mouse_drug_v2_20201105

Julse Jiang

2020/11/5

读取表格，转置

```
library(ggplot2)
rm(list = ls())
dat = read.csv('1.csv',stringsAsFactors = F)
dat=dat[2:138]
tdat = t(dat)
tdat = as.data.frame(tdat)
head(dat)[1:5]
```

```
##     control.1   control.2   control.3   control.4   control.5
## 1 0.195569448 0.073117695 0.120875238 0.173960315 0.113318837
## 2 0.010709432 0.017912476 0.084914379 0.011524871 0.002690949
## 3 0.000271813 0.000217450 0.000271813 0.000298994 0.000081500
## 4 0.036531666 0.072628432 0.017640663 0.080157652 0.025767872
## 5 0.009459092 0.008833922 0.018293014 0.027453112 0.003696657
## 6 0.019951074 0.087985866 0.032699103 0.008344659 0.058983419
```

```
head(tdat)[1:5]
```

```
##                    V1          V2          V3         V4          V5
## control.1 0.19556945 0.010709432 0.000271813 0.03653167 0.009459092
## control.2 0.07311769 0.017912476 0.000217450 0.07262843 0.008833922
## control.3 0.12087524 0.084914379 0.000271813 0.01764066 0.018293014
## control.4 0.17396032 0.011524871 0.000298994 0.08015765 0.027453112
## control.5 0.11331884 0.002690949 0.000081500 0.02576787 0.003696657
## control.6 0.34028269 0.009594999 0.000679532 0.07616200 0.008181571
```

分类统计

一共有12个组，每个组的样本数目如下

```
# 获取组名***********************************
tmp = rownames(tdat)
tmp = strsplit(unlist(tmp),'.',fixed = TRUE)
# 丢弃最后一个元素,合并
#paste(head(tmp[[134]],-1),collapse = '.')

myfun = function(x){
  x = head(x,-1)
}
group_type =sapply(lapply(tmp, myfun), paste, collapse = ".")
group_count = as.data.frame(table(group_type))
group_count
```

```
##    group_type Freq
## 1     control   20
## 2       D.G17    9
## 3       D.G75    9
## 4       D.Rb1   10
## 5       D.Rg2   10
## 6       D.Rg3   10
## 7         DSS   19
## 8         G17   10
## 9         G75   10
## 10        Rb1   10
## 11        Rg2   10
## 12        Rg3   10
```

分组统计

dist:计算欧氏距离 分组统计异常小鼠："DSS.4" "G17.5" "Rg2.3" "Rg3.1" "Rg3.4"

```r
#分组统计------------------------------------------------
mylist = c()
par(mfrow=c(2,6))
for (y in 1:12){
  # 1:20
  #取某一类别的名字
  tmp_group = group_count[y,1]
  #20
  tmp_num = as.integer(group_count[which(group_count$group_type %in% tmp_group),][2])
  # control.1~20
  tmp_list = paste(tmp_group,1:tmp_num,sep = '.')
  control_tdat = tdat[tmp_list,]
  t = colMeans(control_tdat)
  tmp = data.frame()
  for (x in 1:tmp_num){
    tmp = rbind(tmp,dist(rbind(t,colMeans(control_tdat[-x,1:2485]))))
  }

  colnames(tmp) = c(tmp_group)
  sp = boxplot(tmp,xlab=tmp_group)
  sp$out
    xi <- 0.1 + seq(sp$n)
  for (a in which(tmp==as.double(sp$out))){
    # text(xi,tmp[a,], a)
    mylist = append(mylist,paste(tmp_group,a,sep = '.'))
    # text(xi,tmp[a,], paste(a,tmp[a,],':'))
    # print(paste(tmp_group,a,sep = '.'))
  }
}
```
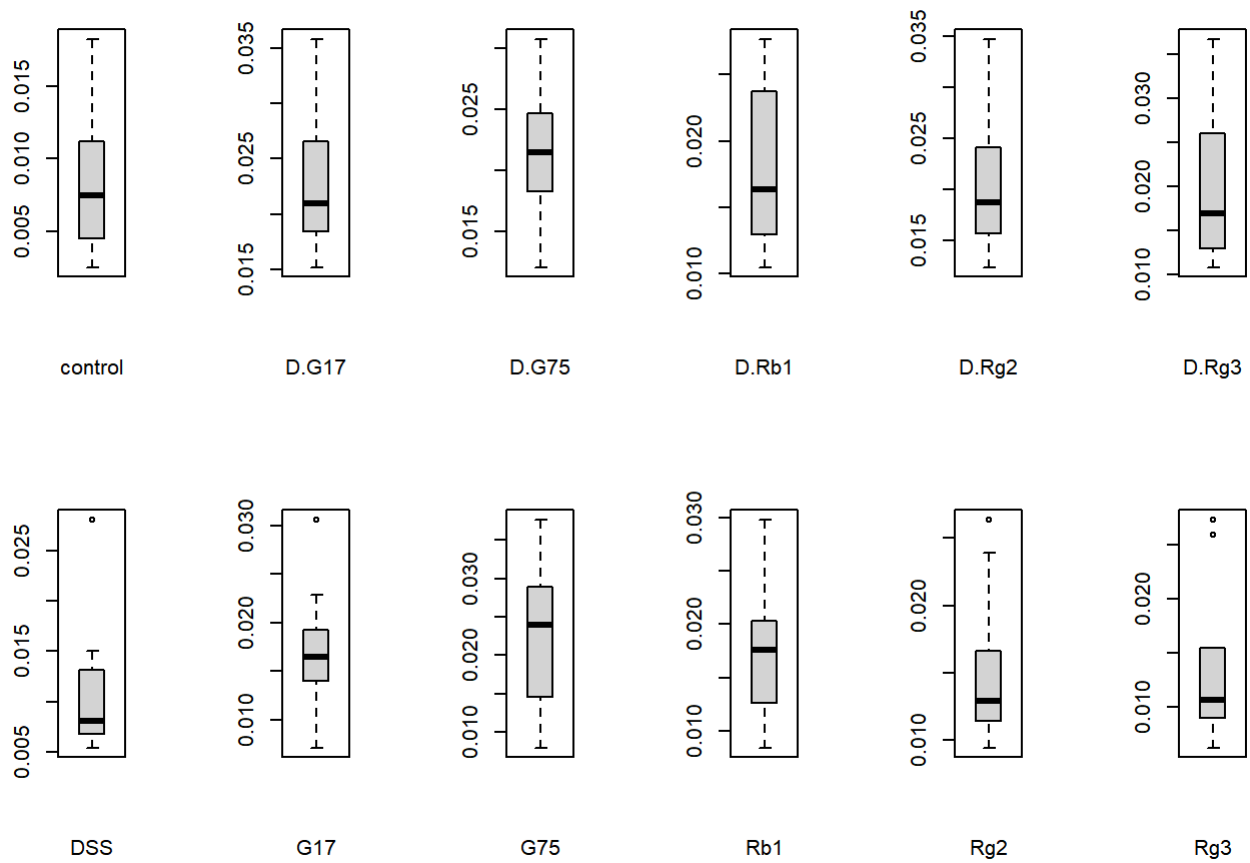
```
mylist
```

```
## [1] "DSS.4" "G17.5" "Rg2.3" "Rg3.1" "Rg3.4"
```