report

ZJH

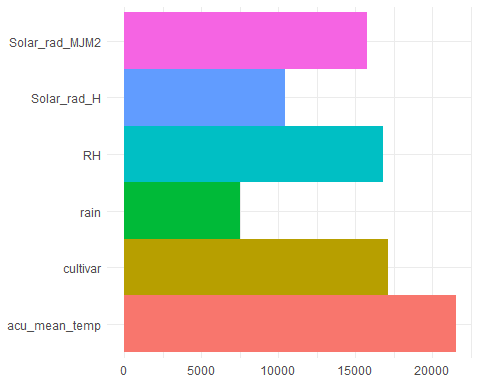
2022/3/29

analyze\_data <- readRDS('analyze.RDS')   
numeric\_dat <- analyze\_data %>% select(.,-c(key,observe\_ys,observe\_date:sample\_label,G1L:O9L,year,temp\_differ))  
train\_idx <- sample(1:138,100)  
test\_idx <- !(1:138 %in%train\_idx)  
train <- numeric\_dat[train\_idx,]  
test <- numeric\_dat[test\_idx,]  
  
 importanceplot <- function(pp=and\_rf) {  
 i\_scores <- varImp(pp, conditional=TRUE)  
 #Gathering rownames in 'var' and converting it to the factor  
 #to provide 'fill' parameter for the bar chart.   
 i\_scores <- i\_scores %>% tibble::rownames\_to\_column("var")   
 i\_scores$var<- i\_scores$var %>% as.factor()  
   
 #Plotting the bar and polar charts for comparing variables  
 i\_bar <- ggplot(data = i\_scores) +   
 geom\_bar(  
 stat = "identity",#it leaves the data without count and bin  
 mapping = aes(x = var, y=Overall, fill = var),   
 show.legend = FALSE,  
 width = 1  
 ) +   
 labs(x = NULL, y = NULL)  
 i\_bar + coord\_flip() + theme\_minimal()  
 }

xtrain <- model.matrix(polyphenol~.+acu\_mean\_temp\*RH\*Solar\_rad\_H\*Solar\_rad\_MJM2, train)[,-1]  
ytrain <- train$polyphenol  
ytest <- test$polyphenol  
xtest <- model.matrix(polyphenol~.+acu\_mean\_temp\*RH\*Solar\_rad\_H\*Solar\_rad\_MJM2, test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_polyphenol <- analyze\_data %>% select(., c(cultivar,rain,polyphenol,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(polyphenol~.+acu\_mean\_temp\*RH\*Solar\_rad\_H\*Solar\_rad\_MJM2, data = data\_polyphenol[train\_idx,])  
and\_pred <- predict(and\_rf,data\_polyphenol[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.356668

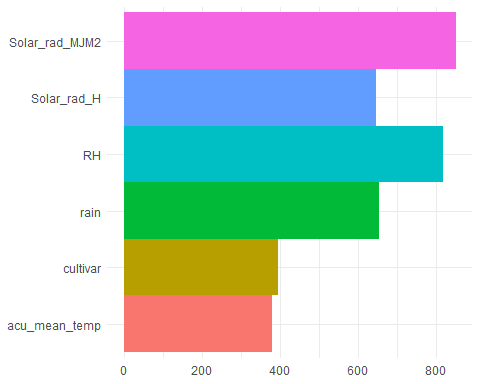
importanceplot()



xtrain <- model.matrix(FAA~., train)[,-1]  
ytrain <- train$FAA  
ytest <- test$FAA  
xtest <- model.matrix(FAA~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_FAA <- analyze\_data %>% select(., c(cultivar,FAA,rain,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(FAA~., data = data\_FAA[train\_idx,])  
and\_pred <- predict(and\_rf,data\_FAA[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.4748558

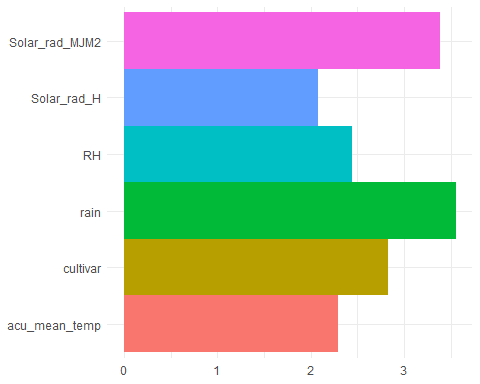
importanceplot()



xtrain <- model.matrix(Gallic\_Acid~., train)[,-1]  
ytrain <- train$Gallic\_Acid  
ytest <- test$Gallic\_Acid  
xtest <- model.matrix(Gallic\_Acid~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_Gallic\_Acid <- analyze\_data %>% select(., c(cultivar,rain,Gallic\_Acid,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(Gallic\_Acid~., data = data\_Gallic\_Acid[train\_idx,])  
and\_pred <- predict(and\_rf,data\_Gallic\_Acid[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.4251999

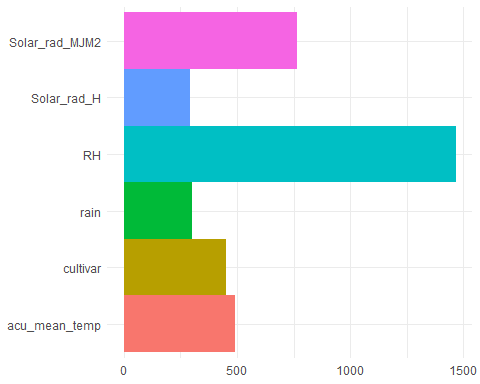
importanceplot()



xtrain <- model.matrix(GC~., train)[,-1]  
ytrain <- train$GC  
ytest <- test$GC  
xtest <- model.matrix(GC~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_GC <- analyze\_data %>% select(., c(cultivar,rain,GC,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(GC~., data = data\_GC[train\_idx,])  
and\_pred <- predict(and\_rf,data\_GC[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.3226176

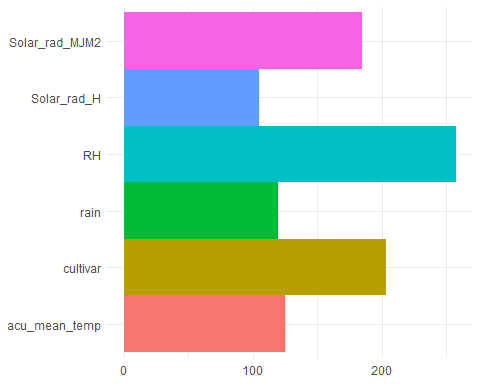
importanceplot()



xtrain <- model.matrix(GCG~., train)[,-1]  
ytrain <- train$GCG  
ytest <- test$GCG  
xtest <- model.matrix(GCG~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_GCG <- analyze\_data %>% select(., c(cultivar,GCG,rain,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(GCG~., data = data\_GCG[train\_idx,])  
and\_pred <- predict(and\_rf,data\_GCG[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.2057175

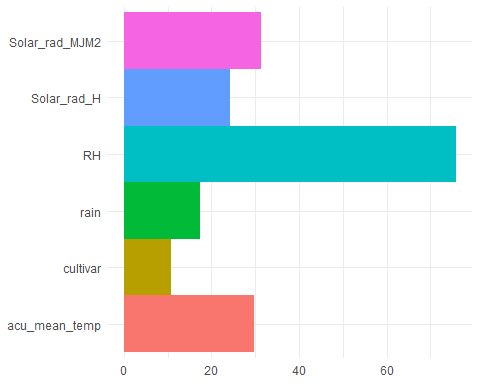
importanceplot()



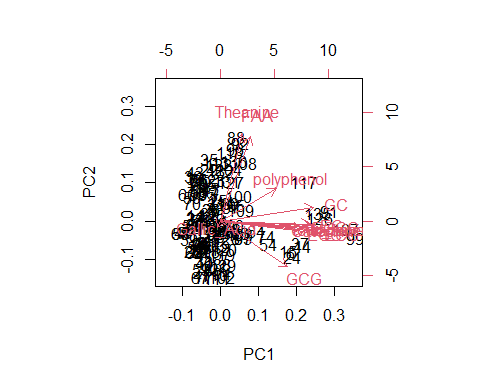
xtrain <- model.matrix(Catechin~., train)[,-1]  
ytrain <- train$Catechin  
ytest <- test$Catechin  
xtest <- model.matrix(Catechin~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_caffeine <- analyze\_data %>% select(., c(cultivar,Catechin,rain,acu\_mean\_temp,RH,Solar\_rad\_H,Solar\_rad\_MJM2))  
and\_rf <- randomForest(Catechin~., data = data\_caffeine[train\_idx,])  
and\_pred <- predict(and\_rf,data\_caffeine[test\_idx,])  
(mean((and\_pred-ytest)^2)/var(ytest))^0.5

## [1] 0.5386441

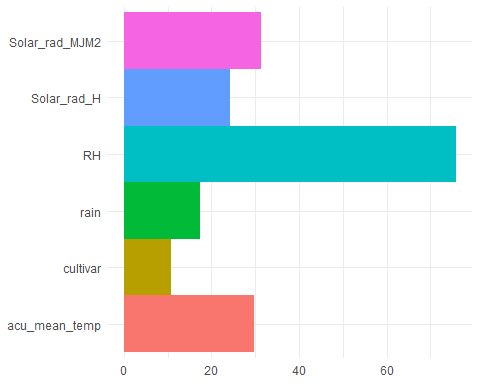
importanceplot()



chem <- analyze\_data %>% select(,polyphenol:total\_catechins)  
pr\_chem <- prcomp(scale(chem) )  
biplot(pr\_chem)



importanceplot()



xtrain <- model.matrix(level~., train)[,-1]  
ytrain <- train$level  
ytest <- test$level  
xtest <- model.matrix(level~., test)[,-1] %>% as\_tibble() %>% select(.,season2nd:season6th,season7th,everything()) %>% as.matrix()  
data\_level <- analyze\_data %>% select(., c(cultivar,level,polyphenol:total\_catechins))  
data\_level$level <- factor(data\_level$level)  
and\_rf <- randomForest(level~., data = data\_level[train\_idx,])  
and\_pred <- predict(and\_rf,data\_level[test\_idx,])  
table(data\_level[test\_idx,]$level,and\_pred)

## and\_pred  
## B C  
## B 8 3  
## C 4 23

importanceplot()

