load data

```
load(file="bottomly.Rdata")

edata <- as.matrix(exprs(bottomly.eset))
edata <- edata[rowMeans(edata) > 10, ]
edata <- log2(as.matrix(edata) + 1)
```

Homework Problem 1: Make one heatmap of the Bottomly data with the following options: a) both rows and columns are clustered, b) show a dendrogram only on the columns., and c) scale in the column direction. Send only one heatmap. If you are unsure, check the help document on this function by typing ?heatmap.2

```
my_palette <- colorRampPalette(c("blue", "white", "orange"))(n = 299)

pdf(file = "Fedorczyk_problem1_a.pdf",height=9,width=9)
heatmap.2(edata,
          main = "Bottomly et al. rows columns clustered", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(12,9),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          dendrogram="none",     # only draw a row dendrogram
          scale = "row")
dev.off()

pdf("Fedorczyk_problem1_b.pdf",height=9,width=9)
heatmap.2(edata,
          main = "Bottomly et al. - with dendrogram", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(12,9),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          dendrogram="column",     # only draw a column dendrogram
          scale = "row")
dev.off()

pdf("Fedorczyk_problem1_c.pdf",height=10,width=10)
heatmap.2(edata,
          main = "Bottomly et al. - scaled in the column direction", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(12,9),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          dendrogram="column",     # only draw a row dendrogram
          scale = "column")
dev.off()
```

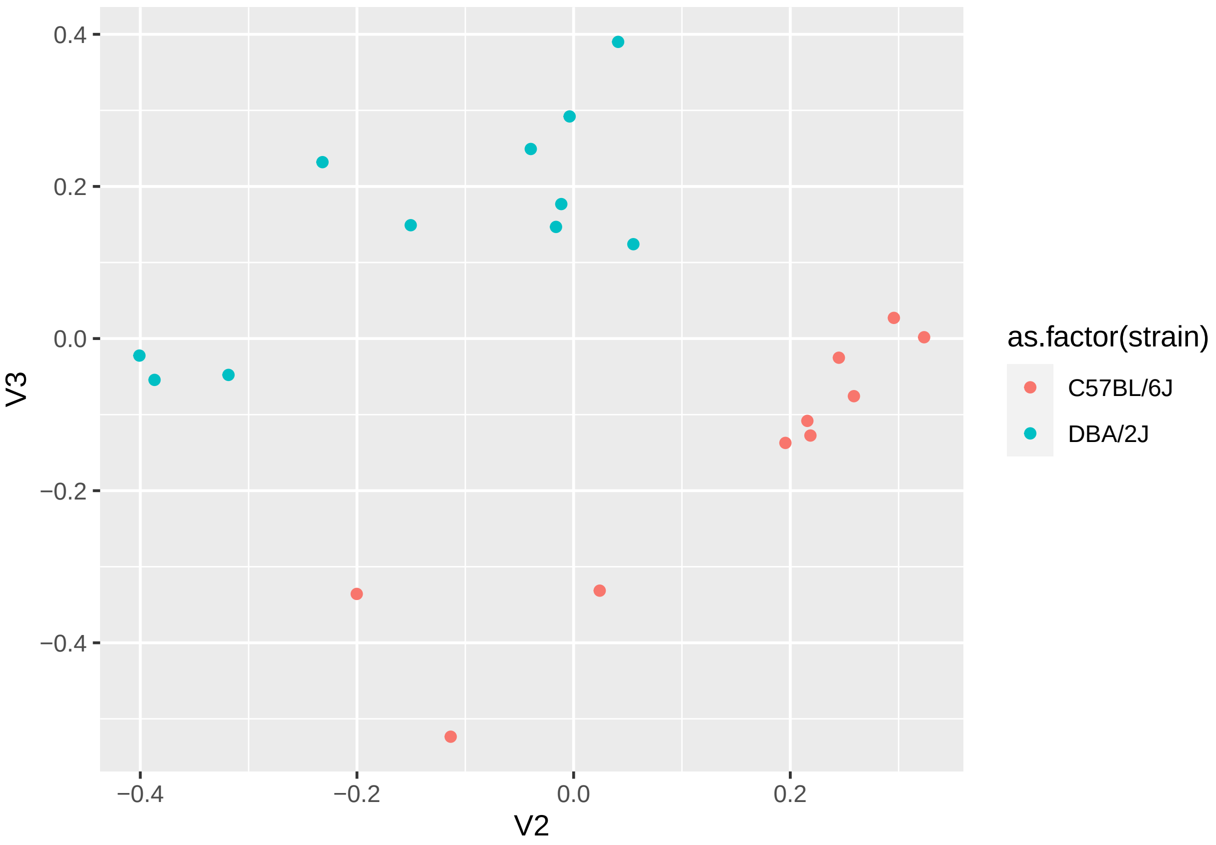
Homework Problem 2: Explore different combinations of PCs in scatter plots while coloring the data points by the genetic strains. Find a combination of PCs that separate the strains well. Send only one scatterplot.

```
edata <- t(scale(t(edata), scale=FALSE, center=TRUE))
svd.out <- svd(edata)

PC = data.table(svd.out$v,pData(bottomly.eset))

# combinations <-  combn(PC[,1:21], 2, simplify = FALSE)
# ncomb <- length(combinations)
#
# for (i in 1:ncomb){
#   df <- as.data.frame(combinations[i])
#   df['strain'] <- PC$strain
#   comb_plot <- ggplot(df) + geom_point(aes_string(x=names(df)[1], y=names(df)[2], colour = "strain"))
#   print(comb_plot)
# }

#chosen plot
ggplot(PC) + geom_point(aes(x=V2, y=V3, col=as.factor(strain)))
```

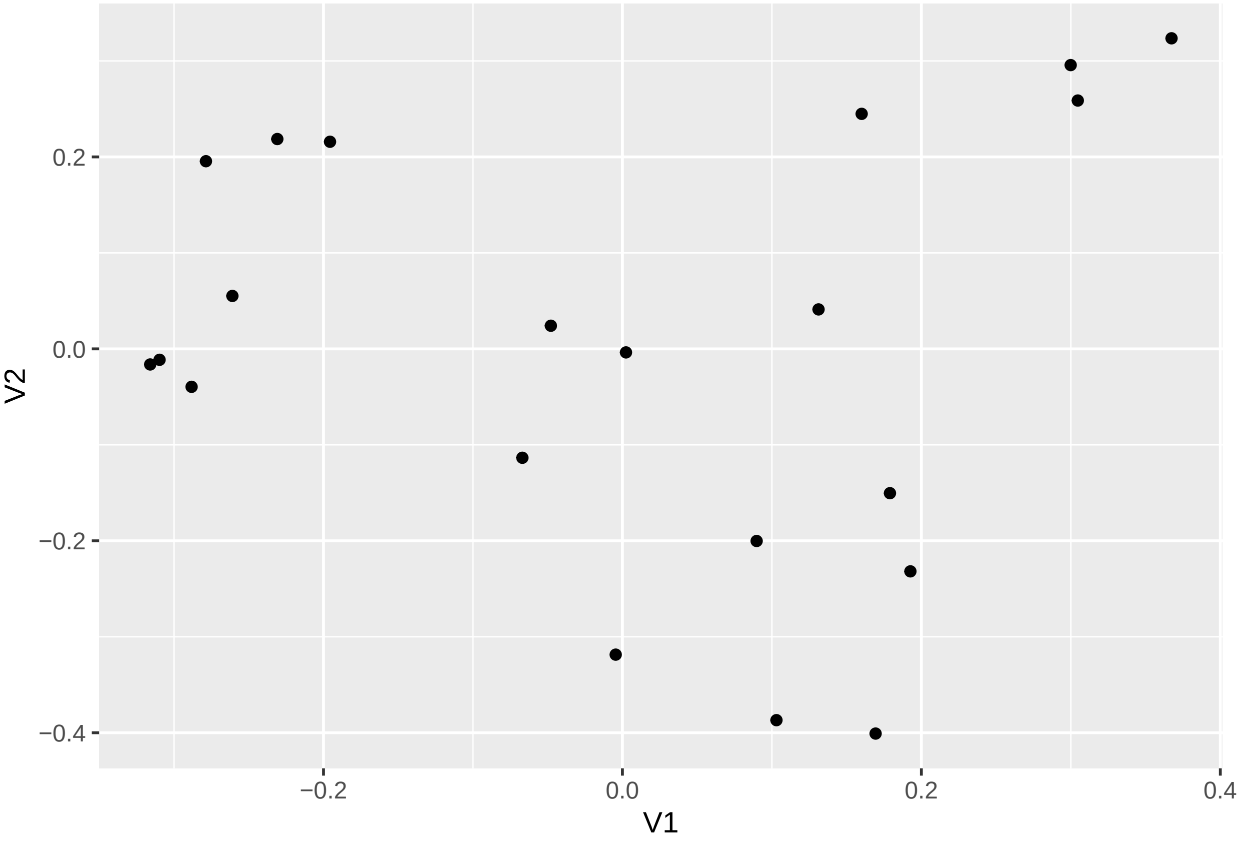


```
pdf("Fedorczyk_problem2.pdf")
ggplot(PC) + geom_point(aes(x=V2, y=V3, col=as.factor(strain)))
dev.off()
```

```
## pdf
##   2
```

Homework Problem 3: Make a scatter plot of the top 2 left singular vectors.

```
ggplot(PC) + geom_point(aes(x=V1, y=V2))
```

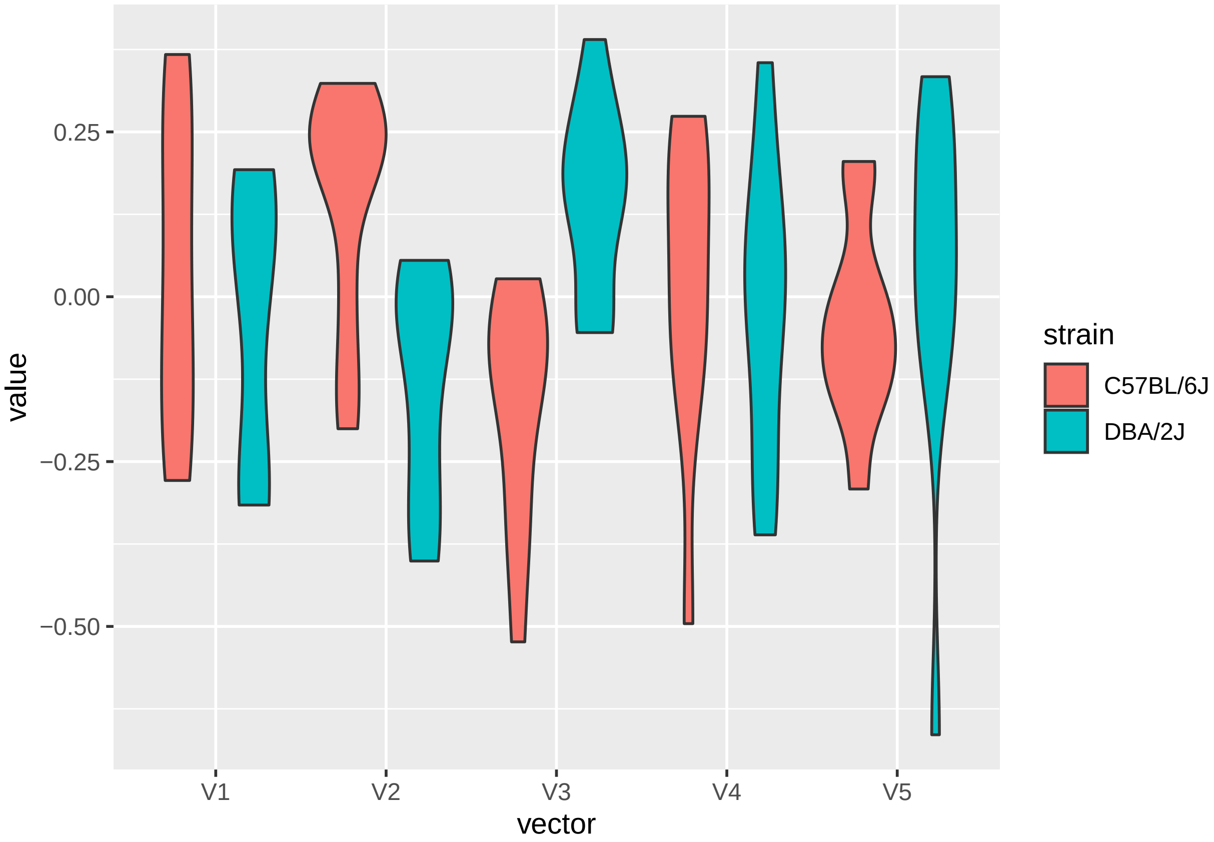


```
pdf("Fedorczyk_problem3.pdf")
ggplot(PC) + geom_point(aes(x=V1, y=V2))
dev.off()
```

```
## pdf
##   2
```

Homework Problem 4: Make one figure that contains violin plots of the top 5 left singular vectors (loadings). Hint/To-do: Make sure turn the top 5 left singular vectors into a data.table (or a data.frame) and ggplot2 to plot them altogether. Do not send 5 figures!

```
five_top <- data.frame(strain = rep(PC$strain, 5))
five_top$value <- PC[, 1:5] %>% as.matrix %>% as.vector
five_top$vector <- rep(c("V1", "V2", "V3", "V4", "V5"), each = 21)
ggplot(five_top) + geom_violin(aes(x = vector, y=value, fill = strain))
```



```
pdf("Fedorczyk_problem4.pdf")
ggplot(five_top) + geom_violin(aes(x = vector, y=value, fill = strain))
dev.off()
```

```
## pdf
##   2
```

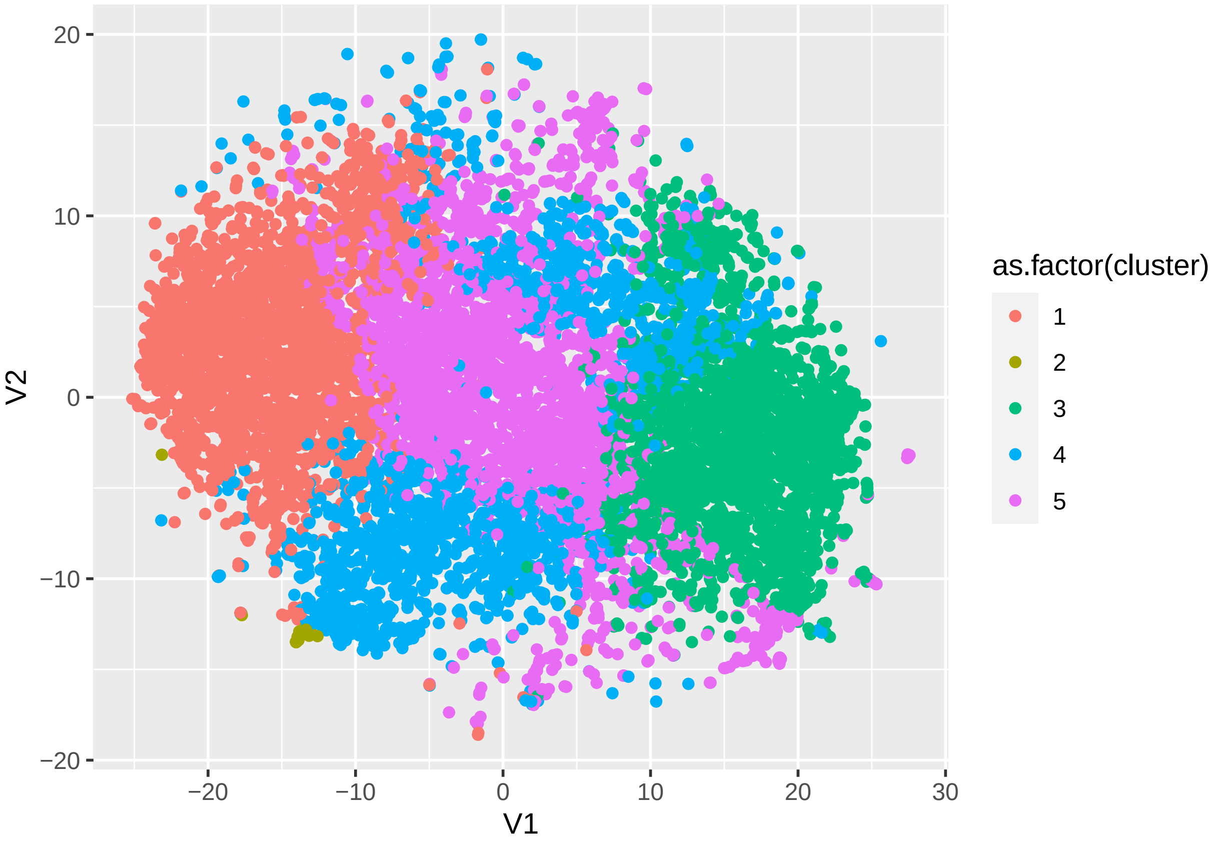
Homework Problem 5: Cluster the genes (rows) using K-means clustering (function `kmeans()`) on the original data, with `k=5` clusters. Then, create a 2-dimensional t-SNE projection (as done previously) while using the 5 clusters to color the data points corresponding to genes.

```
k5 <- kmeans(edata, centers = 5)

set.seed(1)
tsne_out <- Rtsne(edata,pca=FALSE,perplexity=60)
tsne_out = data.table(tsne_out$Y)

tsne_out$cluster <- factor(k5$cluster)

ggplot(tsne_out) + geom_point(aes(x=V1, y=V2, col = as.factor(cluster)))
```



```
pdf("Fedorczyk_problem5.pdf")
ggplot(tsne_out) + geom_point(aes(x=V1, y=V2, col = as.factor(cluster)))
dev.off()
```

```
## pdf
##   2
```