# Solutions: Box Plots and Scatter Plots

# 1 Load the Hot Dog Data

```
myFile <- "hotdog.csv"
hd <- read.csv(file = myFile,
               skip = 1,
               nrows = 54,
               header = TRUE)

hd$Day <- factor(hd$Day,
                 levels = c("Mon", "Tue", "Wed", "Thu", "Fri"))
```

## 1.1 Solutions

1. What do the arguments 'skip=', 'nrows=' and 'header=' do in the read.csv() function?

   - skip=: number of lines to skip from the beginning of csv file
   - nrows=: number of rows to read (after the skip) from the csv file
   - header=: first row that is read contains column names

2. Why do we reorder the levels of the Day column?

   - The levels are automatically set to alphabetical order when using the read.csv() function. We prefer them to be in a meaningful order.

3. **True or False**: We reorder the levels of the Day column by overwriting the original column?
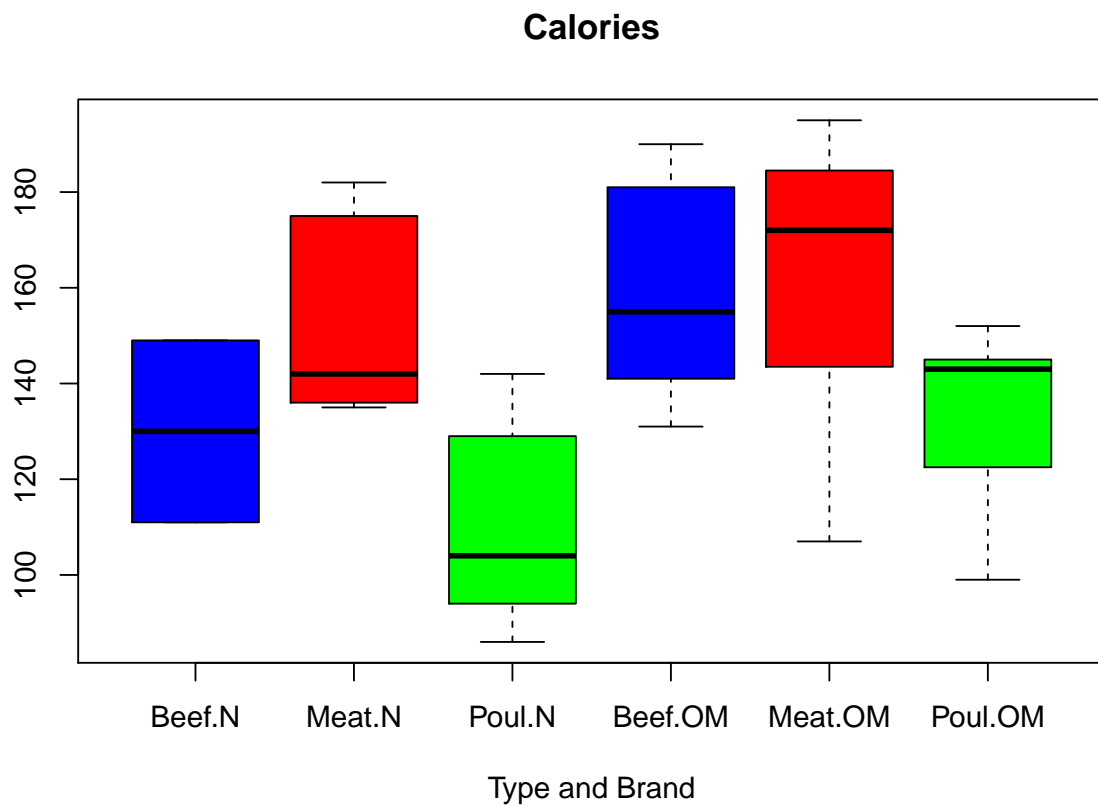
   - TRUE

# 2  Box Plots

Blanks should be filled in as:

```
boxplot(Calories ~ Type * Brand,
        data = hd,
        main = "Calories",
        xlab = "Type and Brand",
        col  =  c("blue"  ,"red"  ,"green"  ,"blue"  ,"red"   , "green"),
        xaxt = "n")

x.ticks <- c("Beef.N", "Meat.N", "Poul.N", "Beef.OM", "Meat.OM", "Poul.OM")

axis(side = 1,
     at = 1:6,
     labels = x.ticks)
```

## 2.1 Solutions

1. What does the argument 'axt = "n"' do in the `boxplot()` function?

   - Removes the x-axis tick marks and tick labels.

2. What do the arguments 'side = 1' and 'at = 1:6' do in the `axis()` function?

   - `side = 1`: references the bottom axis (i.e., x-axis)

   - `at = 1:6`: tells it to place a tick mark at the points 1 to 6

3. Create a box plot of `Sodium` by `Type`. Include the following:

   - A main title

   - Custom x and y axes titles

   - Make the box plots horizontal

   - Different colors for each hot dog type

```r
boxplot(Sodium ~ Type,
        data = hd,
        horizontal = TRUE,
        main = "Sodium",
        xlab = "mg",
        ylab = "Hot Dog Type",
        col = c("burlywood", "cornflowerblue", "springgreen"))
```
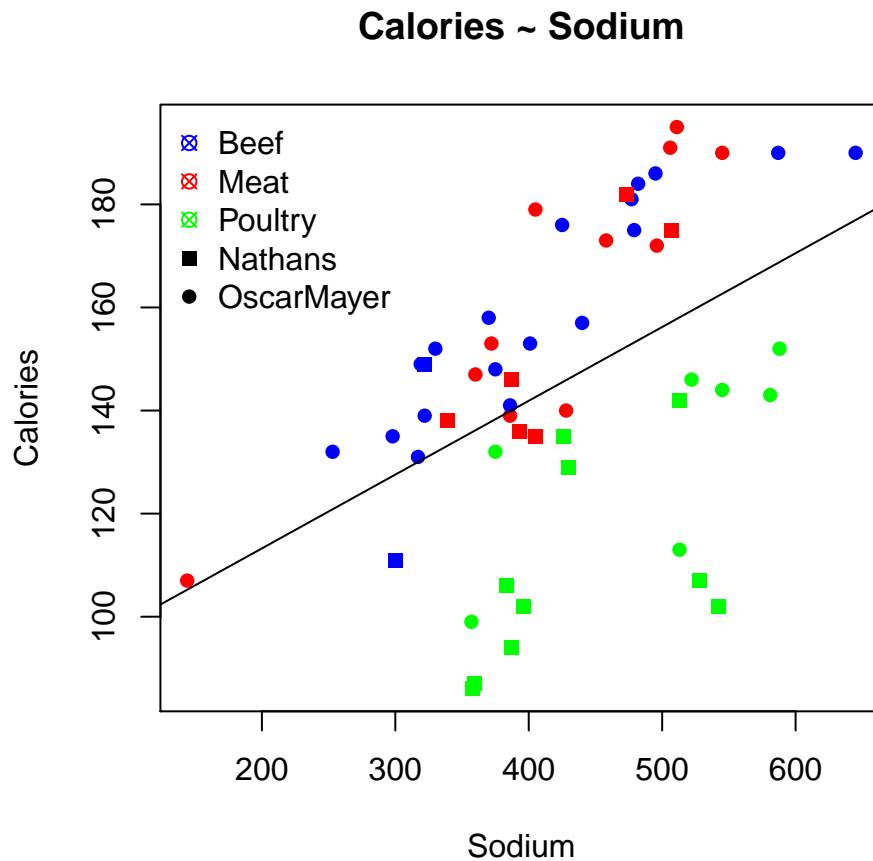
**Sodium**

Hot Dog Type

Poultry

Meat

Beef

mg

200    300    400    500    600
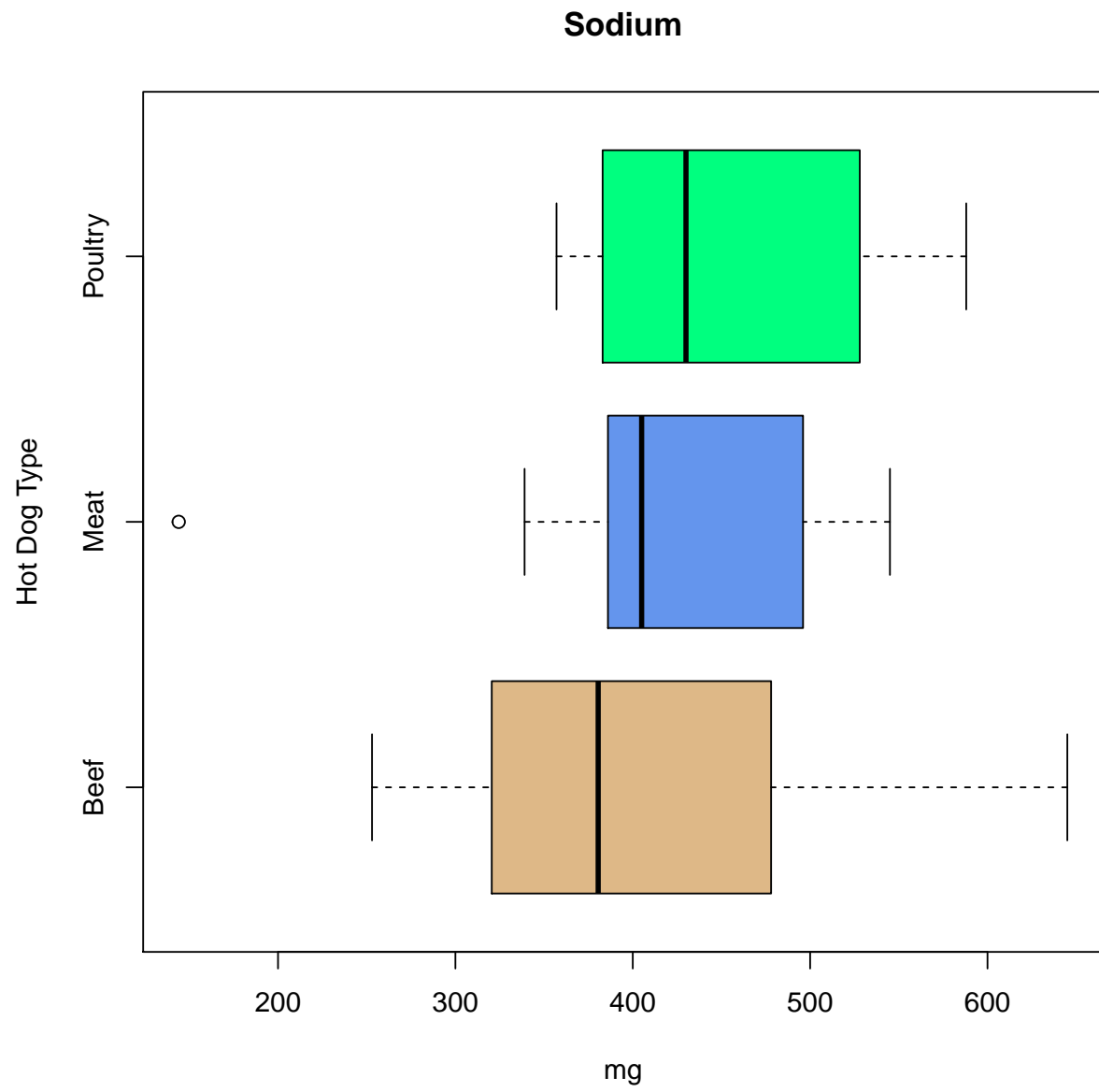
# 3 Scatter Plots

Blanks should be filled in as:

```r
plot(Calories ~ Sodium,
     data = hd,
     pch = c(15, 16)[hd$Brand],
     col = c("blue", "red", "green")[hd$Type],
     main = "Calories ~ Sodium")

abline(lm(Calories ~ Sodium, data = hd))

legend(x = "topleft",
       legend = c(levels(hd$Type), levels(hd$Brand)),
       pch = c(rep(13, 3), 15, 16),
       col = c("blue", "red", "green", "black", "black"),
       bty = "n")
```

## 3.1 Questions and Challenges

1. What does the argument 'pch=' do in the plot() function?

   - Defines the point character (i.e., the point's shape/symbol).

2. What does the command 'c(15, 16)[hd$Brand]' do?

   - Creates a vector where a 15 or 16 is repeated for each level of hd$Brand.

3. Instead of using 'x = ''topleft''' in the legend() function what can you use?

   - Another string: "topright", "bottomleft" or "bottomright"

   - Use both 'x=' and 'y=' arguments to specify a point on plot

4. What does the argument 'bty=' do in the legend() function?

   - Specifies whether or not a box should be drawn around the legend ("o" = box, "n" = no box).

5. Create a scatter plot of Sodium and Type. Include the following:

   - Sodium on the y-axis, Calories on the x-axis

   - A main title

   - Different shapes for each day (hint: pch)

   - Different colors for each brand

   - **Bonus**: Add a line of best fit and a legend

```r
plot(Sodium ~ Calories,
     data = hd,
     main = "Sodium ~ Calories",
     pch = c(15, 16, 17, 18, 19)[hd$Day],
     col = c("cyan", "salmon")[hd$Brand])
# BONUS
abline(lm(Sodium ~ Calories, data = hd))
legend(x = "topleft",
       legend = c(levels(hd$Brand), levels(hd$Day)),
       pch = c(13, 13, 15:19),
       col = c("cyan", "salmon", rep("black", 5)),
       bty = "n")
```

**Sodium ~ Calories**
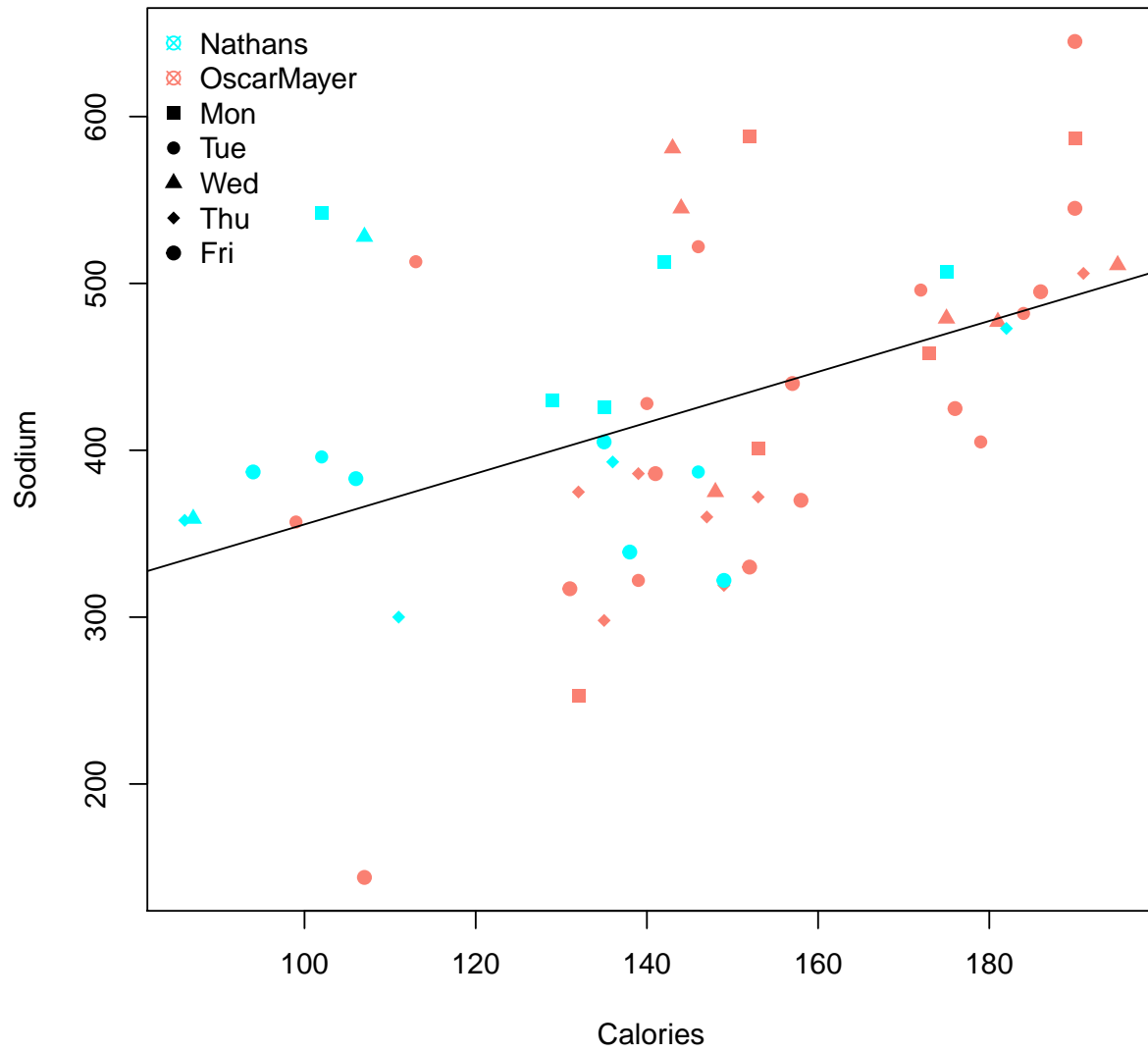
# R Session Information

```
R version 3.0.2 (2013-09-25)
Platform: x86_64-apple-darwin10.8.0 (64-bit)

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] knitr_1.5

loaded via a namespace (and not attached):
[1] digest_0.6.4   evaluate_0.5.1 formatR_0.10   highr_0.3
[5] stringr_0.6.2  tools_3.0.2
```