Homework due May 9, 2021 07:01 +03

K-means Exercises #1

1/1 point (graded)

Run `kmeans()` with 5 centers for the blood RNA data:

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

Set the seed to 10, `set.seed(10)`, right before running `kmeans()` with 5 centers.

Explore the relationship of clusters and information in `sampleInfo`. Which of the following best describes what you find:

- ( ) `sampleInfo$group` is driving the clusters as the 0s and 1s are in completely different clusters
- ( ) The year is driving the clusters
- (•) Date is driving the clusters
- ( ) The clusters don't depend on any of the columns of `sampleInfo`

✔

Explanation

```
mds=cmdscale(dist(t(geneExpression)))
set.seed(10)
result=kmeans(t(geneExpression),5)
mypar(1,1)
plot(mds,bg=result$cl,pch=21)
table(sampleInfo$group,result$cluster)
table(sampleInfo$date,result$cluster)
##looks better if we re-order:
table(sampleInfo$date,result$cluster)[,c(4,1,5,3,2)]
```

Submit | You have used 1 of 2 attempts

Answers are displayed within the problem