Lab2

Data visualization

Jessica Temporal 7547611

August 27, 2016

Contents

1. Use the Spellman yeast cell cycle dataset (spellman.txt).

```
file <- "spellman.txt"
```

2. a) Read into R (Hint: using the read.table() function with a "header" argument is one method to do this).

```
spellman_data <- read.table(file, header = T)
```

2. b) Set the row names to the first column, then remove this first column.

```
rownames(spellman_data) <- spellman_data$row.names
spellman_data$row.names <- NULL
```

3. a) Look at the dimensions of the data frame and make sure that there are 6,178 genes and 77 arrays/sample.

```
dim(spellman_data)
```

```
## [1] 6178   77
```

3. b) Isolate only the cdc15 experiment (samples 23-46), pick a gene with some missing values (I use gene #2/YAL002W in the solutions), and impute with the row mean (save as something).

```
cdc15 <- subset(spellman_data, select = c(23:46))
cdc15["YAL004W",]
```

```
##         cdc15_10 cdc15_30 cdc15_50 cdc15_70 cdc15_80 cdc15_90 cdc15_100
## YAL004W       NA       NA       NA     -1.5    -0.03     -1.2     -0.06
##         cdc15_110 cdc15_120 cdc15_130 cdc15_140 cdc15_150 cdc15_160
## YAL004W     -1.78      0.14     -1.13     -0.13     -1.27     -0.27
##         cdc15_170 cdc15_180 cdc15_190 cdc15_200 cdc15_210 cdc15_220
## YAL004W     -0.94      0.14        NA      1.04      0.48      1.94
##         cdc15_230 cdc15_240 cdc15_250 cdc15_270 cdc15_290
## YAL004W      1.62      1.73      1.22        NA        NA
```

```
yal004w_mean <- rowMeans(cdc15["YAL004W",], na.rm = T)
yal004w_mean
```

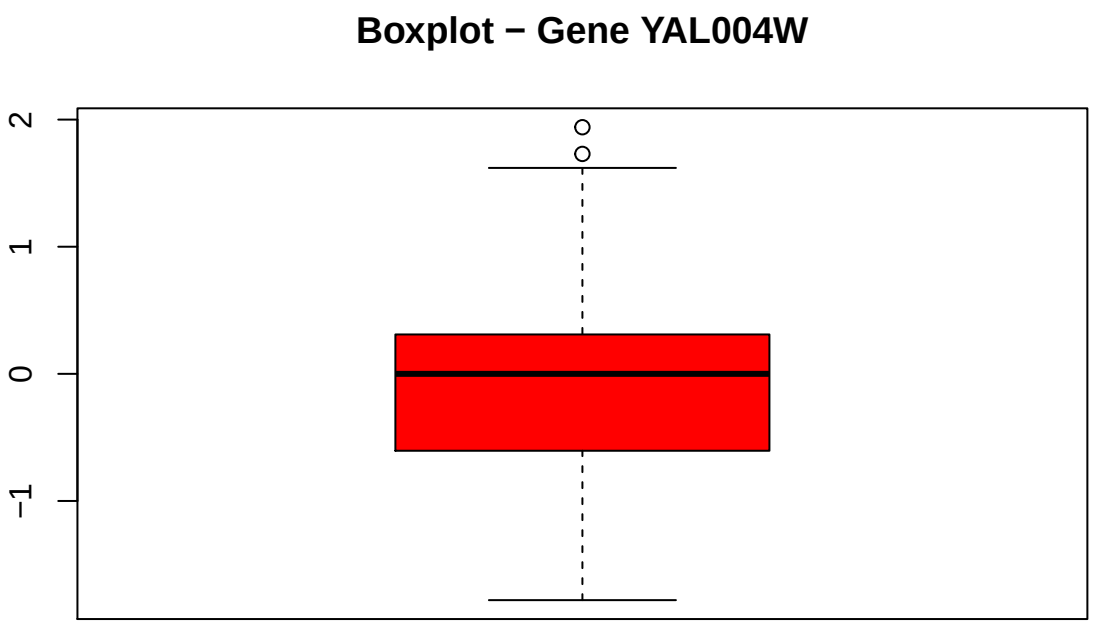
```
##      YAL004W
## 1.079383e-17
```

```
for (x in 1:length(cdc15["YAL004W",])) {
    if (is.na(cdc15["YAL004W",x])) {
        cdc15["YAL004W",x] <- yal004w_mean
    }
}
yal <- cdc15["YAL004W",]
yal
```

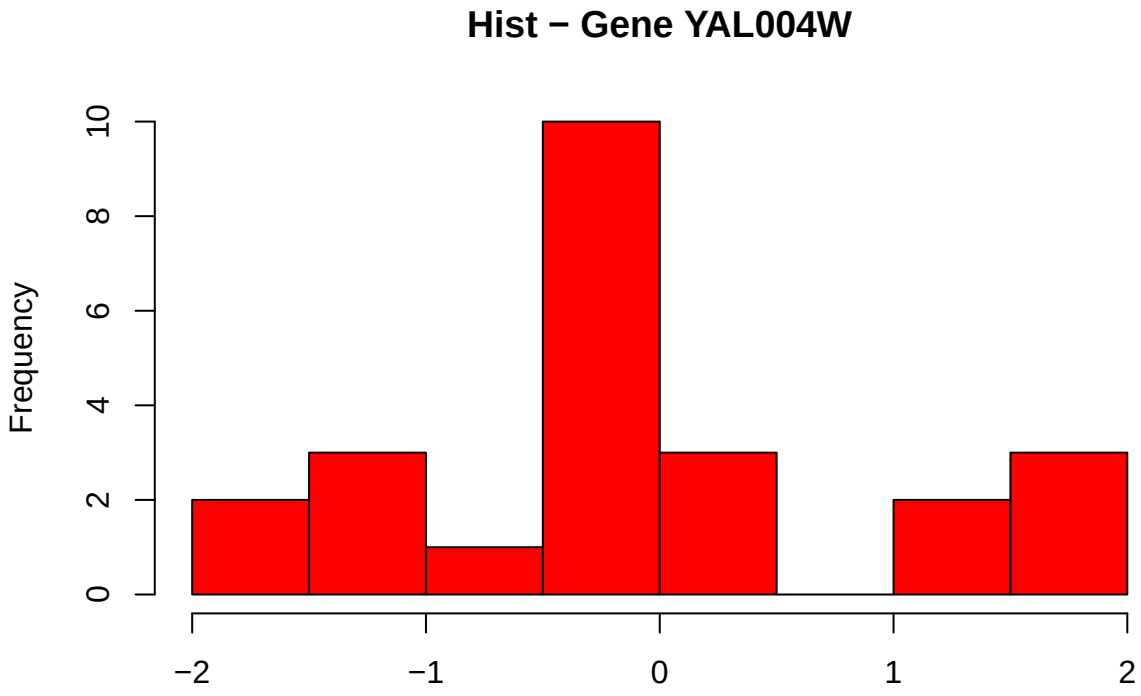
```
##             cdc15_10     cdc15_30     cdc15_50 cdc15_70 cdc15_80 cdc15_90
## YAL004W 1.079383e-17 1.079383e-17 1.079383e-17     -1.5    -0.03     -1.2
##         cdc15_100 cdc15_110 cdc15_120 cdc15_130 cdc15_140 cdc15_150
## YAL004W     -0.06     -1.78      0.14     -1.13     -0.13     -1.27
##         cdc15_160 cdc15_170 cdc15_180    cdc15_190 cdc15_200 cdc15_210
## YAL004W     -0.27     -0.94      0.14 1.079383e-17      1.04      0.48
##         cdc15_220 cdc15_230 cdc15_240 cdc15_250    cdc15_270    cdc15_290
## YAL004W      1.94      1.62      1.73      1.22 1.079383e-17 1.079383e-17
```

4. Look up the functions for boxplot and hist and plot the gene. Color the plots red and title them. Make sure that the vector is numeric (as.numeric).

```
boxplot(as.numeric(yal), col = "red", main = "Boxplot - Gene YAL004W")
```




```
hist(as.numeric(yal), freq = T, col = "red",
    main = "Hist - Gene YAL004W", xlab = "")
```



5. Generate a profile plot of the same gene. Title the plot. Use lwd in the plot command (lwd=line width).

```
plot(
    x = 1:length(yal),
    y = as.numeric(yal),
    type = "o",
    lwd = 1,
    main = "Profile of YAL004W",
    col = "red",
    xlab = "",
    ylab = ""
)
```