Homework due May 9, 2021 07:01 +03

Heat Maps Exercises #1

1/1 point (graded)

Load the data:

```
library(GSE5859Subset)
data(GSE5859Subset)
```

Pick the 25 genes with the highest across sample variance. This function might help

```
install.packages("matrixStats")
library(matrixStats)
?rowMads ##we use mads due to a outlier sample
```

While a heatmap function is included in R, we recommend the `heatmap.2` function from the **gplots** package on CRAN because it is a bit more customized. For example, it stretches to fill the window.

```
library(gplots)
```

Use `heatmap.2()` to make a heatmap showing the `sampleInfo$group` with color, the date as labels, the rows labelled with chromosome, and scaling the rows.

What do we learn from this heatmap?

- ○ The data appears as if it was generated by `rnorm()`
- ○ Some genes in chr1 are very variable
- ◉ A group of chrY genes are higher in group 0 and appear to drive the clustering. Within those clusters there appears to be clustering by month.

- ◯ A group of chrY genes are higher in October compared to June and appear to drive the clustering. Within those clusters there appears to be clustering by `sampleInfo$group`.

✔

Explanation

```
##load libraries
library(rafalib)
library(gplots)
library(matrixStats)
library(RColorBrewer)
##make colors
cols = colorRampPalette(rev(brewer.pal(11,"RdBu")))(25)
gcol=brewer.pal(3,"Dark2")
gcol=gcol[sampleInfo$g+1]

##make lables: remove 2005 since it's common to all
labcol= gsub("2005-","",sampleInfo$date)

##pick highly variable genes:
sds =rowMads(geneExpression)
ind = order(sds,decreasing=TRUE)[1:25]

## make heatmap
heatmap.2(geneExpression[ind,],
          col=cols,
          trace="none",
          scale="row",
          labRow=geneAnnotation$CHR[ind],
          labCol=labcol,
          ColSideColors=gcol,
          key=FALSE)
```

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Heat Maps Exercises #2

1/1 point (graded)

Create a large data set of random data that is completely independent of `sampleInfo$group` like this:

```
set.seed(17)
m = nrow(geneExpression)
n = ncol(geneExpression)
x = matrix(rnorm(m*n),m,n)
g = factor(sampleInfo$g )
```

Create two heatmaps with these data. Show the group g either with labels or colors.

1. Taking the 50 genes with smallest p-values obtained with `rowttests`

2. Taking the 50 genes with largest standard deviations.

Which of the following statements is true:

- ( ) These two techniques produced similar heatmaps.
- ( ) Selecting genes with the t-test is a better technique as it permits us to detect the two groups. It appears to find hidden signals.
- (•) There is no relationship between `g` and `x` but with 8,793 tests some will appear significant by chance. Selecting genes with the t-test gives us a deceiving result.
- ( ) The genes with the largest standard deviation add variability to the plot and do not let us find the differences between the two groups.

✔

Explanation

```
library(gplots)
library(matrixStats)
library(genefilter)
library(RColorBrewer)
cols = colorRampPalette(rev(brewer.pal(11,"RdBu")))(25)

ttest = rowttests(x,g)
sds = rowSds(x)
Indexes = list(t=order(ttest$p.value)[1:50], s=order(-sds)[
for(ind in Indexes){
 heatmap.2(x[ind,],
           col=cols,
           trace="none",
           scale="row",
           labCol=g,
           key=FALSE)
}
```

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem