : 1.00   
##   
## Sensitivity : 0.9467   
## Specificity : 0.5000   
## Pos Pred Value : 0.9342   
## Neg Pred Value : 0.5556   
## Prevalence : 0.8824   
## Detection Rate : 0.8353   
## Detection Prevalence : 0.8941   
## Balanced Accuracy : 0.7233   
##   
## 'Positive' Class : No SARS-Cov-2   
##

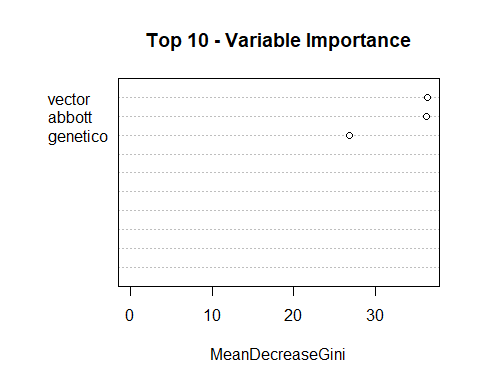
plot(rf)



hist(treesize(rf),  
 main = "No. of Nodes for the Trees",  
 col = "green")



varImpPlot(rf,  
 sort = T,  
 n.var = 10,  
 main = "Top 10 - Variable Importance")



importance(rf)

## MeanDecreaseGini  
## vector 36.36597  
## abbott 36.21787  
## genetico 26.79918

pred\_prob\_rf <-as.data.frame( predict(rf, test, type="prob"))  
  
roc\_rf <- pROC::roc(test$`MNT 1:80`, pred\_prob\_rf$`SARS-Cov-2`)

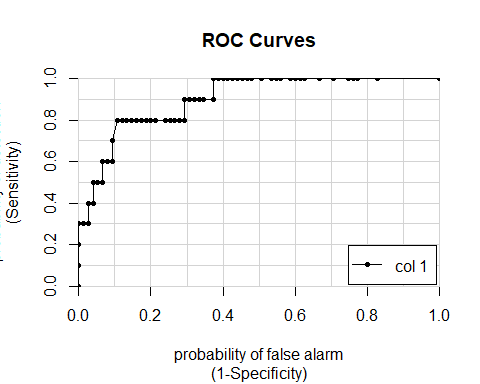
## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

roc\_rf

##   
## Call:  
## roc.default(response = test$`MNT 1:80`, predictor = pred\_prob\_rf$`SARS-Cov-2`)  
##   
## Data: pred\_prob\_rf$`SARS-Cov-2` in 75 controls (test$`MNT 1:80` No SARS-Cov-2) < 10 cases (test$`MNT 1:80` SARS-Cov-2).  
## Area under the curve: 0.9007

caTools::colAUC(pred\_prob\_rf$`SARS-Cov-2`, test$`MNT 1:80`, plotROC = T)



## [,1]  
## No SARS-Cov-2 vs. SARS-Cov-2 0.9006667

#Проделаем тоже самое с каждой тест-системой отдельно

set.seed(100)  
  
ind <- sample(2, nrow(datasetrf), replace = TRUE, prob = c(0.8, 0.2))  
  
train <- datasetrf[ind==1,]  
test <- datasetrf[ind==2,]  
  
rf\_vector <- randomForest(`MNT 1:80`~ vector, data=train, proximity=TRUE)  
  
print(rf\_vector)

##   
## Call:  
## randomForest(formula = `MNT 1:80` ~ vector, data = train, proximity = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 1  
##   
## OOB estimate of error rate: 21.47%  
## Confusion matrix:  
## No SARS-Cov-2 SARS-Cov-2 class.error  
## No SARS-Cov-2 261 32 0.1092150  
## SARS-Cov-2 44 17 0.7213115

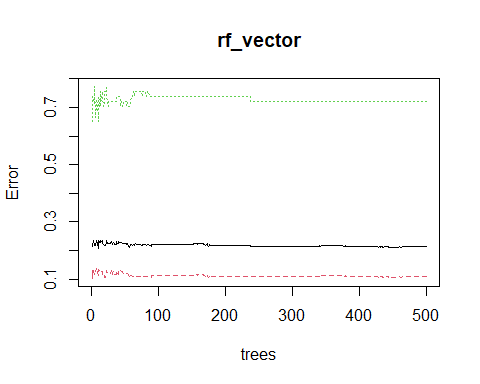
p1\_vector <- predict(rf\_vector, train)  
  
confusionMatrix(p1\_vector, train$ `MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 293 22  
## SARS-Cov-2 0 39  
##   
## Accuracy : 0.9379   
## 95% CI : (0.9074, 0.9606)  
## No Information Rate : 0.8277   
## P-Value [Acc > NIR] : 6.756e-10   
##   
## Kappa : 0.7458   
##   
## Mcnemar's Test P-Value : 7.562e-06   
##   
## Sensitivity : 1.0000   
## Specificity : 0.6393   
## Pos Pred Value : 0.9302   
## Neg Pred Value : 1.0000   
## Prevalence : 0.8277   
## Detection Rate : 0.8277   
## Detection Prevalence : 0.8898   
## Balanced Accuracy : 0.8197   
##   
## 'Positive' Class : No SARS-Cov-2   
##

p2\_vector <- predict(rf\_vector, test)  
  
confusionMatrix(p2\_vector, test$`MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 67 2  
## SARS-Cov-2 8 8  
##   
## Accuracy : 0.8824   
## 95% CI : (0.7943, 0.9421)  
## No Information Rate : 0.8824   
## P-Value [Acc > NIR] : 0.5832   
##   
## Kappa : 0.5503   
##   
## Mcnemar's Test P-Value : 0.1138   
##   
## Sensitivity : 0.8933   
## Specificity : 0.8000   
## Pos Pred Value : 0.9710   
## Neg Pred Value : 0.5000   
## Prevalence : 0.8824   
## Detection Rate : 0.7882   
## Detection Prevalence : 0.8118   
## Balanced Accuracy : 0.8467   
##   
## 'Positive' Class : No SARS-Cov-2   
##

plot(rf\_vector)



pred\_prob\_rf\_vector <-as.data.frame( predict(rf\_vector, test, type="prob"))  
  
roc\_rf\_vector <- pROC::roc(test$`MNT 1:80`, pred\_prob\_rf\_vector$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

roc\_rf\_vector

##   
## Call:  
## roc.default(response = test$`MNT 1:80`, predictor = pred\_prob\_rf\_vector$`SARS-Cov-2`)  
##   
## Data: pred\_prob\_rf\_vector$`SARS-Cov-2` in 75 controls (test$`MNT 1:80` No SARS-Cov-2) < 10 cases (test$`MNT 1:80` SARS-Cov-2).  
## Area under the curve: 0.8867

set.seed(100)  
  
ind <- sample(2, nrow(datasetrf), replace = TRUE, prob = c(0.8, 0.2))  
  
train <- datasetrf[ind==1,]  
test <- datasetrf[ind==2,]  
  
rf\_abbott <- randomForest(`MNT 1:80`~ abbott, data=train, proximity=TRUE)  
  
print(rf\_abbott)

##   
## Call:  
## randomForest(formula = `MNT 1:80` ~ abbott, data = train, proximity = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 1  
##   
## OOB estimate of error rate: 21.47%  
## Confusion matrix:  
## No SARS-Cov-2 SARS-Cov-2 class.error  
## No SARS-Cov-2 262 31 0.1058020  
## SARS-Cov-2 45 16 0.7377049

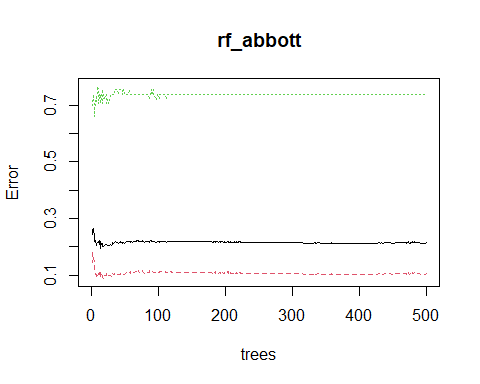
p1\_abbott <- predict(rf\_abbott, train)  
  
confusionMatrix(p1\_abbott, train$ `MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 293 14  
## SARS-Cov-2 0 47  
##   
## Accuracy : 0.9605   
## 95% CI : (0.9345, 0.9782)  
## No Information Rate : 0.8277   
## P-Value [Acc > NIR] : 1.284e-14   
##   
## Kappa : 0.8475   
##   
## Mcnemar's Test P-Value : 0.000512   
##   
## Sensitivity : 1.0000   
## Specificity : 0.7705   
## Pos Pred Value : 0.9544   
## Neg Pred Value : 1.0000   
## Prevalence : 0.8277   
## Detection Rate : 0.8277   
## Detection Prevalence : 0.8672   
## Balanced Accuracy : 0.8852   
##   
## 'Positive' Class : No SARS-Cov-2   
##

p2\_abbott <- predict(rf\_abbott, test)  
  
confusionMatrix(p2\_abbott, test$`MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 67 8  
## SARS-Cov-2 8 2  
##   
## Accuracy : 0.8118   
## 95% CI : (0.7124, 0.8884)  
## No Information Rate : 0.8824   
## P-Value [Acc > NIR] : 0.9805   
##   
## Kappa : 0.0933   
##   
## Mcnemar's Test P-Value : 1.0000   
##   
## Sensitivity : 0.8933   
## Specificity : 0.2000   
## Pos Pred Value : 0.8933   
## Neg Pred Value : 0.2000   
## Prevalence : 0.8824   
## Detection Rate : 0.7882   
## Detection Prevalence : 0.8824   
## Balanced Accuracy : 0.5467   
##   
## 'Positive' Class : No SARS-Cov-2   
##

plot(rf\_abbott)



pred\_prob\_rf\_abbott <-as.data.frame( predict(rf\_abbott, test, type="prob"))  
  
# ROC value  
roc\_rf\_abbott <- pROC::roc(test$`MNT 1:80`, pred\_prob\_rf\_abbott$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

## Setting direction: controls < cases

roc\_rf\_abbott

##   
## Call:  
## roc.default(response = test$`MNT 1:80`, predictor = pred\_prob\_rf\_abbott$`SARS-Cov-2`)  
##   
## Data: pred\_prob\_rf\_abbott$`SARS-Cov-2` in 75 controls (test$`MNT 1:80` No SARS-Cov-2) < 10 cases (test$`MNT 1:80` SARS-Cov-2).  
## Area under the curve: 0.732

set.seed(100)  
  
ind <- sample(2, nrow(datasetrf), replace = TRUE, prob = c(0.8, 0.2))  
  
train <- datasetrf[ind==1,]  
test <- datasetrf[ind==2,]  
  
rf\_genetico <- randomForest(`MNT 1:80`~ genetico, data=train, proximity=TRUE)  
  
print(rf\_genetico)

##   
## Call:  
## randomForest(formula = `MNT 1:80` ~ genetico, data = train, proximity = TRUE)   
## Type of random forest: classification  
## Number of trees: 500  
## No. of variables tried at each split: 1  
##   
## OOB estimate of error rate: 25.71%  
## Confusion matrix:  
## No SARS-Cov-2 SARS-Cov-2 class.error  
## No SARS-Cov-2 251 42 0.1433447  
## SARS-Cov-2 49 12 0.8032787

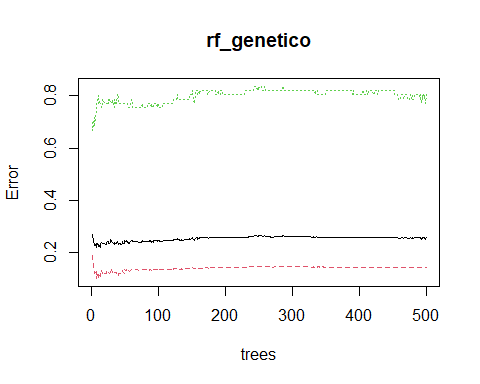
p1\_genetico <- predict(rf\_genetico, train)  
  
confusionMatrix(p1\_genetico, train$ `MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 287 18  
## SARS-Cov-2 6 43  
##   
## Accuracy : 0.9322   
## 95% CI : (0.9008, 0.9561)  
## No Information Rate : 0.8277   
## P-Value [Acc > NIR] : 6.094e-09   
##   
## Kappa : 0.7422   
##   
## Mcnemar's Test P-Value : 0.02474   
##   
## Sensitivity : 0.9795   
## Specificity : 0.7049   
## Pos Pred Value : 0.9410   
## Neg Pred Value : 0.8776   
## Prevalence : 0.8277   
## Detection Rate : 0.8107   
## Detection Prevalence : 0.8616   
## Balanced Accuracy : 0.8422   
##   
## 'Positive' Class : No SARS-Cov-2   
##

p2\_genetico <- predict(rf\_genetico, test)  
  
confusionMatrix(p2\_genetico, test$`MNT 1:80`)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No SARS-Cov-2 SARS-Cov-2  
## No SARS-Cov-2 65 10  
## SARS-Cov-2 10 0  
##   
## Accuracy : 0.7647   
## 95% CI : (0.6603, 0.85)  
## No Information Rate : 0.8824   
## P-Value [Acc > NIR] : 0.9993   
##   
## Kappa : -0.1333   
##   
## Mcnemar's Test P-Value : 1.0000   
##   
## Sensitivity : 0.8667   
## Specificity : 0.0000   
## Pos Pred Value : 0.8667   
## Neg Pred Value : 0.0000   
## Prevalence : 0.8824   
## Detection Rate : 0.7647   
## Detection Prevalence : 0.8824   
## Balanced Accuracy : 0.4333   
##   
## 'Positive' Class : No SARS-Cov-2   
##

plot(rf\_genetico)



pred\_prob\_rf\_genetico <-as.data.frame( predict(rf\_genetico, test, type="prob"))  
  
roc\_rf\_genetico <- pROC::roc(test$`MNT 1:80`, pred\_prob\_rf\_genetico$`SARS-Cov-2`)

## Setting levels: control = No SARS-Cov-2, case = SARS-Cov-2

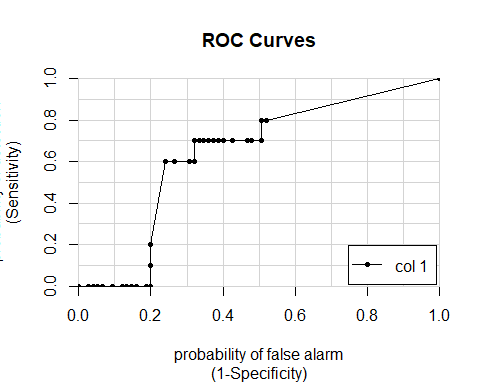
## Setting direction: controls < cases

roc\_rf\_genetico

##   
## Call:  
## roc.default(response = test$`MNT 1:80`, predictor = pred\_prob\_rf\_genetico$`SARS-Cov-2`)  
##   
## Data: pred\_prob\_rf\_genetico$`SARS-Cov-2` in 75 controls (test$`MNT 1:80` No SARS-Cov-2) < 10 cases (test$`MNT 1:80` SARS-Cov-2).  
## Area under the curve: 0.6373

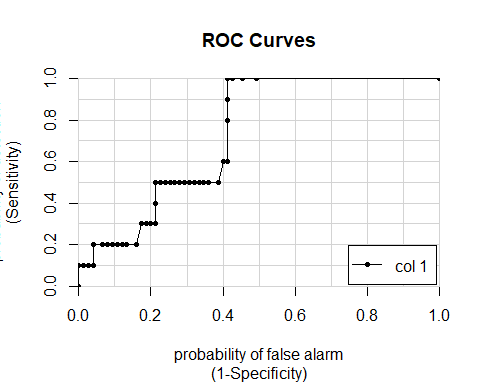
#Построим ROC-кривую по каждой из моделей

caTools::colAUC(pred\_prob\_rf\_genetico$`SARS-Cov-2`, test$`MNT 1:80`, plotROC = T)



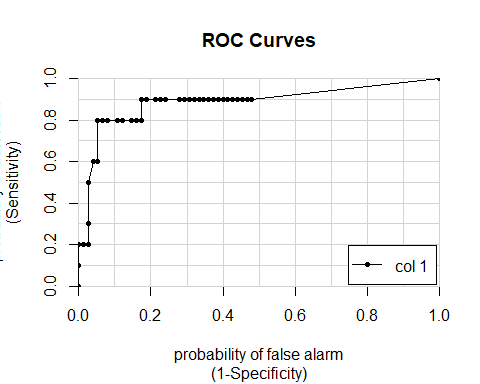
## [,1]  
## No SARS-Cov-2 vs. SARS-Cov-2 0.6373333

caTools::colAUC(pred\_prob\_rf\_abbott$`SARS-Cov-2`, test$`MNT 1:80`, plotROC = T)



## [,1]  
## No SARS-Cov-2 vs. SARS-Cov-2 0.732

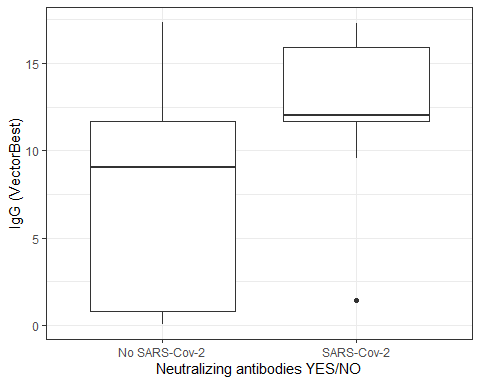
caTools::colAUC(pred\_prob\_rf\_vector$`SARS-Cov-2`, test$`MNT 1:80`, plotROC = T)



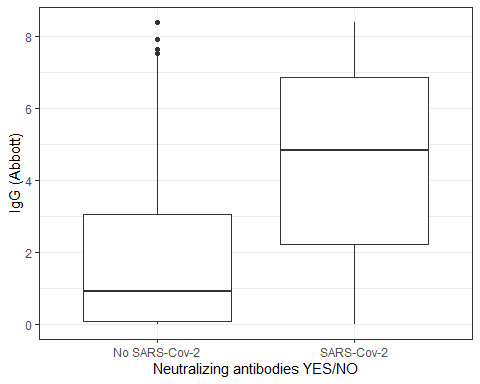
## [,1]  
## No SARS-Cov-2 vs. SARS-Cov-2 0.8866667

#Сравним получившиеся результаты с визуализацией с помощью боксплотов

ggplot(test\_nab, aes(y = vector, x = `MNT 1:80` )) +  
 geom\_boxplot() +   
 ylab("IgG (VectorBest)") +   
 xlab("Neutralizing antibodies YES/NO") +   
 theme\_bw()



ggplot(test\_nab, aes(y = abbott, x = `MNT 1:80` )) +  
 geom\_boxplot() +   
 ylab("IgG (Abbott)") +   
 xlab("Neutralizing antibodies YES/NO") +   
 theme\_bw()



ggplot(test\_nab, aes(y = genetico, x = `MNT 1:80` )) +  
 geom\_boxplot() +   
 ylab("IgG (Genetico)") +   
 xlab("Neutralizing antibodies YES/NO") +   
 theme\_bw()

