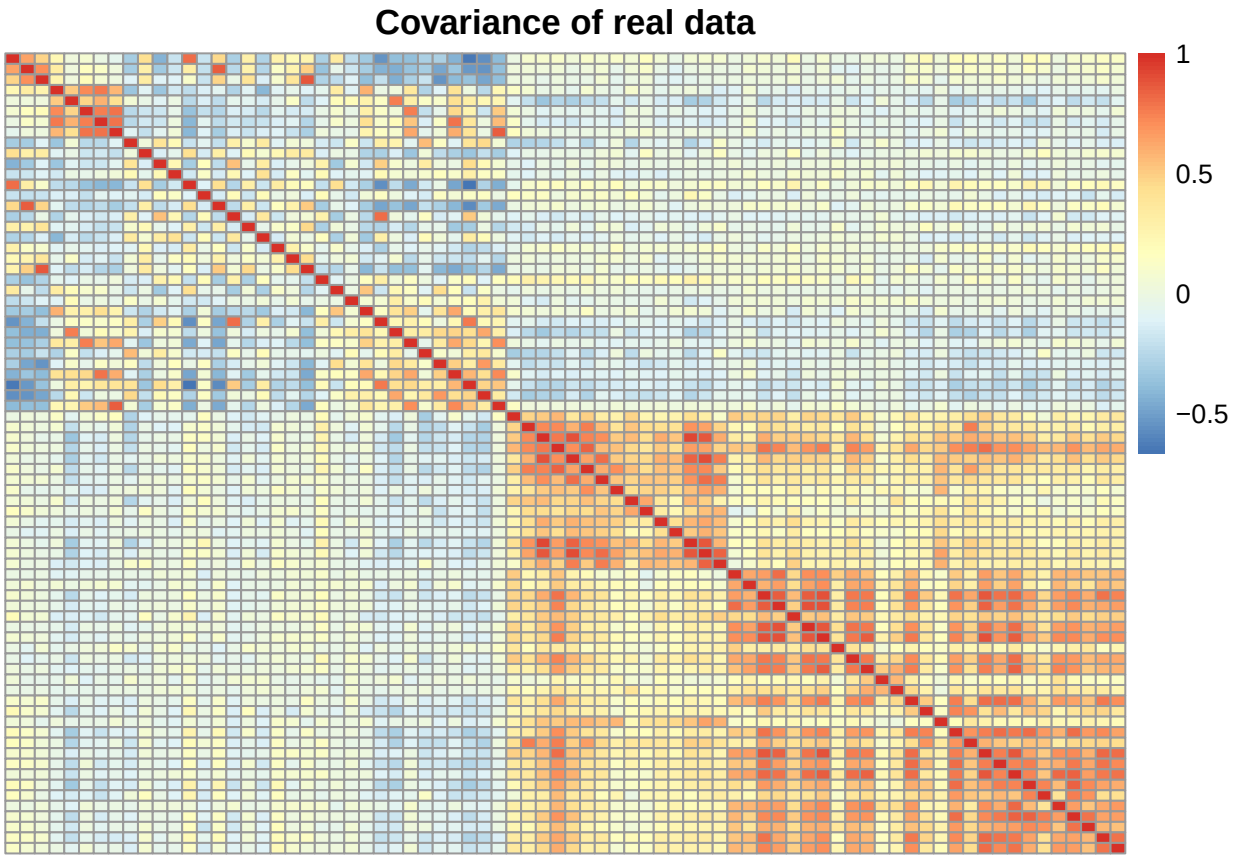
Simulation data generation

Rum Wei

12/18/2017

1.Import real data and calculate covariance

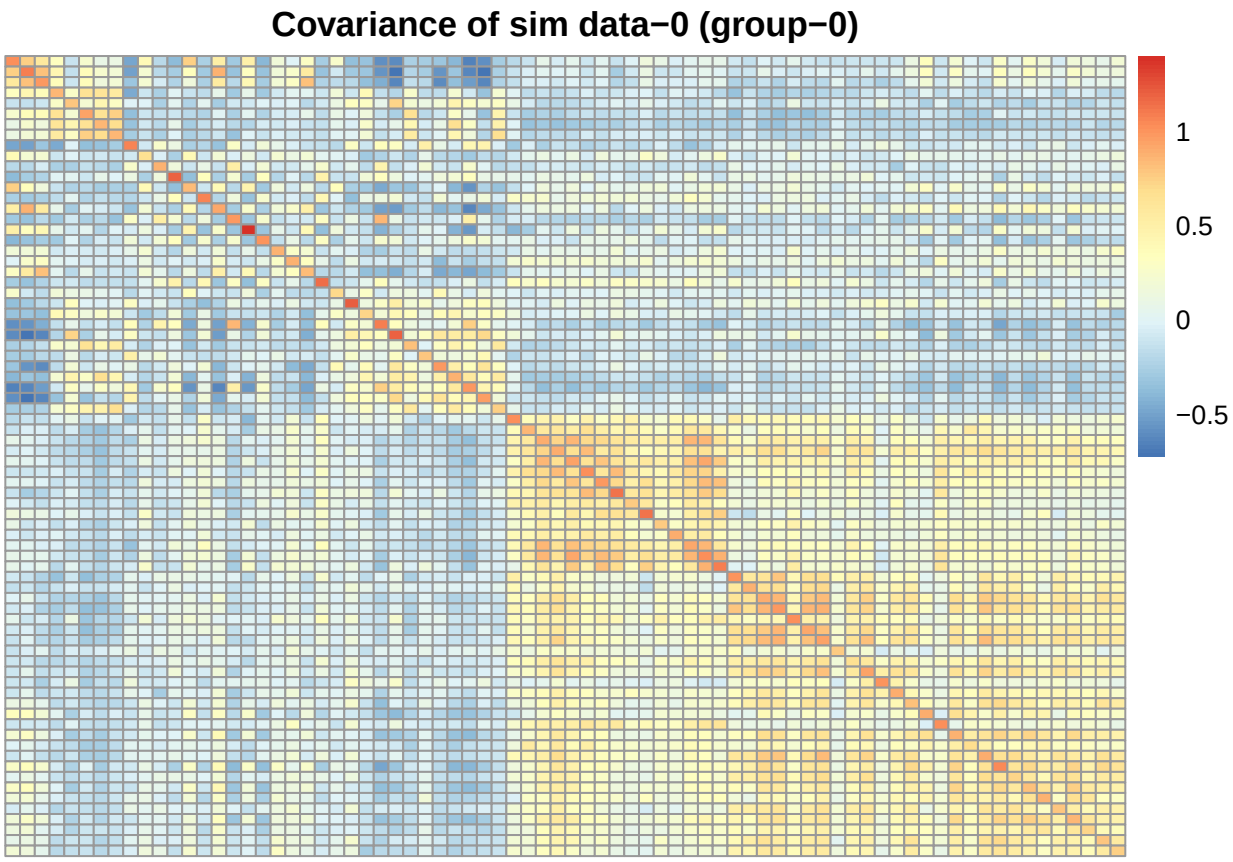
```
options(stringsAsFactors = F)
require(magrittr)
require(pheatmap)
require(mvtnorm)

data_raw <- read.csv('real_data.csv', row.names=1)
data_lg_sc <- data_raw %>% log %>% scale
cov_mat <- cov(data_lg_sc)
pheatmap(cov_mat, cluster_rows=F, cluster_cols=F, show_rownames=F, show_colnames=F,
         main='Covariance of real data')
```



2.Simulation dataset generation (part-1, first 80 samples as first group)

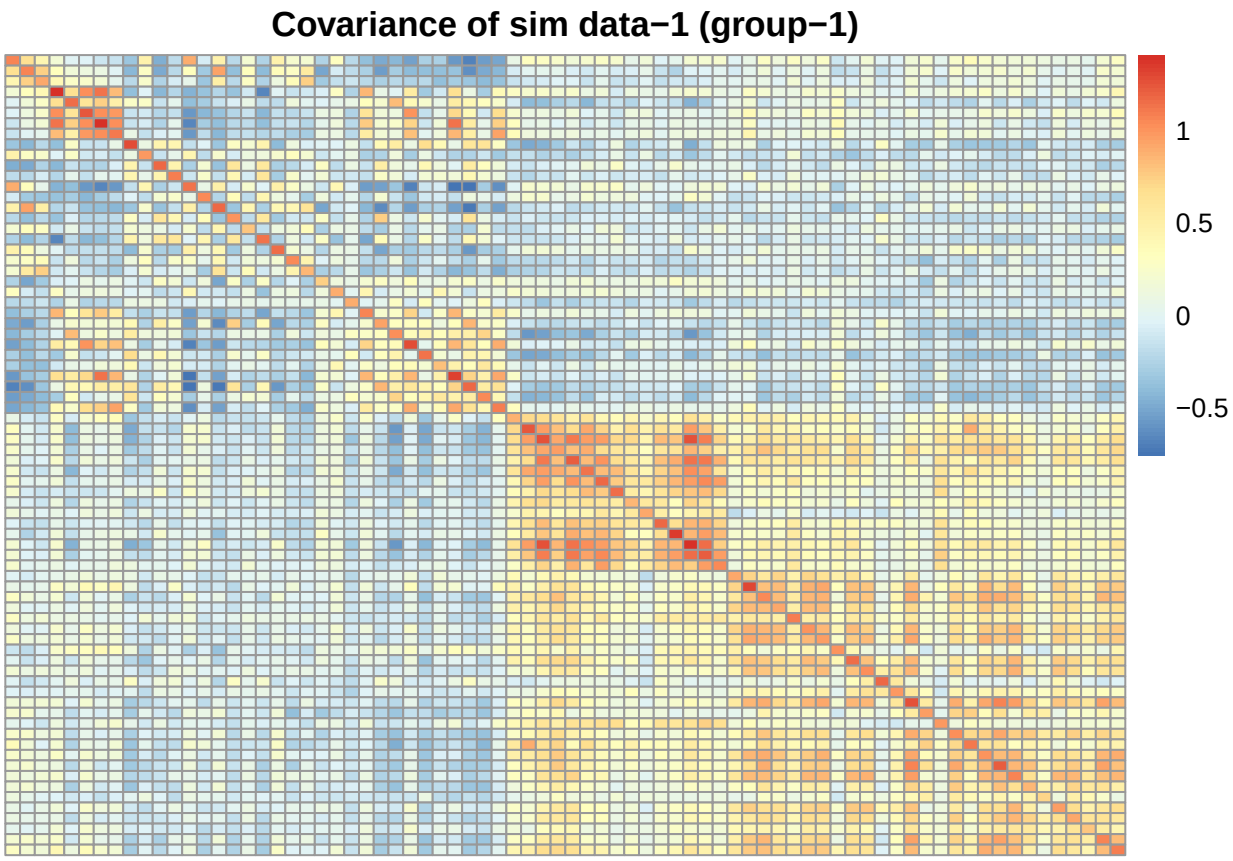
```
set.seed(123)
data_sim_0 <- rmvnorm(80, mean=rnorm(nrow(cov_mat), sd=.5), sigma=cov_mat, method='svd')
cov_mat_sim_0 <- cov(data_sim_0)
```

```
pheatmap(cov_mat_sim_0, cluster_rows=F, cluster_cols=F, show_rownames=F, show_colnames=F,
         main='Covariance of sim data-0 (group-0)')
```



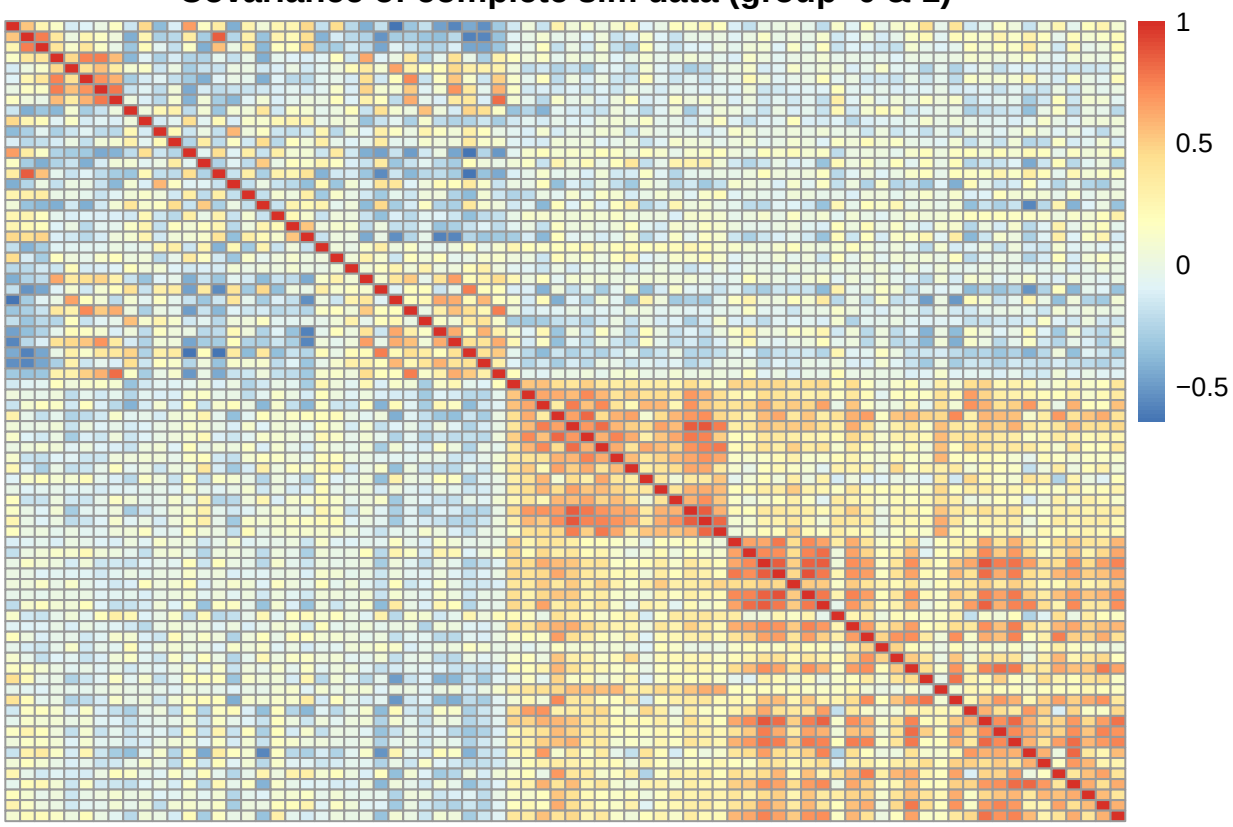
3.Simulation dataset generation (part-2, second 80 samples as second group)

```
set.seed(321)
data_sim_1 <- rmvnorm(80, mean=rnorm(nrow(cov_mat), sd=.5), sigma=cov_mat, method='svd')
cov_mat_sim_1 <- cov(data_sim_1)
pheatmap(cov_mat_sim_1, cluster_rows=F, cluster_cols=F, show_rownames=F, show_colnames=F,
         main='Covariance of sim data-1 (group-1)')
```



4.Simulation dataset generation (part-3, stack two groups together)

```
data_sim <- rbind(data_sim_0, data_sim_1)
data_sim_sc <- scale(data_sim)
cov_mat_sim <- cov(data_sim_sc)
pheatmap(cov_mat_sim, cluster_rows=F, cluster_cols=F, show_rownames=F, show_colnames=F,
         main='Covariance of complete sim data (group-0 & 1)')
```



5.T-test on two groups

```
group <- rep(c(0, 1), each=80) %>% as.factor
sim_pvals <- apply(data_sim_sc, 2, function(x) t.test(x ~ group)$p.value)
sim_pvals[1:10]
```

```
##  [1] 7.746669e-11 1.963778e-01 8.522197e-06 9.885972e-02 1.165720e-01
##  [6] 5.295838e-06 5.510601e-01 3.338875e-06 1.934161e-04 1.385230e-01
```