Long Problem Set 01

```
# load data and source files needed for this long problem set
load("LPS01.RData")
source("iqr.R")
```

1. For NOX as an analyte, report the mean, the median, the standard deviation, the IQR, and the MAD for the month of January and for the month of July. Comment briefly on your results.

```
# results for January
mean(jan.data$NOX, na.rm = TRUE)
```

```
## [1] 383.5581
```

```
median(jan.data$NOX)
```

```
## [1] 339.1
```

```
sd(jan.data$NOX, na.rm = TRUE)
```

```
## [1] 184.0559
```

```
iqr(jan.data$NOX)
```

```
## [1] 204.4
```

```
mad(jan.data$NOX, constant = 1)
```

```
## [1] 89.19998
```

```
# results for July
mean(jul.data$NOX, na.rm = TRUE)
```

```
## [1] 257.2806
```

```
median(jul.data$NOX)
```

```
## [1] 270.8
```

```
sd(jul.data$NOX, na.rm = TRUE)
```

```
## [1] 96.69015
```

```
iqr(jul.data$NOX)
```

```
## [1] 155.8
```

```
mad(jul.data$NOX, constant = 1)
```

```
## [1] 57.6
```

Note the use of our iqr function in place of R's IQR function, and note that we add constant = 1 to the mad() function to give values consistent with how we defined the median absolute difference. The most interesting observations from these results is that the concentration of NOX is higher in January than in July (mean and median), and that the variation in the daily concentrations of NOX is greater in January than in July (standard deviation, IQR, and MAD). This certainly suggests that temperature most play a role of some sort in determining the concentration of NOX.

2. For NOX as an analyte, construct a histogram and a stripchart for the month of January and for the month of July. Comment briefly on your results.

```
# split the plotting window in a 2 by 2 matrix
old.par = par(mfrow = c(2, 2))

# make the plots: jitter shifts the points randomly in the verticl direction
# cex is used to make the points smaller
hist(jan.data$NOX, main = "January",
     xlab = "micrograms NOX/cubic meter")
hist(jul.data$NOX, main = "July",
     xlab = "micrograms NOX/cubic meter")
stripchart(jan.data$NOX, method = "jitter", jitter = 1, pch = 19,
        cex = 0.25, main = "January",
        xlab = "micrograms NOX/cubic meter")
stripchart(jul.data$NOX, method = "jitter", jitter = 1, pch = 19,
        cex = 0.25, main = "July",
        xlab = "micrograms NOX/cubic meter")
```
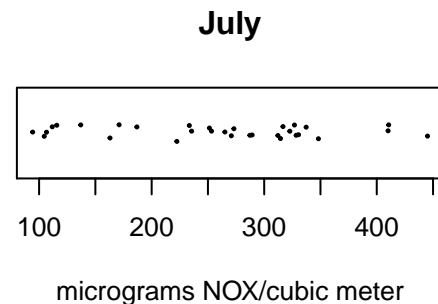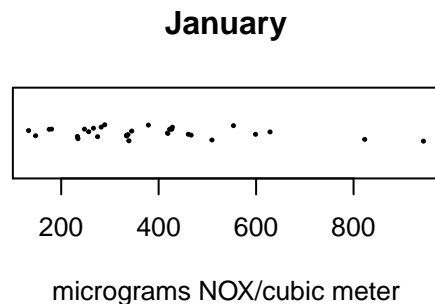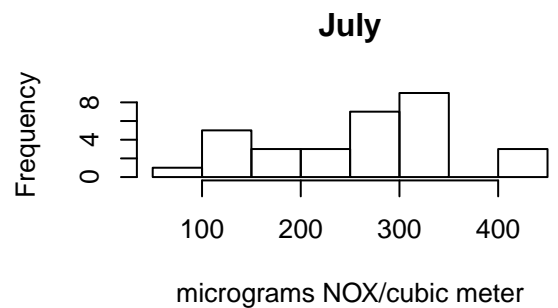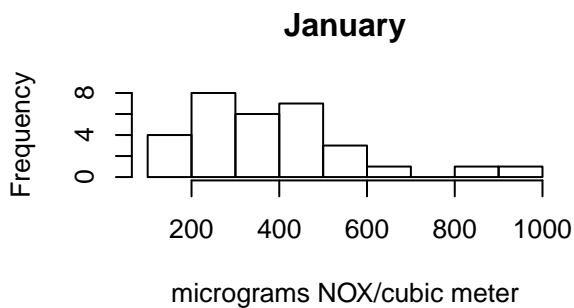


```
# restore the old plotting parameters
par(old.par)
```

Both the histograms and the stripcharts show the greater variation in the daily concentrations of NOX for January relative to those for July. We also see that the results for July are somewhat more symmetically distributed relative to those for January, which show a long tail toward higher values. This is somewhat easier to see in the stripchart below which shows the data for both months on a single scale (note that the two data sets are passed to the function as a list). This suggests, again, that there must be some relationship between NOX and temperature.

```
# it is helpful to examine two sets of data on the same scale; thus
stripchart(x = list(jan.data$NOX, jul.data$NOX), method = "jitter", pch = 19,
           cex = 0.5, xlab = "micrograms NOX/cubic meter",
           ylab = "1: January and 2: July")
```
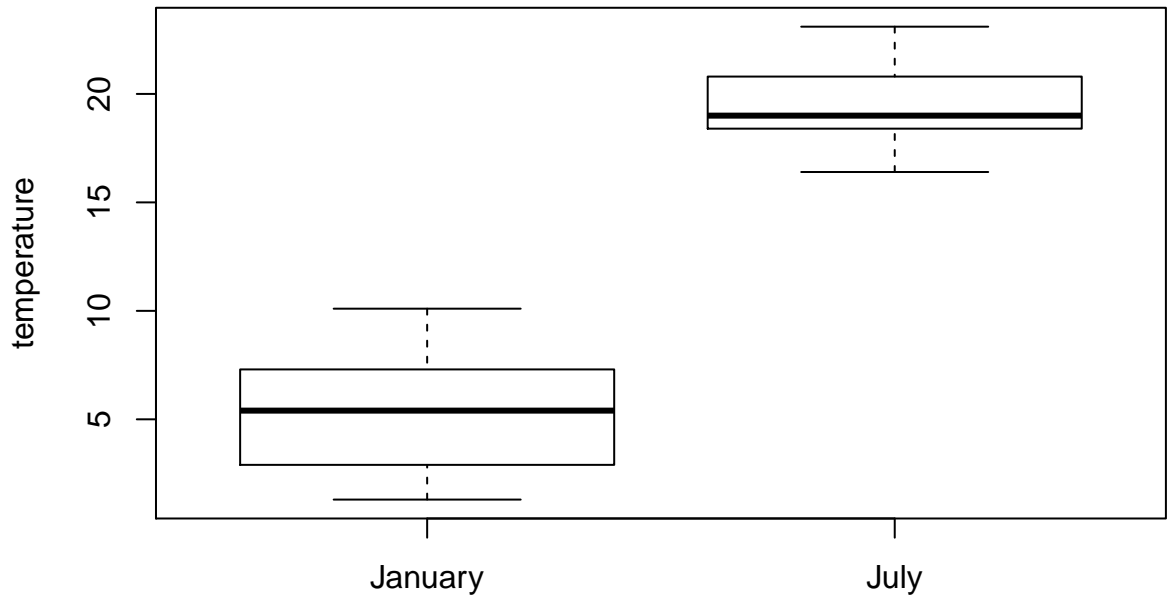
microgramas NOX/cubic meter
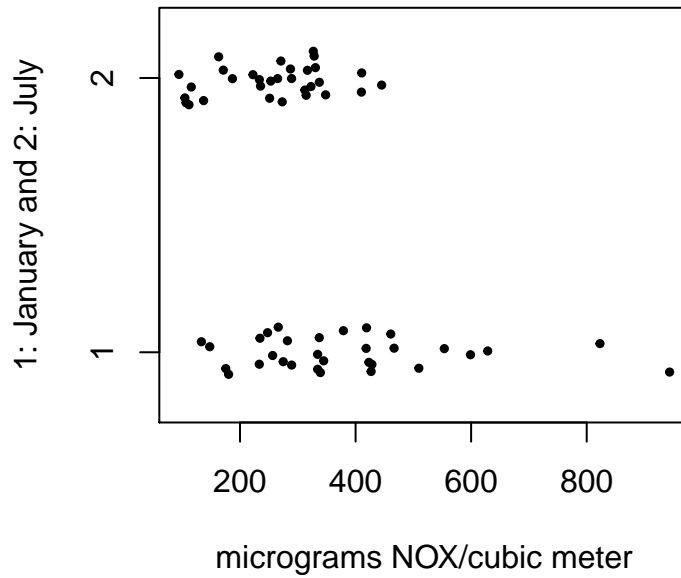
3. For NOX as an analyte, create a single panel that has separate boxplots for January and for July. Repeat using temperature in place of NOX. Comment briefly on your results.

```
boxplot(jan.data$NOX, jul.data$NOX, names = c("January", "July"),
        ylab = "micrograms NOX/cubic meter")
```
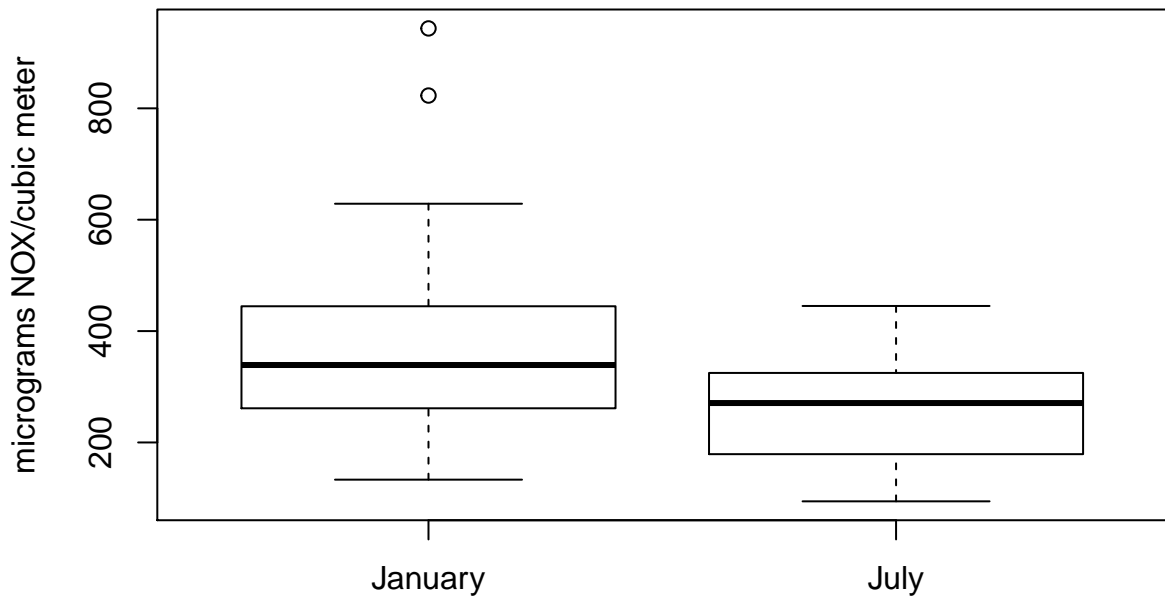


```
boxplot(jan.data$TEMP, jul.data$TEMP, names = c("January", "July"),
        ylab = "temperature")
```

The boxplot for NOX show nicely that there were two days in January with unusually high concentrations. We also see here in the overlap of the two boxes, as we do above for the stripcharts, that there are many days across the two months where the concentration of NOX are similiar in value. Contrast this with the boxplots for temperature that show, as we might well expect, that there is a significant difference in seasonal temperatures.

4. Create a single scatterplot of temperature (on the x-axis) vs. NOX (on the y-axis) using the data for January and for July. Use different point symbols and different colors for the two months. Add a legend to your plot and a set of grid lines. Set the lower limit of each axis to zero and choose upper limits that will display all points. Give the axis labels appropriate names and add a suitable main title to your plot. See the document *Creating Plots Using R's Base Graphics* for hints on how to create and modify plots. *Note: The `plot()` command plots a point only when both variables for a day are numeric; if either variable's value is missing (`NA`), the function simply ignores that day when plotting your data.*

```r
# find max NOX and max TEMP to set upper limit for x and y axes
max.nox = max(c(jan.data$NOX, jul.data$NOX), na.rm = TRUE)
max.temp = max(c(jan.data$TEMP, jul.data$TEMP), na.rm = TRUE)

# now we create the plot
plot(x = jan.data$TEMP, y = jan.data$NOX, pch = 19, col = "blue",
     xlim = c(0, max.temp), ylim = c(0,max.nox),
     xlab = "temperature",
     ylab = "[NOX] in micrograms per cubic meter",
     main = "NOX vs. temperature at Marlyebone")
points(x = jul.data$TEMP, y = jul.data$NOX, pch = 15, col = "red")
grid()
legend(x = "topright", legend = c("January", "July"),
       pch = c(19, 15), col = c("blue", "red"), bty = "n")
```

NOX vs. temperature at Marlyebone