EDA Quiz R/Statistics

James Bullard

January 20, 2008

Please do this quiz using Sweave. This will allow for easy presentation of the results!

1. Given the following time-series microarray data construct a barplot for the gene expression through the 12 time points. Add a line to the barplot which is the trimmed average with a window size of $K = 3$. Note, one thing to remember is that barplot returns a vector of x positions, it does so with the function invisible, we haven't seen this so check out the help: ?invisible. (see figure: (1))

```
> states <- numeric(12)
> states[1] <- rpois(1, lambda = 4) + 1
> for (i in 2:12) states[i] <- rpois(1, lambda = states[i - 1]) +
+     1
> dta <- rexp(12, 1/states)
```

2. Add other smooth lines to the plot above. Namely, try a lowess smoothed line. Are there transformations of the data which help make any patterns stand out?

3. As in most microarray experiments the measurements above were actually produced from a summary statistic of a number of lower-level measurements, i.e., we have J probes for each gene and we summarize these to produce a number which is our "gene-level" measure. We have taken averages above to produce the data, but we are concerned that this is to dramatically effected by outliers. Using the data below explore the distributions of the probe level measures for the timepoints, produce a plot like figure (2). Below the red line is the boxplot means and the yellow line is the boxplot trimmed means - it should be easy to produce both if you have solved question 1.

```
> rawData <- sapply(states, function(s) {
+     rexp(22, 1/s)
+ })
```
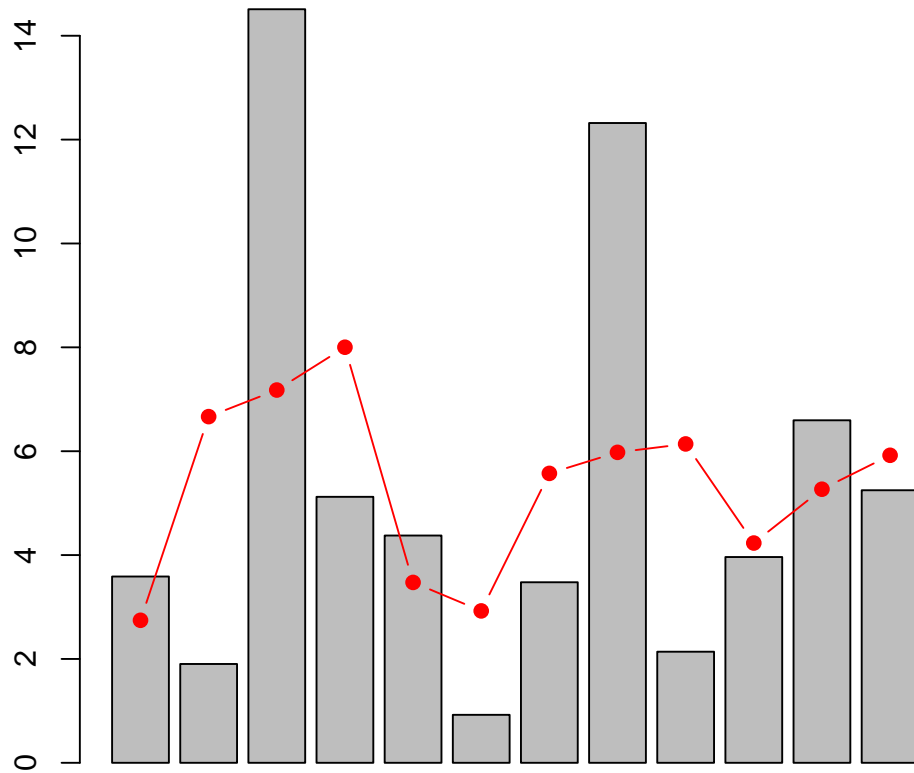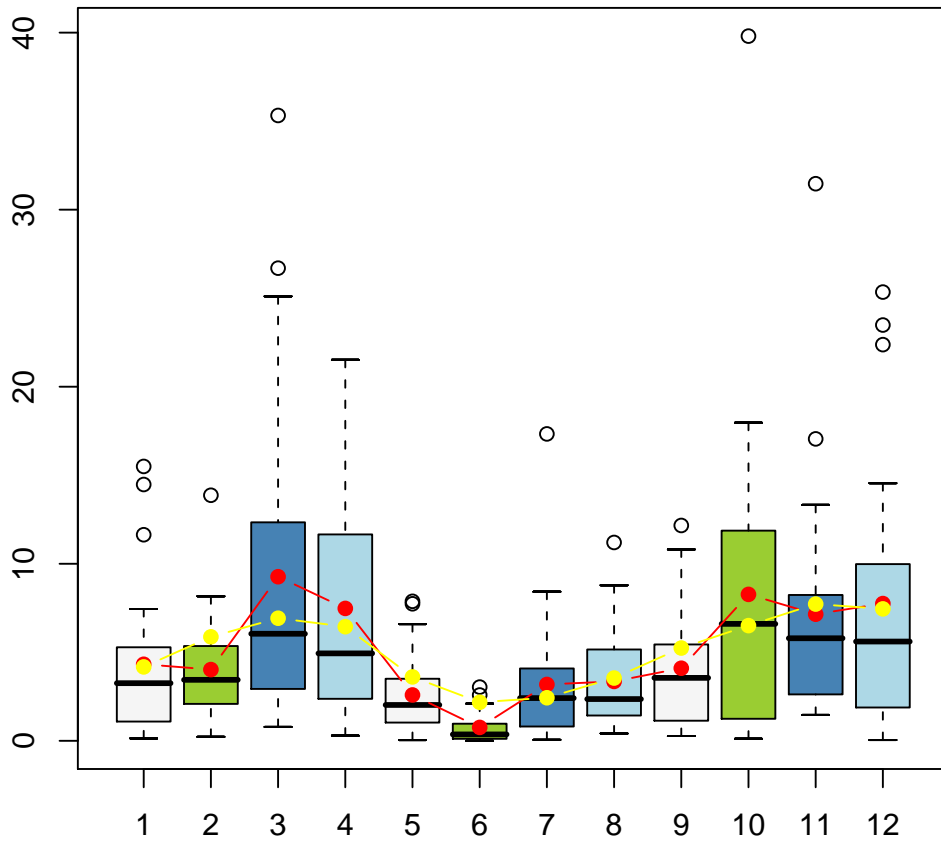
Figure 1: barplot of gene level expression measures from a microarray experiment

Figure 2: running box plots