readData = function(path = "G:\\data\\D\\data project D.xlsx"){

#check if it is csv of xlsx

if(grepl("xlsx", path)){

d <- openxlsx::read.xlsx(path, sheet = 1,colNames = FALSE)

}else if(grepl("csv", path)){

# file = "C:\\Users\\Sili Fan\\Downloads\\val (18).csv"

d <- data.table::fread(path)

}

# make "" as NA

d[d==""] <- NA

#### fData

fData <- d[!is.na(d[,1]),c(which(is.na(d[1,])),sum(is.na(d[1,]))+1)] # The first row and column is critical of formating the data.

colnames(fData) = as.character(fData[1,]); fData = data.frame(fData[-1,],stringsAsFactors = F,check.names = FALSE);rownames(fData) = 1:nrow(fData);

# following steps keeps the column type.

fData.=lapply(fData,function(x){

if(sum(!is.na(as.numeric(x))) == length(x)){

as.numeric(x)

}else{

x

}

})

fData. = do.call(cbind, lapply(fData., data.frame, stringsAsFactors=FALSE))

colnames(fData.) = colnames(fData)

fData = fData.

fData = fData[,c(ncol(fData),2:ncol(fData)-1)]

fData[[1]] = make.unique(fData[[1]], sep = '\_')

#### pData

pData <- d[c(which(is.na(d[,1])),max(which(is.na(d[,1])))+1) ,!is.na(d[1,])]

pData <- t(pData); colnames(pData) = pData[1,]; pData = data.frame(pData[-1,],stringsAsFactors = F,check.names = FALSE)

# following steps keeps the column type.

pData.=lapply(pData,function(x){

if(sum(!is.na(as.numeric(x))) == length(x)){

as.numeric(x)

}else{

x

}

})

pData. = do.call(cbind, lapply(pData., data.frame, stringsAsFactors=FALSE))

colnames(pData.) = colnames(pData)

pData = pData.

pData = pData[,c(ncol(pData),2:ncol(pData)-1)]

pData[[1]] = make.unique(make.names(pData[[1]]), sep = '\_')

#### eData

eData <- d[!is.na(d[,1]),!is.na(d[1,])][-1,-1]

eData <- sapply(eData, as.numeric)

eData <- data.frame(eData,stringsAsFactors = F)

colnames(eData) = pData[[1]]; rownames(eData) = fData[[1]]

# # remove any unwanted character in columns of eData, fData and pData to \_.

# colnames(eData) = gsub("([\_])|[[:punct:]]", "\_", colnames(eData))

# colnames(fData) = gsub("([\_])|[[:punct:]]", "\_", colnames(fData))

# colnames(pData) = gsub("([\_])|[[:punct:]]", "\_", colnames(pData))

# remove all the NA. And replace NA with "NA" Otherwise DataTables will give error.datatables warning requested unknown parameter

# eData[is.na(eData)]="NA"

# fData[is.na(fData)]="NA"

# pData[is.na(pData)]="NA"

# remove unwanted character in p.

# for(i in 1:nrow(pData)){

# for(j in 1:ncol(pData)){

# pData[i,j] = gsub("\\+|~|-", " ", pData[i,j])

# }

# }

return(list(e = eData, f = fData, p = pData))

}