improved leverage

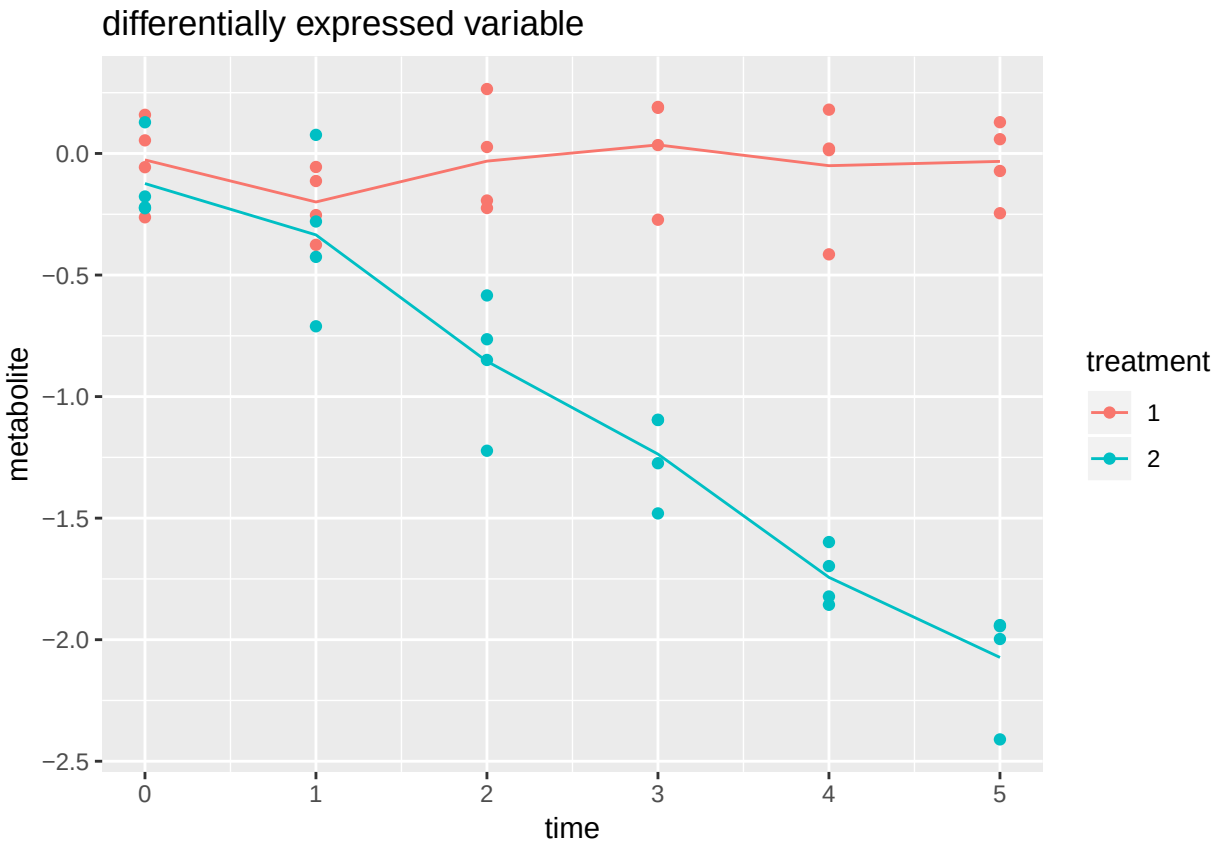
zhu

23/12/2018

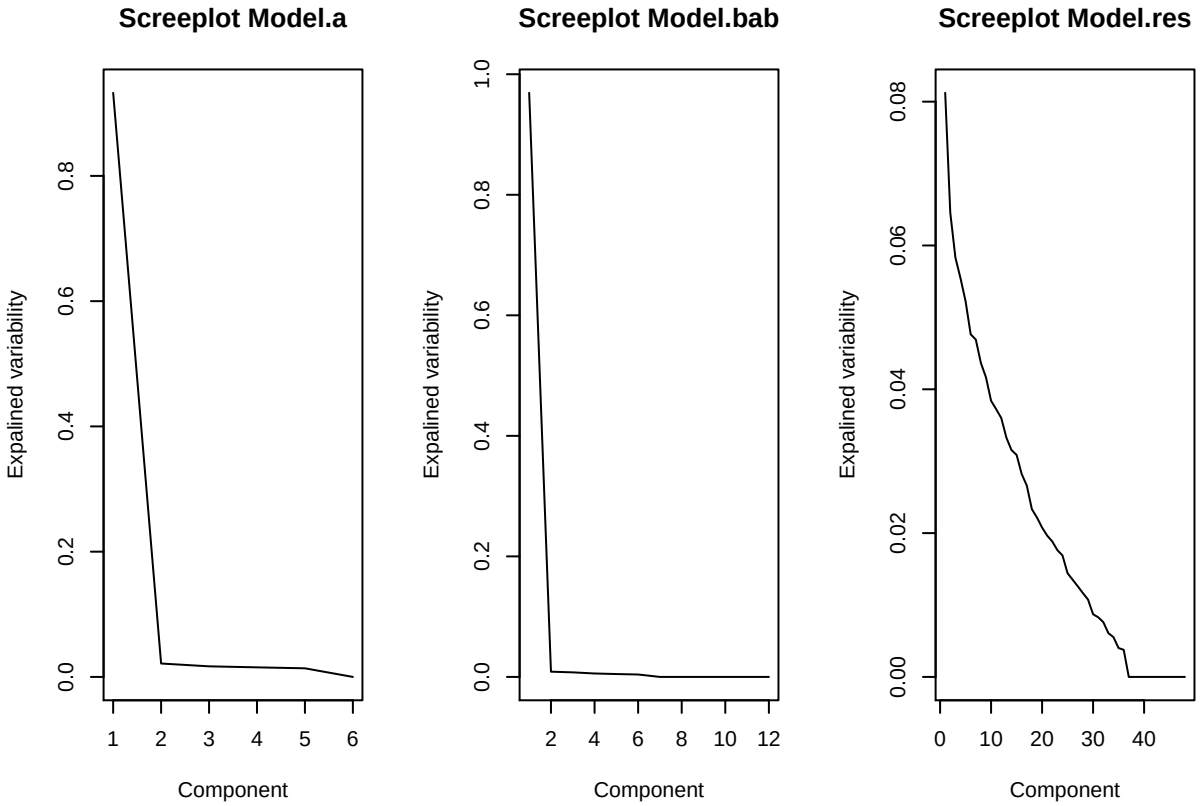
Data

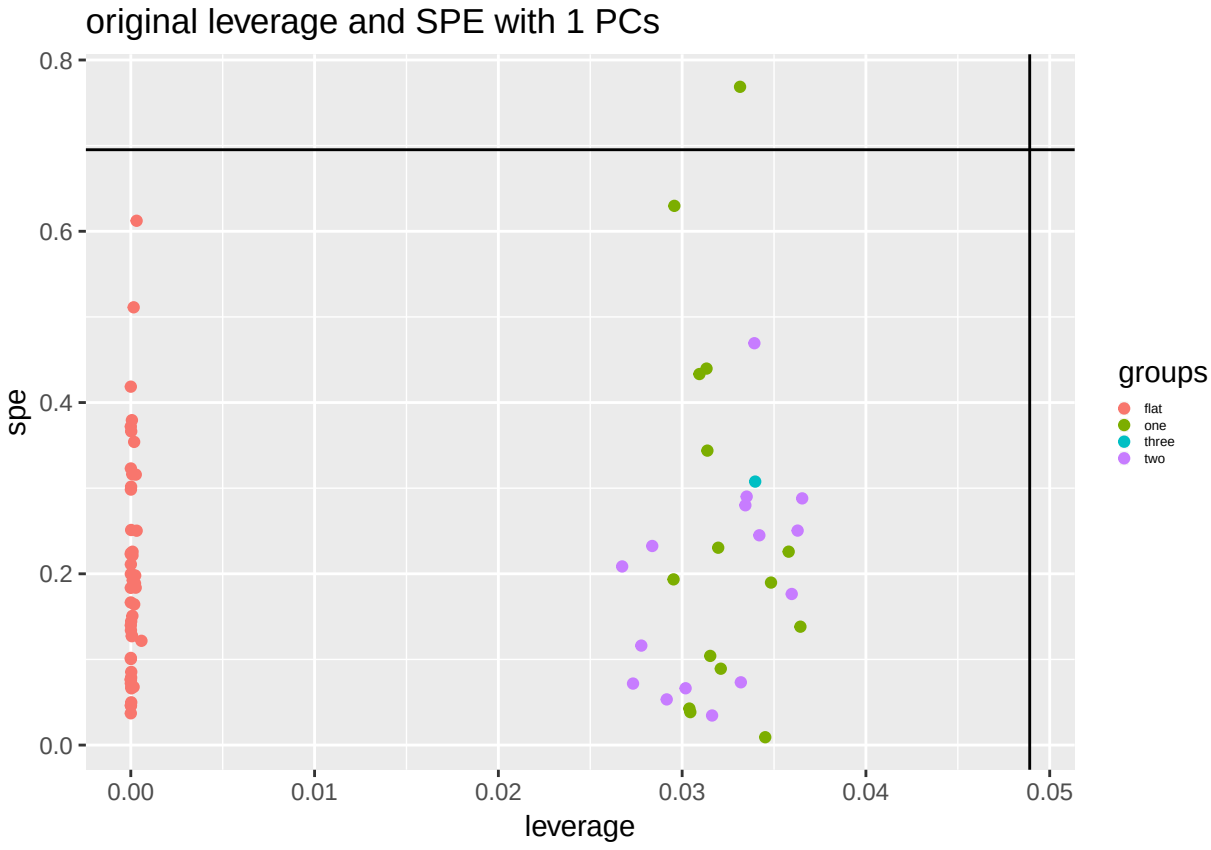
there are 2 treatments, 6 time points and 4 replicates. replicates were drawn as independent observations from normal distribution N(0,0.2). The simulation dataset consists of 48 rows(samples) and 81 columns(31 differential expressed variable and 50 flat variables).

example of differentially expressed variable is shown below.



Original leverage





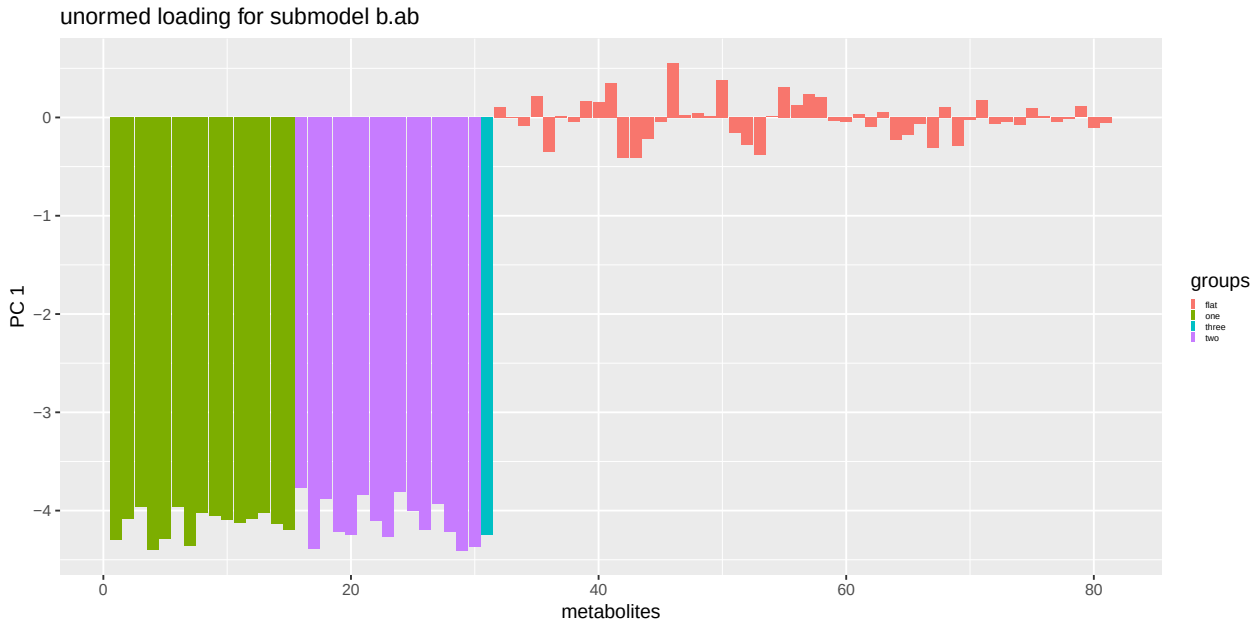
```
## character(0)
```

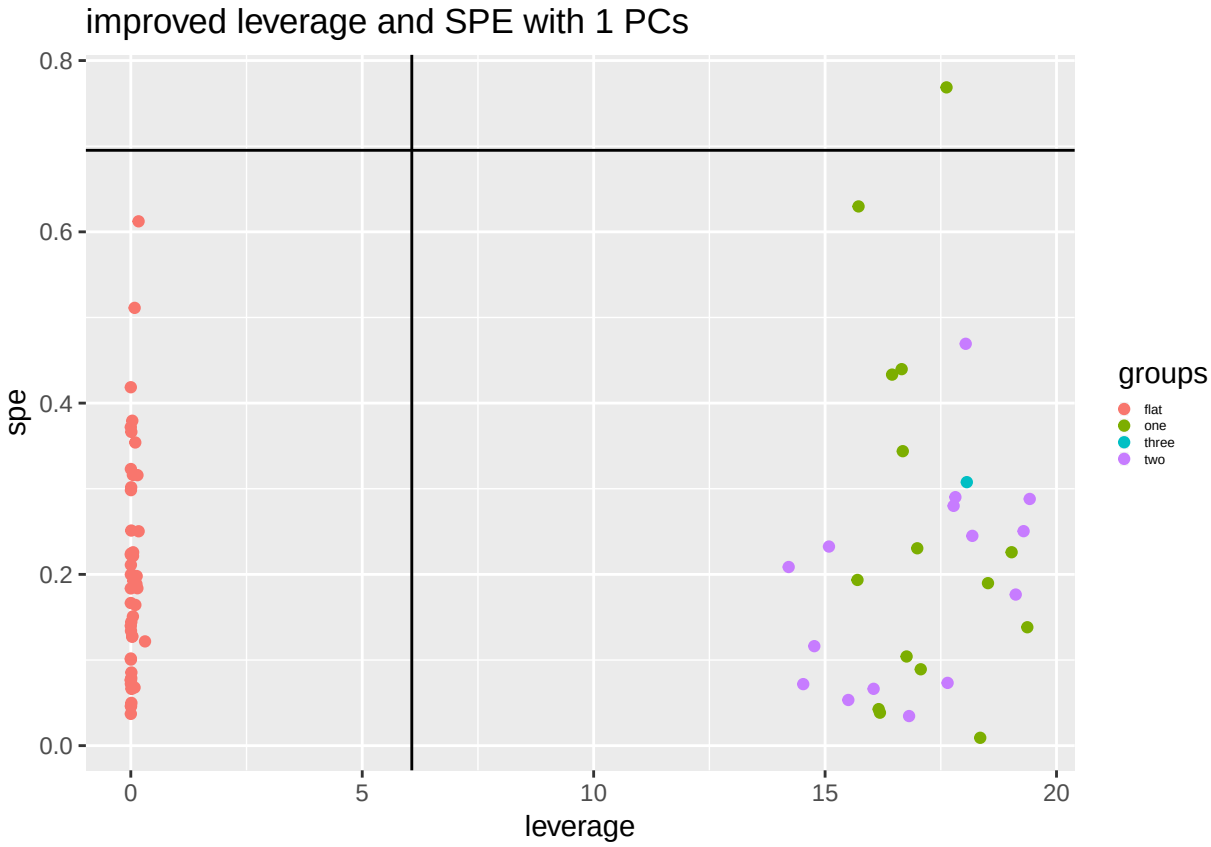
As expected, with 1 PC selected for submodel treatment and submodel interaction, leverage cutoff line set excluded all of the variabes to be significant.

Improved leverage

As discussed previously, this is due to the loadings were normalized to unit vector during calculation. Alternatively, we could normalize score to unit vector. Modification was done in the asca-gene algotirhm to achieve this. The formulus as well as the code is shown below:

```
PCA.GENES.unorm_loading<-function(X)
{
  #PCA.GENES is very useful to obtain principal components to a matrix that has more variables than ind
  #R can not apply princomp is such case and when there are a lot of variables eigen(t(X)%*%X) can not

  #X is a matrix that has on columns the genes considered as variables in the PCA analysis.
  #First we center the matrix by columns (Xoff) and then we obtain the eigenvalues and the eigenvectors


  n<-ncol(X)
  p<-nrow(X)
  offset<-apply(X,2,mean)
  Xoff<-X-(cbind(matrix(1,p,1))%*%rbind(offset))

  eigen<-eigen(Xoff%*%t(Xoff)/(p-1))#calculates eigen value and unit eigen vector of c*cT
  var<-cbind(eigen$values/sum(eigen$values),cumsum(eigen$values/sum(eigen$values)))

  scores.normed<-eigen$vectors
  loading.unormed<-t(Xoff)%*%scores.normed



  output<-list(eigen,var,scores.normed,loading.unormed)
  names(output)<-c("eigen","var.exp","scores","loadings.unormed")
  output
}
```





the new loading and leverage and spe plot are shown above. simulation data are correctly indentified as designed.