

```

## R code for data splitting for PLSR calibration and validation
## Stratified sampling considering a range of trait
## 'Spectral wavelength range influences the performance of chemometric models
estimating various foliar functional traits'
## Authors: Minjee Park, Lorenzo Cotrozzi, Geoffrey M Williams, Matthew D Ginzel,
Michael V Mickelbart, Douglass F Jacobs, John J Couture

setwd("D:/Purdue_Research")
library(plyr)

inVar="Sugars" # Foliar trait, e.g., Sugars

Volume=read.csv(paste("2018+2019_ChemicalData_",inVar,".csv", sep=""), sep=',',
header=T)

id <- c(1:nrow(Volume))
Volume <- cbind(id=id, Volume)

grouped = ddply(Volume, .(id),transform, quant_num =
                ifelse(get(inVar) <= quantile(Volume[,inVar], .25), 1,
                ifelse(get(inVar) <= quantile(Volume[,inVar], .5), 2,
                ifelse(get(inVar) <= quantile(Volume[,inVar], .75),
3, 4)))
)

ExtVal <- ddply(grouped, .(quant_num), function(x)
x[sample(nrow(x),round(nrow(Volume)*0.2/4)),]) # Randomly take ~20 % of the original
data
CalVal<-Volume[-ExtVal[,1],] # Randomly take ~80 % of the original data

ExtVal <- ExtVal[,-c(1,ncol(ExtVal))]
CalVal <- CalVal[,-1]

write.csv(CalVal, paste("2018+2019_ChemicalData_",inVar,"_CalVal.csv", sep=""),
row.names=FALSE)
write.csv(ExtVal, paste("2018+2019_ChemicalData_",inVar,"_ExtVal.csv", sep=""),
row.names=FALSE)

```