

Preparation

Cnv.tsv from

https://xenabrowser.net/datapages/?dataset=TCGA-GBM.htseq_counts.tsv&host=https%3A%2F%2Fgdc.xenahubs.net&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443

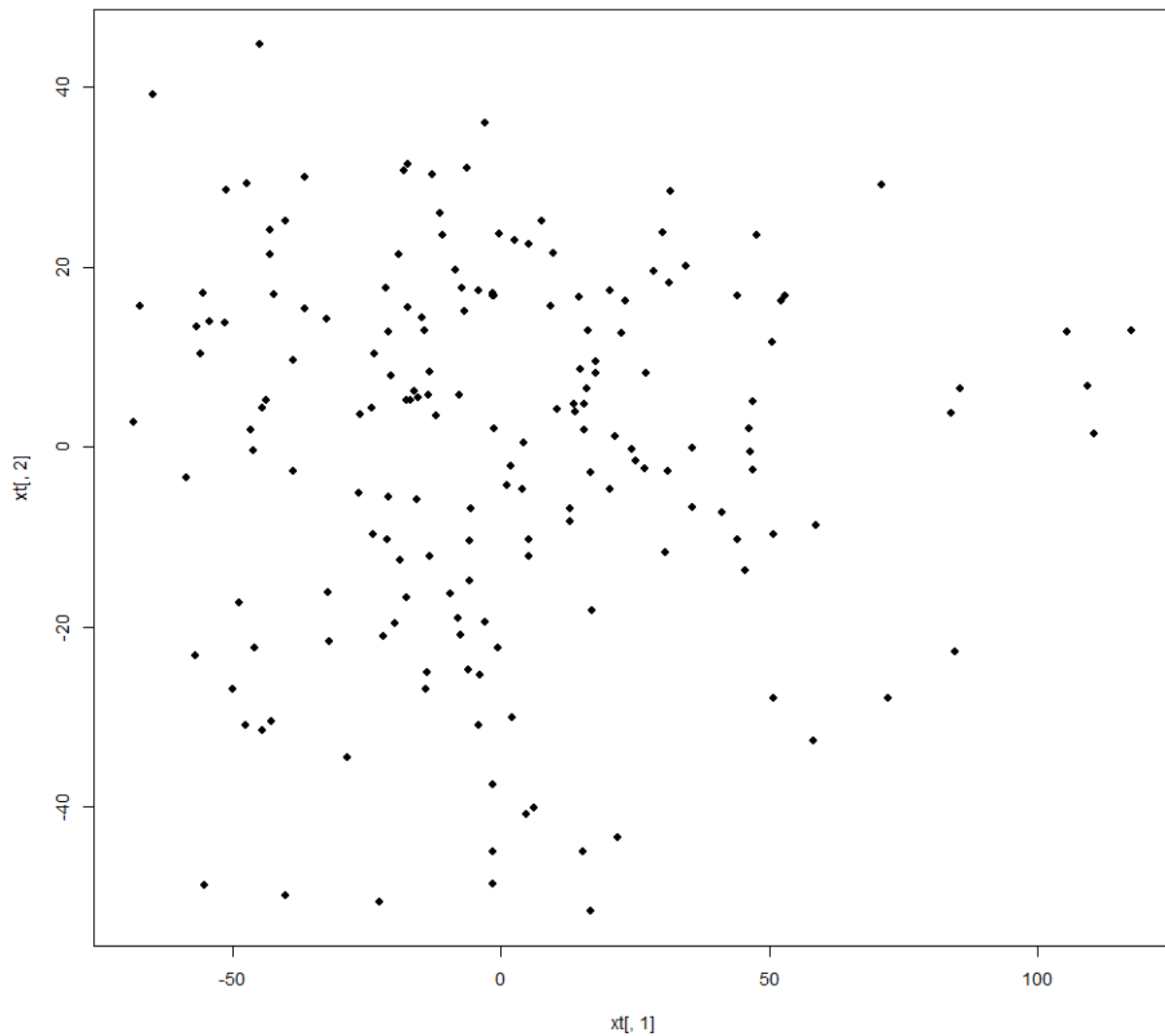
Phe.tsv from

https://xenabrowser.net/datapages/?dataset=TCGA-GBM.GDC_phenotype.tsv&host=https%3A%2F%2Fgdc.xenahubs.net&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443

```
1 dir_project1 <- "C:/Users/choyo/Desktop/CODEDATA/biomeddatasci/II/hw1/"
2 dir_res1 <- dir_project1
3
4 # load in Data
5 f_cnv1 <- paste(dir_project1, 'cnv.tsv', sep='/')
6 f_phe1 <- paste(dir_project1, 'phe.tsv', sep='/')
7
8 data1 <- read.delim(f_cnv1, sep='\t')
9 phe1 <- read.delim(f_phe1, sep='\t')
10
11 # Count null values in phe1
12 null_count <- apply(phe1, 2, function(x) sum(is.na(x)))
13 remove_cols <- names(null_count[null_count > 100])
14
15 # Create a logical vector to select columns for removal
16 cols_to_remove <- names(phe1) %in% remove_cols
17
18 # Remove selected columns
19 phe1 <- subset(phe1, select = !cols_to_remove)
20 phe1 <- phe1[complete.cases(phe1),]
21
22 phe1 <- phe1[,-1] #drop the 1st column which are TCGA-*
23 data1 <- data1[,-1] #drop the 1st column which are TCGA-*
24
25 phe1 <- as.matrix(phe1) #drop the 1st column which are TCGA-*
26 data1 <- as.matrix(data1) # convert data into numeric matrix
27
28 phe1 <- apply(phe1, 2, as.numeric) #drop the 1st column which are TCGA-*
29 data1 <- apply(data1, 2, as.numeric)
```

PCA Demo

```
29 #PCA for count numbers
30 r_sd1 <- apply(data1, 1, sd) # row (probe/gene) standard deviation (sd)
31 idx_row1 <- which(r_sd1 >= (sort(r_sd1, decreasing=T)[1000]))
32 data2 <- data1[idx_row1,]
33 dim(data2)
34 data2 <- t(data2)
35 dim(data2)
36 data2 <- apply(data2, 2, as.numeric)
37 y_pca <- prcomp(data2)
38 attributes(y_pca)
39 summary(y_pca)
40 xt <- y_pca$x
41 plot(xt[,1], xt[,2], pch=16)
```



NMF Demo

```
62 #NMF
63 install.packages('NMF')
64 library('NMF')
65 y_nmf <- nmf(t(data2), 2)
66 w1 <- y_nmf@fit@W
67 h1 <- y_nmf@fit@H
68 plot(h1[1,], h1[2,], pch=16)
```

